

Essays in Urban and Transport Economics

A THESIS SUBMITTED TO THE UNIVERSITY OF DUBLIN, TRINITY COLLEGE
IN APPLICATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY

Elisa Maria Tirindelli

Supervised by

Ronan Lyons

2023

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

Chapter 2 of this thesis is co-authored with Ronan Lyons. Chapter 4 of this thesis is co-authored with Paolo Beria.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Elisa Maria Tirindelli

Non-Technical Summary

This dissertation consists of three essays at the intersection of urban and transport economics. While they do not use common data or methods, they all tackle fundamental issues concerning the effects of transport on people's life and on urban and extra-urban dynamics.

Chapter 2 contributes to an intensively discussed topic, namely the emergence of Zipf's Law and Gibrat's Law. The former posits that cities size and its rank are inversely proportional, while the latter points out that there is no correlation between cities size and their growth level. It focuses on a sample of cities in Britain, between 1801 and 2011, and examines their size, rank and growth level to show that conclusions depend on city definition, sample cutoff and regression methods. It also points out that Zipf's Law cannot be rejected under the strongest combination of data and methods, unlike if other data or methods are used. Finally, across Zipf, Gibrat and Gini analyses, it finds that urban concentration in Britain peaked in the mid-19th century but fell 1861-1911 and 1951-1991.

Chapter 3 examines the impact of investments in transport infrastructure in an urban context. It focuses in particular on the case of Dublin and, through the calibration of a structural model, it retrieves the necessary parameters to obtain a measure of improved market access, which reflects the reduced commuting costs, induced by the presence of a new light rail metro system, the *Luas*. Subsequently, it inspects the relation between improved market access and population density, skill composition and rent level. It finds that improved market access increases the share of resident population, however it points out that the increase is driven by high-skilled workers; in fact the share of resident low-skilled workers decreases with improved market access. The opposite is true for workplace population, i.e. population working in an area. In this case, more low-skilled workers are likely to work in an area better served by transport, while no significant effect is found for high-skilled workers. Finally, it does not find an effect on long-term rents, however, because of lack of reliable data so far, it cannot rule out yet an effect on short-term, market rents.

Chapter 4 explores the effects of the process of rail liberalisation in Europe, which started more than two decades ago. While it focuses on the case of Italy, it reflects well the results

obtained in other countries for other transport markets and it therefore allows to generalise the results. This chapter points out that a reduction in prices linked to the presence of competition is very short-lived and becomes inexistent in less than a year, with little exceptions. The frequency and the level of service, on the other hand, benefit, however, in a weakly correlated manner with prices.

Acknowledgements

First and foremost, I am indebted to my supervisor Ronan Lyons, whom I want to thank for insightful comments, questions and support. His continuous guidance enabled me to progress this work the way I did. I am also thankful to Paolo Beria, who provided his advice and knowledge of the transport sector in the process of writing this thesis. Further, I want to thank Guillaume Daudin, who has been my mentor and guided me to the path of a PhD. I probably would not have started this work, had it not been for his encouragement.

I also thank Nicola Fontana, Gaia Narciso, Gabriel Ahlfeldt, Marvin Suesse, Michael Wycherly, Martina Kirchberger and Andrea Guariso for very interesting, motivating and helpful discussions. This work has benefited from comments of seminar participants in Trinity College Dublin, for which I am really grateful. I owe special thanks to Colette Ding for her indispensable assistance with numerous administrative matters.

This research was funded by the Grattan Scholarship of the Department of Economics, Trinity College Dublin, and I am grateful to Annrai O'Toole, my Grattan donor, for his essential support of this thesis.

I am grateful to my two colleagues and great friends at TRISS for making my PhD life so enjoyable. To Luca Hennig, for our great bike trips and to Benoit Voudon for our numerous movie nights, and to both for all the amazing moments spent together laughing and eating. I will cherish the memories of the time spent together.

I am also grateful to my “Parisian” friend Zahra, for her ability to read my mind and never judge but understand and give loving and rational advice. For her support to my dubious choice to follow my love for mathematics.

To Fra, for understanding who I was before I ever did, and for helping me become what I am with love and patience for so many years. Also, for her excellent advice on colours and graphs in this thesis, despite her unfamiliarity with the topic; experts in the field could not have done better.

To my sister Mary, I pretended to take care of for so many years, while she was in fact always taking care of me. For her huge support and encouragement in my choices, much as they were different from hers.

To Tommi, for all his patience and nothing more, because patience alone was already

huge on his part. In particular, to his patience in explaining math, despite his skepticism on its economic applications.

To Nene, my lifelong friend and soul-sister, who has been supporting me through all the chapters of this thesis and of my life. For all our moments spent hiking and talking and skiing and saving each other from our black holes, since very early on. There are no words to express how immensely grateful I am to her for always being there.

Finally, to all of them together for being in my life.

I dedicate this dissertation to a person who did not get to see me start it but I am sure would have been genuinely curious to read it.

A mia madre, Stefania

Contents

1	General Introduction	1
2	The Rise & Fall of Urban Concentration in Britain: Zipf, Gibrat and Gini across two centuries	5
2.1	Introduction	7
2.2	Theory & Evidence	9
2.2.1	Power Laws	9
2.2.2	Cutoffs and Exponents	11
2.2.3	Literature	14
2.3	Data	16
2.3.1	Administrative Context	16
2.3.2	Spatial Taxonomies	17
2.3.3	Samples by Cut-off	18
2.4	Analysis & Results	22
2.4.1	Pareto Exponent	22
2.4.2	Robustness check	27
2.4.3	Gibrat's Law	29
2.5	Conclusion	32
3	Skill-biased transportation change: Population, gentrification and Dublin's <i>Luas</i> light rail	35
3.1	Introduction	37
3.2	Theory & Hypotheses	39
3.2.1	Workers	40
3.2.2	Firms	42
3.2.3	Model Predictions	43
3.3	Data	45
3.3.1	The <i>Luas</i> lines	45
3.3.2	Residential & Workplace Populations	47
3.3.3	Residential Rents	48
3.3.4	Commute time	51

3.3.5	Instrumental Variables	52
3.4	Analysis	53
3.4.1	Estimating RCMA	53
3.4.2	Estimating FCMA	55
3.4.3	Results	58
3.5	Conclusion	61
4	Market segmentation and long-run competition: the case of the Italian rail market	65
4.1	Introduction	67
4.2	Italian Rail Market	69
4.3	Data	70
4.4	Trenitalia and Italo: a comparison	73
4.5	Competition effects	76
4.5.1	Competition effects on core routes	76
4.5.2	Competition on tourist routes	81
4.6	Mechanism of price increase	84
4.7	Conclusion	84
5	General Conclusion	87
	Bibliography	91
	Appendices	101
A.1	The Rise & Fall of Urban Concentration in Britain: Zipf, Gibrat and Gini across two centuries Appendix	103
A.1	Literature	103
A.1.1	Short-run analyses	103
A.1.2	Long-run analyses	105
A.2	Tables	109
A.3	Figures	115
A.4	Market segmentation and long-run competition: the case of the Italian rail market: Appendix	126
A.5	Skill-biased transportation change: Population, gentrification and Dublin's Luas light rail: Appendix	128

List of Tables

3.1	Summary statistics	46
3.2	Summary statistics of housing stock in 2002 across EDs	50
3.3	Elasticity of commute flows to commute shares	54
3.4	Estimation of kappa	57
3.5	Effect of introduction of LUAS on probability of commuting	59
3.6	Effect of introduction of LUAS on share on probability of commuting . .	60
3.7	Effect of introduction of LUAS on ED rents	61
4.1	Summary Statistics	73
4.2	Market segments characteristics and differences	76
4.3	Market segmentation	77
4.4	Averaged Diff in diff estimation on core segment	78
4.5	Pooled difference in difference on Core segment	79
4.6	Pooled difference in difference on Tourist segment	82
A.1	Review of Key Papers in the Literature	109
A.2	Summary statistics for the four units	110
A.3	Local Government District	111
A.4	Unitary Authority	111
A.5	Primary Urban Area	112
A.6	Gini Coefficient	112
A.7	Gini Coefficient	113
A.8	Spline regression for Primary Urban Area and Conservative cutoff - 1 . .	113
A.9	Spline regression for Primary Urban Area and Conservative cutoff - 2 . .	114
A.10	Spline regression for Primary Urban Area and Conservative cutoff - 3 . .	114
A.11	Pooled difference in difference on both segments	127
A.12	Effect of introduction of LUAS on ED rents	130
A.13	Placebo Effect of introduction of LUAS on probability of commuting . .	131
A.14	Effect of introduction of LUAS on share on probability of commuting . .	131

List of Figures

2.1	Local Government District	19
2.2	District/Unitary Authority	20
2.3	Primary Urban Area	21
2.4	Pareto exponent, Deviation and Conservative cutoff for Primary Urban Areas	23
2.5	Pareto exponent across spatial units, by cutoff	24
2.6	Pareto exponent across cutoff, by spatial units	25
2.7	Pareto exponent, with and without correction for bias and standard errors	26
2.8	Gini coefficient across spatial units, <i>Conservative</i> cutoff	27
2.9	Gini coefficient by cutoffs, Primary Urban Areas	28
2.10	Kernel regression: Primary Urban Area with Conservative cutoff	31
2.11	Gini coefficient and Zipf coefficient for Primary Urban Area with Con- servative cutoff	33
3.1	Luas Extensions	47
3.2	Average rent per room by ED, CSO	49
3.3	Commute time 2002	51
3.4	Commute time 2016	52
3.5	Commuting Market Access calibration	56
4.1	Passenger per km in billion	71
4.2	Summary statistics	72
4.3	Minimum Tariff - Core segment	74
4.4	Average daily frequency - Core segment	75
4.5	LR competition vs routes in competition since 2018	80
4.6	Routes in competition since July 2020 vs routes only operated by Trenitalia	83
4.7	Share of partial flexibility tickets one day before departure	85
A.1	Size-rank coefficient: Local Government Districts	115
A.2	Size-rank coefficient: Local Government Districts	115
A.3	Size-rank coefficient: District/Unitary Authority	116
A.4	Size-rank coefficient: Primary Urban Area	116

A.5	Gini coefficient for four different units	117
A.6	Gini coefficient for four different cutoffs	117
A.7	Lorenz curve for Conservative cutoff - Local Government District	118
A.8	Lorenz curve for Fraction cutoff - Local Government District	119
A.9	Lorenz curve for Level cutoff - Local Government District	120
A.10	Lorenz curve for Deviation cutoff - Local Government District	121
A.11	Lorenz curve for Conservative cutoff - Unitary Authority	122
A.12	Lorenz curve for Fraction cutoff - Unitary Authority	123
A.13	Lorenz curve for Level cutoff - Unitary Authority	124
A.14	Kernel regression: Unitary Authority with Conservative cutoff	125
A.15	Train type share on tourist segment	126
A.16	CSO rents vs Daft rents	128
A.17	Estimation of $\kappa\epsilon$	128
A.18	Wage calibration	129

Chapter 1

General Introduction

This dissertation is a collection of three essays at the intersection of transport and urban economics. While the research topics are diverse, some common elements exist. Within the area of urban economics, chapter 2 and chapter 3 are both concerned with city dynamics; the former focuses on the evolution of population size across cities while the latter examines it at within-city level. Chapter 4 and chapter 3, on the other hand, are both concerned with transport infrastructure and how they influence their users, either via prices or by reduced commuting costs for urban dwellers. Finally, even without establishing a direct link, we can think of these three chapters as a collection of studies on the effects of transport infrastructure investments. Historically and indirectly, like in chapter 2, where we study population dynamics in a period of big infrastructural changes, and directly, like in chapter 4 and 3, where we are concerned with a direct effect on urban or on overall population welfare.

The first essay (Chapter 2) provides an insight on the extent to which Zipf's Law and Gibrat's Law hold across space and time. The fraction of the global population living in cities is estimated to have been just 10% in 1800 but, by 2010, more than half the world's population lived in cities. Understanding patterns of city growth, therefore, is central in accommodating an additional three billion city-dwellers over coming decades. At the heart of patterns of city size and growth are Zipf's Law and Gibrat's Law. Zipf's Law posits that the relationship between a city's size and its rank is unit elastic – in other words, a 10% increase in a city's population leads to a 10% fall in its rank. Gibrat's Law holds that there is no systematic relationship between a city's initial size and its subsequent growth rate.

This chapter examines the relationship between city size and both city rank and subsequent city growth, over the last two centuries in (Great) Britain, supplemented by an analysis of urban concentration using the Gini coefficient. It exploits comprehensive data for all cities in Britain across 19 Censuses between 1801 and 2011, for four different measures of city size and across four different statistical cutoffs for city size. In addition, it performs a comparison across methods suggested by the literature, including parametric and non-parametric methods, and in the case of the "Zipf" exponent, with and without

correcting for bias in the coefficient and also in the standard error.

This chapter contributes to the extensive literature on city size, rank and performance by drawing four main conclusions. Firstly, it highlights the importance of research design in examining whether Zipf's Law holds, with spurious results common where arbitrary city definitions or cutoffs are used, in addition to bias and false precision. Secondly, it finds that we are unable to reject Zipf's Law in any of the 19 Census years under the strongest set-up. Thirdly, the analysis of whether Gibrat's Law holds finds evidence against it holding in the period 1861-1911, where there is a strong negative relationship between initial size and fifty-year growth. In later periods, there is only modest evidence for such a link. Finally, across all three sets of analysis, it finds evidence of rising inequality in city populations in the early 19th century, followed by greater compression across cities in the following century and a half.

The second essay (Chapter 3) uses a general equilibrium model to evaluate the impact of the introduction of a new transport infrastructure in an urban context. In recent decades, congestion and pollution have been plaguing major urban areas, to the extent that cost of time lost in traffic in Europe has been estimated to be around €110 billion in 2012 (Christidis and Ibanez Rivas (2012)) and is projected to grow. In order to mitigate this problem numerous cities and countries around the world have undertaken investments in public transportation, such as buses, light-rail networks and (heavy rail) metro networks. It is unclear however, especially in the context of a European capital city, what is the impact of said investments and how do city dwellers respond to such change.

This chapter investigates the impact of the introduction of a new light-rail metro system, the *Luas*, which was introduced in 2004 in Dublin, Ireland's capital city. It first calibrates a work-live-commute structural model, as in Ahlfeldt et al. (2015), retrieving all city-specific elasticities and obtaining a measure of resident and firm market access, which captures the introduction of the *Luas* through reduced commuting costs. It then combines detailed spatial information on residents and workplace population in different socio-economic segments and residential rents with Census data on commuting patterns, for 2002 (before the first line opened) and for 2016, within a difference in difference instrumental specification, in order to measure the impact of improved market access on population dynamics and rent level.

It finds that improved transportation has an impact both on residential and working population in the areas involved. However the distribution across occupation type is not homogeneous: higher-skill workers tend to relocate in areas close with improved access, while lower-skill workers on the other hand are more prone to work in those areas. A 10% improvement in market access bringing about a 6% increase in a district's share of population. It also points out that this impact was skill-biased, with a district's share of

the higher-skilled population rising by 7% – and its lower-skilled population falling by 5% – in response to a 10% improvement in market access. It does not find a clear effect on rents.

The third essay (Chapter 4) analyses the effects of liberalisation on the Italian rail market on prices and service level. The idea that the rail market should be a protected state monopoly has become increasingly more obsolete in the recent years and liberalisation has been central to the European agenda. The rationale is that competition decreases prices, increases passenger volume and service quality, thus increasing overall consumer benefits. However, whether these are the true effects of liberalisation and in what market segments they are more likely to materialise is still unclear. This chapter sheds light on two main issues. First, how is the Italian rail market segmented and where is competition more likely to happen, once the market is liberalised. Second, what are the long-run consequences of the rail market liberalisation in Italy, on prices and service quality.

Using data on rail prices for summer months between 2017 and 2021 and by performing a cluster analysis, this chapter points out that two distinct market segments can be identified: a higher and a lower willingness to pay segment, which deserve to be studied separately. The former encompasses high speed routes, on the core business arteries of the country, while the latter is composed of lower speed routes, connecting main cities to tourist locations. Competition is more likely to happen, at least initially, on the high willingness to pay segment. Subsequently, in relation to the entrance of a new player on the market, it points out that the true effects are not exactly as expected. Within the segment where competition started earlier, only prices for advanced booking remain reduced in the long-run, while last-minute booking prices ramp up quickly. In the segment where competition started later the effect has the opposite sign, with an increase in prices in relation to entrance of a new player. Effects on frequency of service on the other hand remain slightly positive, also in the long-run, but suffered heavily from the health crisis.

Lastly, Chapter 5 offers a conclusion.

Chapter 2

The Rise & Fall of Urban Concentration in Britain: Zipf, Gibrat and Gini across two centuries¹

¹ This paper is co-authored with Ronan Lyons.

Abstract

While city size and growth are the subjects of substantial literature, consensus is lacking on the extent to which Zipf's Law or Gibrat's Law holds across space and time. We examine city size, rank and growth in Britain 1801-2011 and show conclusions depend on city definition, sample cutoff and regression methods. We find Zipf's Law cannot be rejected under the strongest combination of data and methods, unlike if other data or methods are used. Across Zipf, Gibrat and Gini analyses, we find that urban concentration in Britain peaked in the mid-19th century but fell 1861-1911 and 1951-1991.

2.1 Introduction

Over the last two centuries, cities have become integral to the global economy and to human experience. The fraction of the global population living in cities is estimated to have been just 10% in 1800 but, by 2010, more than half the world’s population lived in cities (UN-DESA, 2018). Understanding patterns of city growth, therefore, will be central in accommodating an additional three billion city-dwellers over coming decades. At the heart of patterns of city size and growth are Zipf’s Law (Zipf, 1965) and Gibrat’s Law (Gibrat, 1931). Zipf’s Law holds that the relationship between a city’s size and its rank is unit elastic – in other words, a 10% increase in a city’s population leads to a 10% fall in its rank. It is described by Krugman (1996) as “one of the most overwhelming empirical regularities in economics”. Gibrat’s Law holds that there is no systematic relationship between a city’s initial size and its subsequent growth rate.

In this paper, we examine the relationship between city size and both city rank and subsequent city growth, over the last two centuries in (Great) Britain. To do this, we assemble for the first time comprehensive data for all cities in Britain across 19 Censuses between 1801 and 2011. Our dataset allows us to measure cities in three different ways – local government districts, unitary authorities and primary urban areas – across four different statistical cutoffs for city size. In addition, we also compare across methods suggested by the literature, including parametric and non-parametric methods, and in the case of the “Zipf” exponent, with and without correcting for bias in the coefficient and also in the standard error.

Both Zipf’s and Gibrat’s Laws are the subject of intense debate among researchers, with deep implications for policymakers. Our paper contributes to this extensive literature on city size, rank and performance and in particular aims to bring clarity to researchers looking to test for the presence of either law in empirical settings. Seminal contributions include Auerbach (1913), who identified a size-rank relationship for cities, Gibrat (1931), who showed that a proportionate growth process delivers a Pareto distribution in the upper tail, and Zipf (1965), who found a similar relationship more broadly, including in word frequency.² Gabaix (1999b) connected Zipf’s Law and Gibrat’s Law, by noting that the random growth in city populations (or population shares) described by Gibrat would lead to the city size-rank relationship described by Zipf. While many authors note the occurrence of Zipf’s Law in very different geographies and time periods, some have emphasised deviations. Using the well-known Bairoch et al. (1988) dataset of European city populations over time, Dittmar (2020) shows the emergence of Zipf’s Law across Western and Eastern Europe at different points in the transition from the medieval period

² For a review of the literature that started with Gibrat’s work, see Sutton (1997).

to the modern period. Soo (2005) uses size-rank distributions at a country level to build a dataset of coefficients and concludes, based principally on municipal definitions of cities and OLS estimators, that the coefficient on the Pareto distribution differs from one in the majority of cases.

However, Soo's results for urban agglomerations – available for a much smaller number of countries – suggest the opposite result, highlighting the importance of city definition and methodology. The literature on Gibrat's Law is equally diverse. An important contribution to this literature is Glaeser et al. (2014), who explore the dynamics of county growth in parts of the USA 1790-2000. While they find evidence in favor of Gibrat's Law for the sample as a whole, this does not hold for long sub-periods, with less populous counties growing faster before 1860 and after 1970. This raises the possibility that occurrences of Gibrat's Law are an artefact of the accidental balancing of centripetal and centrifugal forces over different time periods.

We believe our paper makes five principal contributions. Firstly, ours is the first study in the literature to examine Zipf's and Gibrat's Laws in Britain over the long-run. Britain is of wider interest as it was home to the Industrial Revolution and, related, the world's urbanized economy. Secondly, we highlight the importance of research design in examining whether Zipf's Law holds, with spurious results common where arbitrary city definitions or cut-offs are used, in addition to bias and false precision. Using the same setting, we show the Pareto exponent can vary dramatically across 48 specifications in total: three different definitions of city available in our setting, four different sample cut-offs, and with four sets of results for each of these twelve unit-cutoff pairings – the standard (biased) OLS estimator and the unbiased estimator, each with unadjusted and adjusted standard errors (SEs). Thirdly, we find that we are unable to reject Zipf's Law in any of the 19 Census years under the strongest set-up – using primary urban areas, with a conservative cutoff, the unbiased estimator, and adjusted SEs. This is due not only to the point estimate but also to the necessary lack of precision in an urban system with approximately 60 true cities. Fourthly, our analysis of whether Gibrat's Law holds finds evidence against it holding in the period 1861-1911, where there is a strong negative relationship between initial size and fifty-year growth. In later periods, there is only modest evidence for such a link. Finally, across all three sets of analysis, we find evidence of rising inequality in city populations in the early 19th century, followed by greater compression across cities in the following century and a half. This is seen both in the point estimate of the Pareto exponent, which falls until 1861 then rises, and in the Gini coefficient of urban population, following Henderson and Wang (2007). Similarly, deviations from Gibrat's Law point to smaller cities enjoying faster growth after 1861 (and in particular before 1911 and after

1951).

Our paper is structured as follows. In Section 2.2, we review the underlying theory, both of power laws more generally and relating to exponents and cutoffs for samples, more specifically, and briefly review the existing empirical literature on the city-size distribution, paying particular attention to the city definitions, sample cutoffs and estimation methods used. Thereafter, in Section 2.3, we introduce our data on British city populations since 1801, and in Section 2.4, we outline the results of our empirical analysis, before the paper concludes.

2.2 Theory & Evidence

2.2.1 Power Laws

Power laws are ubiquitous in nature. Newman (2005) lists a number of physical, biological, and man-made phenomena where power laws are observed, including distributions of size frequency and volume. Examples are found in the size of earthquakes, moon craters, solar flares, computer files, and wars; the frequency of use of words in any human language and of personal names in most cultures; and the volume of papers scientists write, of citations received by papers, of visits to web pages, of sales of books and almost all branded commodities.

Mathematically, a power law function is a relationship between two quantities, where a relative change in one quantity results in a proportional relative change in the other. More precisely, given two quantities x and y , one varies as the power of another, independent of their initial size:

$$y = f(x) = ax^{-k} \quad (2.1)$$

The fundamental characteristic of this kind of relationship is scale invariance. This implies that scaling the argument of a power law function causes a proportionate scaling of the function itself:

$$f(cx) = a(cx)^{-k} \quad \text{for } k > 0 \quad (2.2)$$

This is important because it implies that we can observe a linear relation in log-log, which is the most straightforward way to test for the existence of power law behaviour.

Much of the interest in power laws comes from the study of the probability distribution and its application to probability theory and statistics. Strictly, a power law function cannot be a probability distribution as, for any value of the exponent, its integral diverges either in zero or infinity. Consequently, it is necessary to define its support greater than a lower cutoff x_{min} (or smaller than an upper cutoff), and multiply it by a scaling parameter

C , so that its integral meets the necessary unity condition of a probability distribution.

A Pareto distribution is a particular type of power law, which is mostly used in social science. If X is a continuous random variable following a Pareto distribution, then its density distribution function is given by:

$$f_X(x) = \frac{\zeta x_{min}^\zeta}{x^{\zeta+1}} \mathbb{1}_{[x > x_{min}]}(x) \quad (2.3)$$

and its cumulative distribution is:

$$P(X < x) = \int_{x_{min}}^x \frac{\zeta x_{min}^\zeta}{x^{\zeta+1}} dx = 1 - \left(\frac{x_{min}}{x} \right)^\zeta \quad (2.4)$$

Where the Pareto exponent (ζ) takes a value, in absolute terms, of 1, this is called Zipf's law. This regularity is found in a variety of situations and it is so called because it was initially observed by Zipf (1965) in the distribution of words' length.

Our focus is city size. To understand how Zipf's Law can be applied in this context, consider the probability (P) that the size of a given city S is greater than a certain value s . Where G is the survival function of a city's size and r is the city's rank, this probability is proportional to its rank:

$$G(S) = P(S > s) \propto r(S) \quad (2.5)$$

Where $G(S)$ takes the form of a power law and θ refers to its exponent, then:

$$G(S) = \frac{k}{S^\theta} \quad (2.6)$$

Combining equation (2.5) and (2.6) together, it follows that:

$$G(S) = P(S > s) = \frac{k}{S^\theta} \propto r(S) \Rightarrow r(S) = \frac{k}{S^\theta} \quad (2.7)$$

In this way, we obtain a specific inverse relationship between rank and the size. Where $\beta = 1/\theta$, $G(R)$ is the distribution of rank and k is a normalising constant, this relationship can also be expressed in the opposite way, as follows:

$$G(R) = P(R > r) = \frac{k^{1/\theta}}{S^{1/\theta}} \propto S(r) \Rightarrow S(r) = \frac{k^{1/\theta}}{S^{1/\theta}} = \frac{k^\beta}{S^\beta} = \left(\frac{k}{S} \right)^\beta \quad (2.8)$$

This gives the exact same relation as in (2.7) but reversed, i.e. the probability that the rank R is greater than a given r is proportional to city's size. Zipf's Law is said to hold

if $\beta = -1$. An important distinction is needed between rejection of Zipf's Law, where the exponent is statistically different from -1 but the relation may still be log-linear, and rejection of a Pareto distribution, where the log-linear relationship is rejected altogether.³

Before Gabaix (1999b), the unit value of the distribution parameter had remained unexplained. Gabaix's random growth model offers a theoretical explanation for the emergence of Zipf's law, both for cities and for other phenomena. He shows that, if we assume one (proportional) growth process for cities above some minimum size, i.e. that cities grow independent of their size (Gibrat's Law), then the steady state distribution of such a process is a power law with exponent 1. The conditions for this to hold are that cities grow on average at the same pace and with the same variance and that the smallest city in the distribution is very small as a proportion of the total urban population. Most other models in more recent papers that involve the Zipf exponent, such as those by Córdoba (2008) and Dittmar (2020), ultimately rely on the same underlying principle to explain its emergence. An exception is Rossi-Hansberg and Wright (2007), who use a general equilibrium model where cities specialise in particular final goods. In their model, Zipf's Law emerges as cities reach efficient size given their specialization, but only where labour is perfectly mobile and not a factor of production.

2.2.2 Cutoffs and Exponents

As described above, two factors characterize a power law distribution: its exponent and the lower or upper cutoff – in the case of city size distributions, the lower cutoff is the minimum city size. Despite its importance and its relationship with the value of the exponent, there has been very little debate on the estimation of the lower cutoff (Eeckhout, 2004). The majority of the related literature uses a data-driven cutoff (for example: Dobkins et al., 2000; Dobkins and Ioannides, 2001; Ioannides and Overman, 2003; Davis and Weinstein, 2002; Eaton and Eckstein, 1997; Dittmar, 2020; Soo, 2005). An alternative approach, used by Eeckhout (2004), is to set the cutoff at an arbitrary level to encompass different quantiles of the population. Only Muller (2016), in examining air pollution, uses a statistically defined cutoff.

Given the importance of the cutoff, we set out below four different approaches to its definition. These four approaches are then used to inform the empirical analysis.

³ In his paper, for example, Dittmar (2020) fails to reject a Pareto distribution on the Bairoch dataset after 1500 but he does not test the value of the exponent. This implies, strictly speaking, that he is testing for the emergence of a Pareto distribution but not of Zipf's Law, following the glossary provided by Gabaix (2009).

1. **Level cutoff:** This is the most common method employed in the literature so far (see for example Ioannides and Overman, 2003; Dobkins et al., 2000; Dittmar, 2020). Particularly where the empirical setting is the USA, a value of 50,000 is most common, reflecting the fact that the US Census Bureau reports data on Metropolitan Areas of 50,000 inhabitants and above, from 1898. This level cutoff is then constant throughout the period of analysis. While easy to understand, it means that, where population levels change significantly over the period of analysis, the composition of cities included over time can change significantly, relative to the distribution as a whole.
2. **Fraction cutoff:** An alternative heuristic approach is to take a certain fraction of the distribution in each year; Cheshire (1999) mentions this as one of the options to set the cutoff. This has the advantage over a level cutoff of a relatively even sample size, especially over longer periods of analysis, during which city populations may have changed by an order of magnitude. Another feature of this approach is its focus on the top fraction of the distribution; the relationship between city size and rank may differ within this set of cities compared to the full distribution of cities.
3. **Conservative cutoff:** Clauset et al. (2009) describe a process for selecting the cutoff. Firstly, the power law exponent is estimated for each possible x_{min} in the dataset. Secondly, using a Kolmogorov-Smirnov test to calculate distance, the cutoff is chosen as the one in the exponent-cutoff pair which minimises the distance to a theoretical power law distribution with the same parameters. For each x_{min} , the test also returns the p -value for the null hypothesis of a Pareto distribution. Of the tests, it tends to keep a bigger portion of the sample and is, therefore, referred to as the conservative cutoff.
4. **Deviation cutoff:** The fourth cutoff uses the method outlined in Gabaix and Ibragimov (2011) and applied in Rozenfeld et al. (2011). Here OLS is used to estimate the relationship (in logs) between a city's rank and both its size and the square of size minus γ , where $\gamma = (cov(log^2(size), log(size)) / (2var(log(size))))$. More formally:

$$\log(Rank - 0.5) = \beta_0 - \beta_1 \log(size) + \beta_2 \log(size - \gamma)^2 \quad (2.9)$$

As explained in Rozenfeld et al. (2011), the test formalizes the intuition that high values of β_2 indicate deviations from a power law because, in the limit, a true power law will have $\beta_2 = 0$. They outline a critical value for β_2 , at the 1 percent confidence level. If the absolute value of β_2 is greater than this critical value, β_c , the null hypothesis of a power law is rejected. This allows the calculation of a sample size, based on repeatedly extending the sample until the coefficient β_2 is statistically significant.

For a given sample size, the main approaches for estimating the Pareto exponent and the Gibrat's coefficient are described below.

Zipf's Law Despite the problems pointed out by Gabaix and Ioannides (2004), the OLS estimator is the most common method used in the literature for the estimation of the Pareto exponent (see, among others, Rosen and Resnick, 1980; Eeckhout, 2004; Soo, 2005, and Table A.1, where we summarize 22 of the key papers in the literature). The issues arising with this estimator are its biasedness and its underestimation of the standard error. The source of the bias follows from Renyi's theorem (Gabaix and Ioannides, 2004), which proves that $P(S_{(1)}/S_{(2)} > x) = 1/x$ (for $x > 1$). It follows that, for an expected value of $S_{(1)}/S_{(2)}$ of 2, the computed 95 % confidence interval is $[1/0.975, 1/0.025]$ i.e. $[1.03, 40]$. So typically, the value of $S_{(1)}$ will be above the value predicted by the linear regression with slope -1. Intuitively, this is due to the fact that there are few bigger cities and several smaller cities – closer in size – and that observations are ranked. Because of the asymmetry of the confidence interval mentioned before, if the second largest city is smaller than what it should be according to rank-size rule, it can be much smaller – down to 1/20th of the first city instead of 1/2 – without violating Zipf's Law, while it is less likely that such a gap in size exists for smaller cities. In addition, if the second largest city is much smaller, then the third should also be at least as small, because of ranking. This generates larger residuals for cities with smaller ranks (see figure A.1 in Appendix) than for those with higher ranks. Since OLS is minimising residuals, the slope is “dragged down” by the larger residuals of bigger cities, thus biasing downward the final estimate. An easy correction to this problem is provided by Gabaix and Ibragimov (2011), who retrieve the mathematical form of the bias and prove that using the regression $\log(\text{Rank} - 1/2) = a - \log(\text{Size})$ it is removed. The underestimation of the standard error on the other hand derives from the fact that the ranking procedure creates positive correlations between the residuals – if $\log(\text{size})$ of a city is below its supposed $\log(\text{rank})$ level, all smaller cities will also be below their supposed $\log(\text{rank})$ otherwise they would come before in rank – whereas the OLS standard error assumes that the errors are independent. Once again first Gabaix and Ioannides (2004) and then Gabaix and Ibragimov (2011) provide an easy solution to this problem. They show in fact that the true standard error is asymptotic to $\sqrt{2/n}$, where n is the sample size. In this paper, we use the OLS estimator with and without the bias and standard error correction proposed by Gabaix, with our preferred combination being the OLS estimator corrected for bias

and with corrected standard errors.⁴

Gibrat’s Law Concerning investigations of Gibrat’s Law, the literature does include examples of parametric estimations, such as linear spline estimation (see Desmet and Rappaport, 2017) and bin-size dummy regression (see Michaels et al., 2012). However, they are less frequently used than non-parametric estimation and less immediate in their interpretation. Therefore we favor more common non-parametric approaches, and in particular the Nadaraya-Watson estimation. Leading papers in this literature use the Nadaraya-Watson estimator, which provides a visually straightforward interpretation of the results (see, for example, Ioannides and Overman, 2003; Desmet and Rappaport, 2017).

2.2.3 Literature

It is far beyond the scope of this paper to try to summarize the literature in its entirety. Instead, we focus on both seminal and recent contributions to this literature, paying particular attention to tests of Zipf’s and Gibrat’s Laws. We report in this section a brief summary of the more extensive literature review given in Appendix A.1.

We distinguish between short and long-run analysis and, among short-run analyses, between locations. We define the “short-run” literature to be empirical analyses using less than 50 years of data on urban populations and review a range papers that meet these criteria, distinguishing where possible between three sets of analyses, reflecting the choice of urban unit. The first using officially-defined municipalities to delineate cities, the second using metropolitan areas, while the third using definitions of cities built from satellite imagery. The second part of our literature review covers empirical analyses of city growth and size distributions in the long-run, which we define to be over a period of at least half a century. We review these separating into four broad categories based on their (principal) region of analysis: global studies; the Americas; Asia; and Europe.

At first glance there appears to be little consistency in the literature. The cases of India and China are emblematic: Schaffar and Dimou (2012) and Chauvin et al. (2017) both find that Zipf’s Law does not hold in either country, but they then disagree on whether Gibrat’s Law does. Dingel et al. (2019) are not able to reject Zipf’s Law in either economy when they use night-lights to define cities and note that they would have if they had used administratively defined cities. Jiang et al. (2015) make a similar finding and raise

⁴ It is possible to use the Hill estimator, which is the maximum likelihood estimator for a power law, available in closed form, however Gabaix and Ioannides (2004) show that it also delivers a biased standard error. Further alternatives are non-parametric approaches, which allow the Zipf exponent to vary by city size. The latter include the local Zipf exponent (Gabaix, 1999a; Ioannides and Overman, 2003) and the Theil estimator, as per Dittmar (2020).

the importance of scale: Zipf's Law appears to be a weak descriptor of small urban systems but more relevant for larger urban systems. This is intuitive: it seems unreasonable to expect the urban structure of Andorra, home to fewer than 80,000 residents of whom more than half live in the metropolitan area of the largest city, to match precisely the principal characteristics of the urban structures of the continental US or China, vastly larger geographically and economically.

As the examples of India and China suggest, the seeming babel of empirical results relating to Zipf's and Gibrat's Laws stems, at least in part, from differences in the underlying unit being measured, the cutoffs employed on those units, and the empirical methods chosen. As shown in the summary of the short-run literature in the appendix, where larger and less arbitrary urban boundaries are used, it becomes harder to reject Zipf's Law in particular. This is particularly the case in the newer literature using satellite imagery, such as nightlights. Reflecting the literature, as well as economic rather than political realities, our focus is on metropolitan areas, not municipalities. For such units, there is by and large clarity that a power-law relationship exists between rank and size, at least when the appropriate cut-off is employed.

In terms of method for testing Zipf's Law, as outlined above, the state-of-the-art has progressed substantially beyond basic OLS and now includes corrections to both the point estimate and the standard error, as well as non-parametric methods. The latter correction involves use of the sample size itself. When combined with the use of larger urban units, which are by definition fewer in number, one consequence is far less certainty about the parameter estimates. This creates a tension in longer-run analyses between wider confidence intervals, which can mean an inability to reject a parameter estimate of 1 over time, and discerning trends in the point estimate; this is a point we return to in our analysis. For this reason, as a robustness check, we analyse also the trends in the Gini coefficient, as suggested by Henderson and Wang (2007).

The long-run literature typically focuses more on Gibrat's Law than Zipf's. It includes a number of analyses of US urban growth over the long-run, with some evidence in favor of Gibrat's Law but with numerous caveats. Analyses of other urban systems, including Brazil, China and Japan, find support for Zipf's Law but again Gibrat's Law appears to hold only in certain cases. In research that overlaps in setting somewhat with ours, Klein and Leunig (2015) examine urban growth in England 1761-1891 and test Gibrat's Law. They find that it is violated consistently, although this appears to be driven by their choice of unit and lack of any cut-off. Overall, the long-run pattern of urban growth dynamics and the size-rank relationship in Britain, the world's first urbanized industrial economy, remains largely unknown and this is the focus of our analysis.

2.3 Data

2.3.1 Administrative Context

To understand the relationship between a city's size and its rank and future growth, we use detailed data on British city populations from 1801 to 2011. The underlying data stem from the UK's 22 decennial censuses. There are at least two key attributes of the British setting that are worth noting. Firstly, because of its island nature, the geographical scope of the larger political unit is fixed.⁵ Secondly, Britain was the location of the first Industrial Revolution and, consequently, the first economy to transition to majority-urban. This gives Britain the longest-running panel series for an urbanized economy.

The oldest administrative units in Britain are civil parishes, which in many instances have ancient roots, dating back to feudal times. Civil and religious parishes overlapped until the Poor Law Amendment Act (1866), which defined "civil parishes" to be any unit that levied its own rate, for the purposes of poor relief; this definition included not only ecclesiastical parishes but also other units, including townships. The *Local Government Act* (1888) created larger administrative counties (and county boroughs) as units for local government, which often resembled older traditional (or ceremonial) counties. Under a successor Act in 1894, administrative counties were subdivided, into units known as urban and rural districts. The *Local Government Act* of 1972, together with the *London Government Act* of 1963 and the *Local Government (Scotland) Act* 1973, reformed the make-up of districts in Britain. In England, for example, it split 314 districts into metropolitan (34) and non-metropolitan (244) categories, as well as London boroughs (32) and two other *sui generis* districts.

Under the *Local Government Act* of 1992, unitary authorities were formed in England and Wales, with council areas formed in Scotland, after the 1994 *Local Government etc. (Scotland) Act*. These are local authorities responsible for the provision of all local government services within a (local government) district. This means that, since 1992, a new spatial taxonomy has existed that reflects urbanization within districts: under the 1992 Act, larger towns can have separate local authorities from the less urbanised parts of the same districts. This Act continues to be used, with changes in six ceremonial counties between 2019 and 2021. In Dorset, for example, the number of unitary authorities was reduced from eight to two in 2019. This reduction involved the merger of two existing (urban) unitary authorities and one non-metropolitan district into a new authority (Bournemouth, Christchurch and Poole), while the five remaining non-metropolitan dis-

⁵ Relatively few other countries have fixed geographical boundaries over the same time period. The United Kingdom included the full island of Ireland from 1801 to 1921 and Northern Ireland since. We focus here, however, on Britain, which has been in the same political unit since 1707.

districts were merged to form Dorset Council (UK Houses of Parliament, 2018).

The example of Dorset above highlights the challenges of using municipal boundaries to examine the city size-rank relationship. Adjustments to the units used to officially define cities are not incremental but occur instead in less frequent but more substantive reforms, such as those of 1972, 1992 and (in the case of Dorset) 2019. In order to reliably estimate the size-rank relationship, it is important to use spatial units that are not only consistent over time but also reflect the true extent of urban agglomerations. In the case of the Bournemouth, Christchurch and Poole unitary authority established in 2019, the Christchurch area had previously been a non-metropolitan district. Its inclusion was legally defined in the relevant Statutory Instrument by reference to five “electoral divisions”.

Electoral divisions (or wards) are the spatial units used for Britain’s administrative geography: all higher administrative units are built up of whole electoral wards or divisions. They are used for parliamentary constituencies and the EU’s Nomenclature of Territorial Units for Statistics (NUTS) regions, as well as for the unitary authorities (in England and Wales; council areas in Scotland) and the metropolitan and non-metropolitan districts mentioned above. As of the late 2010s, the UK had over 9,000 electoral divisions/wards, with an average population of roughly 5,500 in each. These form the basis of the *Vision of Britain through Time* database (VOB) we use for our analysis (Southall, 2017). The VOB database brings together historical surveys of Britain, in particular Census Reports, to make information on population by geographical unit publicly available. The database involved conversion of named areas into areas with consistent boundaries over time. To do this, VOB combines accurate information from the 2001 Census with County Administrative Diagrams, published from 1900 and extended back in time using Registrar General maps.

2.3.2 Spatial Taxonomies

As noted in the literature review, the choice of urban unit may affect the results of tests of Zipf’s and Gibrat’s Law. For that reason, we use three different geographical units in our analysis, all available from the VOB database and each of which is related to the concept of a city: “Local Government District”, “Unitary Authority” and “Primary Urban Area”.⁶ Summary statistics for all three levels of unit are given in Table A.2.

1. **Local Government Districts (LGDs)** are the most granular unit we use in our analysis and also the unit that varies most over time with administrative changes.

⁶ A fourth unit type, “Travel To Work Area”, exists in the VOB database, but only for three Census years between 1991 and 2011.

Data are available on population by LGD for each British census from 1851 to 2011. The number of urban districts rises from 527 in 1851 (with a median population of 5,109) to a peak of 1,143 in 1921 (median population: 8,595). Thereafter the number falls, in particular after the reforms of 1972, when the number of units falls from 895 to 461. There were 347 LGDs in Britain in 2011.

2. **Unitary Authority (UAs)** are, as described above, a larger spatial unit than LGDs and were reported from 2001, the first Census after the reforms of the 1990s. Using population by electoral division across Censuses, the *Vision of Britain* database has calculated the population for each UA by Census year from 1801 to 2011. This means that, unlike LGDs, UAs are fixed spatial units across the entire period of analysis. There are 379 or 380 UAs in most Census years.⁷
3. **Primary Urban Areas (PUAs)** are a spatial unit defined by the Centre for Cities to reflect the “built-up” area of a city (Centre for Cities, 2021). To do this, they aggregate UAs (for England; for Scotland and Wales, in almost all cases, the corresponding local authority area is used). There are 61 PUAs available for almost every Census year from 1801 to 2011. Given their spatial consistency over time, and given their ability to reflect the full extent of an urban agglomeration, these are our preferred unit.

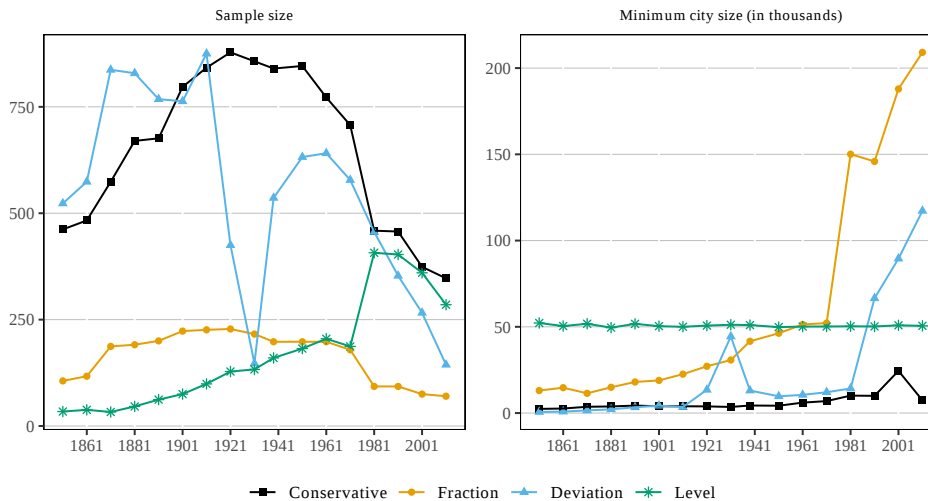
2.3.3 Samples by Cut-off

Section 2.2.2 outlined four possible cutoffs that can be applied to datasets of city populations: what we term the *Level*, *Fraction*, *Conservative* and *Deviation* Cutoffs. Further, Section 2.3.2 outlined three different spatial definitions of city. This gives twelve different combinations of spatial units and cutoff methods, what we term *unit-method pairs*. Figures 2.1-2.3 present for each Census year for which they are available, the number of cities included in the analysis, and the minimum city size, by cutoff, for each of the three levels of spatial units.⁸ In each, the black line represents our preferred unit (PUAs).

⁷ Data for Scottish UAs are missing for 1891.

⁸ Tables A.3-A.5 present the same information in tabular form in the Appendix.

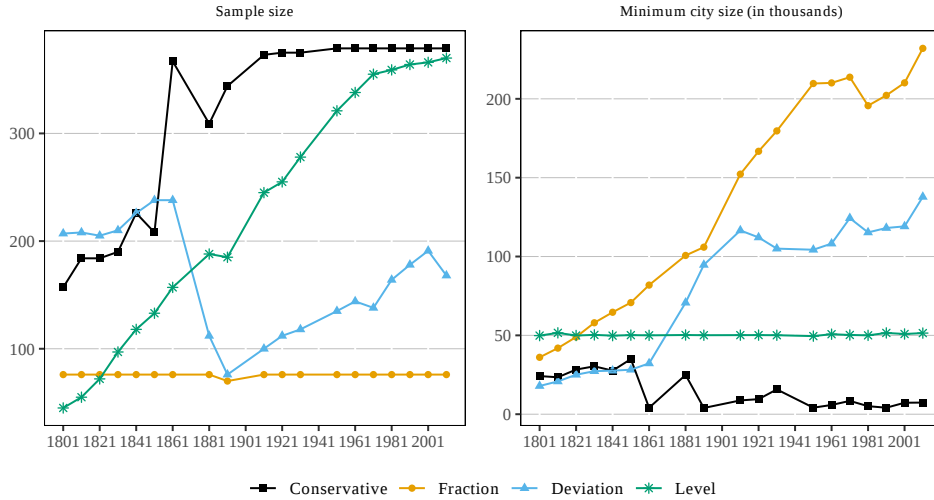
Figure 2.1: Local Government District



Note: The left panel of this figure shows the number of Local Government Districts included in the sample for each census year for each cutoff methods. The right panel shows the size of the smallest Local Government Districts included in the sample for each cutoff method.

Using LGDs (Figure 2.1, Local Governemnt Districts), the sample size to be included in the analysis is quite variable, both over time (given a cutoff method) and across methods. For example, of 347 LGDs in 2011, all would be included under the *Conservative* cutoff, 285 with a 50,000 *Level* cutoff, 144 under the *Deviation* cutoff, and 70 under the *Fraction* cutoff. Because of their nature, and a combination of the trend in the total number of LGDs and rising populations over time, the *Level* and *Fraction* cutoffs trend in different directions: the *Level* cutoff implies a sample of fewer than 100 before 1921, while the *Fraction* cutoff is at or above 200 for the period 1891-1961. While the minimum city size is 50,000 throughout under the *Level* cutoff, it rises from 13,000 in 1851 to 209,000 in 2011 for the *Fraction* cutoff. Under the *Conservative* cutoff, minimum city size is (roughly) between 3,000 and 4,000 until 1931 and close to 7,000 thereafter. The *Deviation* cutoff has a 100-fold change in minimum city size: from 1,000 in 1851 to 117,000 in 2011.

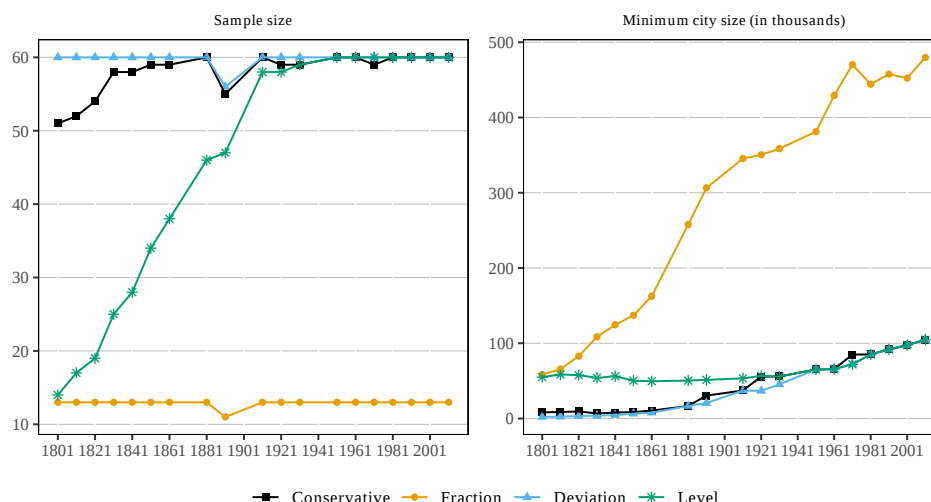
Figure 2.2: District/Unitary Authority



Note: The left panel of this figure shows the number of District/Unitary Authority included in the sample for each census year for each cutoff method. The right panel shows the size of the smallest District/Unitary Authority included in the sample for each cutoff method.

For UAs (Figure 2.2, Unitary Authorities), which in our dataset are constant over time with a full sample size of 380, again under the *Conservative* Cutoff, almost all are included from 1911 on; earlier than this only a smaller share is included (155 in 1801). The minimum city size is quite volatile, at 20,000 or more before 1891 (with one exception) and below 10,000 thereafter (again with one exception). The *Deviation* cutoff yields a sample size of at least 200 before 1881, closer to 100 for the following Censuses, and rising again to close to 200 by the turn of the millennium. Minimum city size increases from less than 30,000 before the 1860s to more than 100,000 from 1911 on. By construction (given the fixed number of spatial units), the *Fraction* cutoff gives a consistent sample size of 76 throughout. There is an order-of-magnitude rise in the minimum city size over the first century, from 18,000 in 1801 to 117,000 by 1911, after which the minimum size is largely stable. Lastly, the *Level* cutoff yields a growing sample over time, from 45 in 1801 to 370 in 2011.

Figure 2.3: Primary Urban Area



Note: The left panel of this figure shows the number of Primary Urban Area included in the sample for each census year for each cutoff methods. The right panel shows the size of the smallest Primary Urban Area included in the sample for each cutoff method..

Lastly, there are the 61 PUAs (Figure 2.3, Primary Urban Areas), our preferred spatial unit. The *Conservative* and *Deviation* cutoffs include almost all of these throughout the two centuries, while (by construction) the *Fraction* cutoff only includes 13 each year. The *Level* cutoff trends up over time, starting at 14 and rising to 58 (or higher) from 1911 on.

There are two features that are appealing *a priori* in judging the various cutoffs and spatial units, given the time and geographical setting. Firstly, the best combination of spatial units and cutoffs are likely to exhibit a consistent sample size over time, reflecting the lack of entirely new urban agglomerations in the setting under consideration. Secondly, the strongest unit-method pairs should show a steadily rising minimum city size, corresponding to Britain's growing urban population over the period.

The minimum size of included cities is typically greatest using the *Fraction* cutoff: only for LGDs, the smallest units, before 1961 and for UAs (before 1831) is this not the case – and in those instances, it is the *Level* cutoff that has the largest minimum city size. Given the less arbitrary nature of the *Conservative* and *Deviation* cutoffs, this suggests that a rule-of-thumb *Fraction* cutoff is likely to miss important parts of the city distribution – as seen, for example, in the right-hand panel of Figure 2.3. Secondly, sample sizes (and related minimum city sizes) are volatile, even for statistically more robust cutoffs, for spatial units based on administrative boundaries. This is most obvious in Figure 2.1, where the minimum city size, under the *Fraction* or *Deviation* cutoffs, increases substantially between 1971 and 1991 – in a way completely inconsistent with Britain's underlying

population dynamics.

Ultimately, of the twelve possible unit-method pairs, it is the Primary Urban Area spatial unit, combined with either the *Conservative* or *Deviation* Cutoff, that reflects the two *a priori* desired attributes: a largely stable number of cities to be included in the analysis, and a minimum city size that grows gradually over time, in line with Britain’s urban population. Indeed, for PUAs in Census years after 1981, the *Conservative* and *Deviation* cutoffs give the same sample of sixty urban areas, representing approximately 60% of Britain’s overall population in 2011. For simplicity, given the nature of its construction means that it is at least as inclusive as the *Deviation* cutoff, our preferred cut-off is the *Conservative* one.

2.4 Analysis & Results

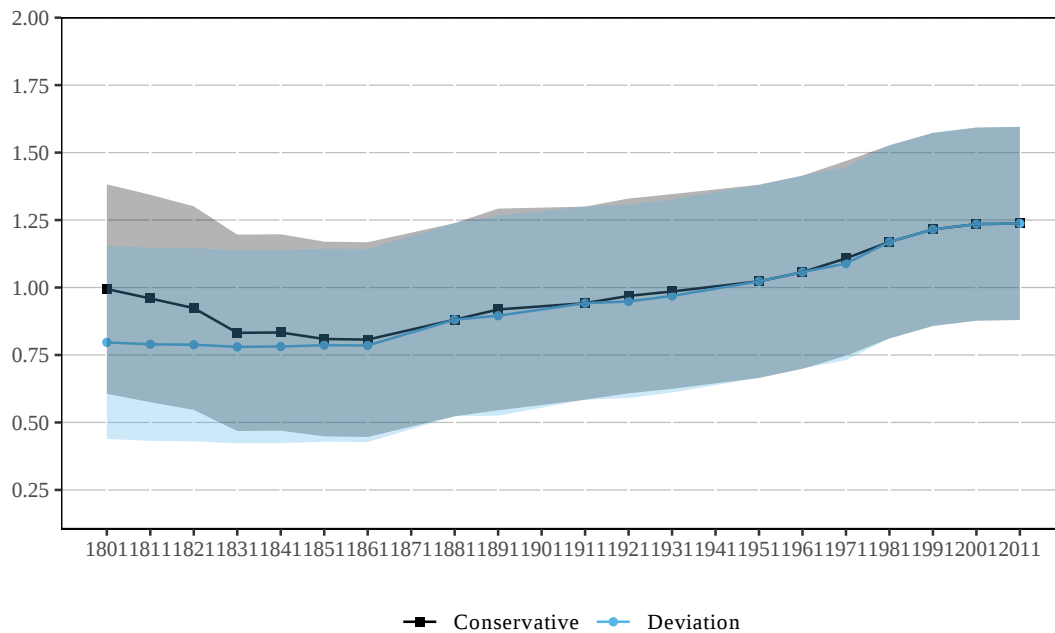
In this section, we outline the results of our analysis. We start by investigating the size of the Pareto exponent, in other words testing whether Zipf’s Law holds, and examine the extent to which the conclusion varies by city definition and sample cut-off. We provide then a robustness check, by presenting Lorenz curves and Gini coefficients for urban population. We turn finally to investigating whether Gibrat’s Law holds, i.e. whether there is any link between initial city size and future growth.

2.4.1 Pareto Exponent

Our first empirical objective is to examine whether Zipf’s Law holds, using the best combination of city definition and city-size cutoff, and how that answer changes when other unit-method pairs are used, including those pairs dominant in the existing literature on the presence of Zipf’s Law and the slope of the Pareto distribution of city size and rank. As explained in Section 2.2.2, in addition to the choice of spatial unit and cutoff, we also examine the impact on the estimated parameter and statistical significance of corrections for bias and the standard error.

Our strategy involves estimating, for all twelve unit-method pairs, the parameters of the city size-rank relationship for each year for which that unit-cutoff pairing is available (up to a maximum of 19 Census years). To start, we convert the size-rank relationship into logs. As per Gabaix and Ibragimov (2011), correcting for bias in the OLS estimator involves using the following adjusted log-log specification: $\log(Rank - 0.5) = \beta_0 + \beta_1 \log(Size)$. We report the absolute value of β , with a greater value corresponding to a steeper downward slope, i.e. a smaller largest city and thus a less concentrated spread of the population. Similarly, as explained in Gabaix and Ioannides (2004) and Gabaix (2009), the standard error estimated by OLS is underestimated by a factor of 5, because

Figure 2.4: Pareto exponent, Deviation and Conservative cutoff for Primary Urban Areas



Note: This figure shows, using two of the four cutoff methods (Conservative and Deviation cutoff), the absolute value of the Pareto coefficient, for each Census year for the preferred unit (Primary Urban Area).

the ranking procedure makes the residuals positively autocorrelated. Thus, we compare the unadjusted standard error with one where the standard error is given by the following:

$$SE = \sqrt{\frac{2}{n}}.$$

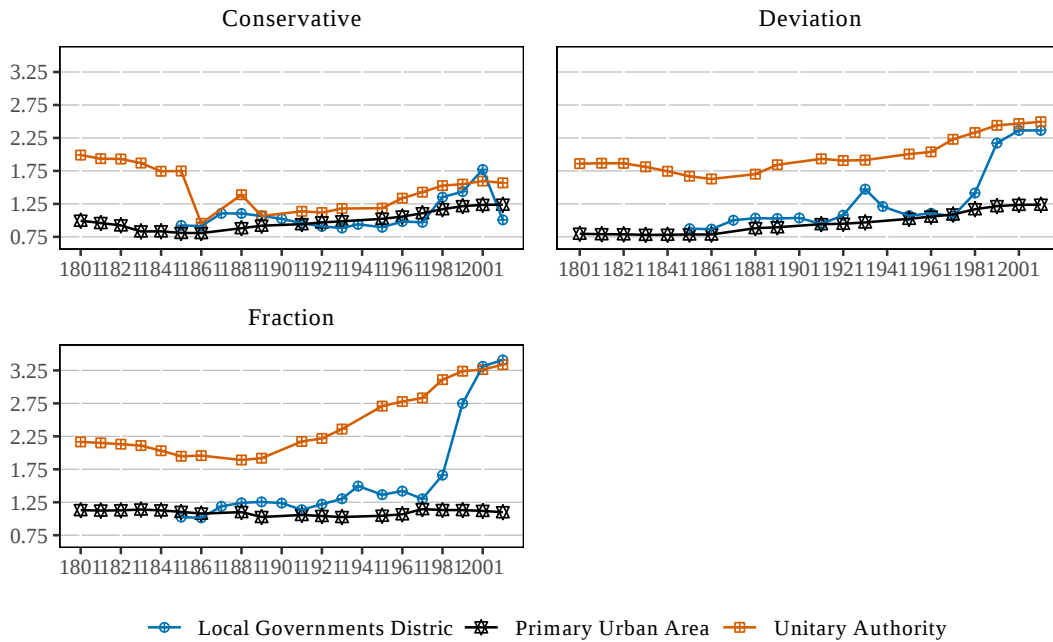
In each of 19 Census years, therefore, there are up to twelve unit-cutoff pairings and four sets of results per pairing: the standard (biased) OLS estimator and the unbiased estimator, each with unadjusted and adjusted SEs. Of the 48 permutations for any given year, we present two – the unbiased estimator, with adjusted SEs, using Primary Urban Areas, with either the *Conservative* or the *Deviation* cutoff – as the most reliable and thus our preferred, based on the discussion in Sections 2.2 and 2.2.3, and contrast the results from alternative permutations against this. The Pareto exponent for these two over time is presented in Figure 2.4, with the appropriate 95% confidence intervals. The preferred combination of data and methods presents three important stylized facts:

1. Firstly, Zipf's Law cannot be rejected in any one of the 19 Census years where we can use the preferred unit (Primary Urban Areas), either of the preferred cutoffs, and the appropriate estimator and standard errors.
2. Secondly, for most of the period (unsurprisingly given the similar samples), the two cutoffs produce very similar results.
3. Lastly, while the estimates are sufficiently imprecise to not rule out an exponent of one each Census year, any trend is upward over time, certainly after 1861: i.e if there is a trend in Britain's city distribution, it is towards a less unequal distribution

over the two centuries covered. At its lowest, the exponent is close to 0.75 while its most recent value is closer to 1.25. For an urban system with a fifth-largest city of one million (roughly the population of the Glasgow metropolitan area in the 2010s), the fall in the exponent implies a significantly smaller largest city: from 15.6m to 5.2m.

We now explore how these findings are affected firstly by changes in city definition, secondly by changes in sample cut-offs, and finally by the omission of correcting for bias in the coefficient or its standard error. Figures 2.5 and 2.6 compare, on panels with standardized axes, the estimated rank-size coefficient across all three definitions of city, for each of the four cutoffs described earlier. In all panels, the coefficients include a correction for bias in the point estimate. For ease of exposition, confidence intervals are not shown but the full set of results is shown graphically in the Appendix, in figures A.2 to A.4.

Figure 2.5: Pareto exponent across spatial units, by cutoff



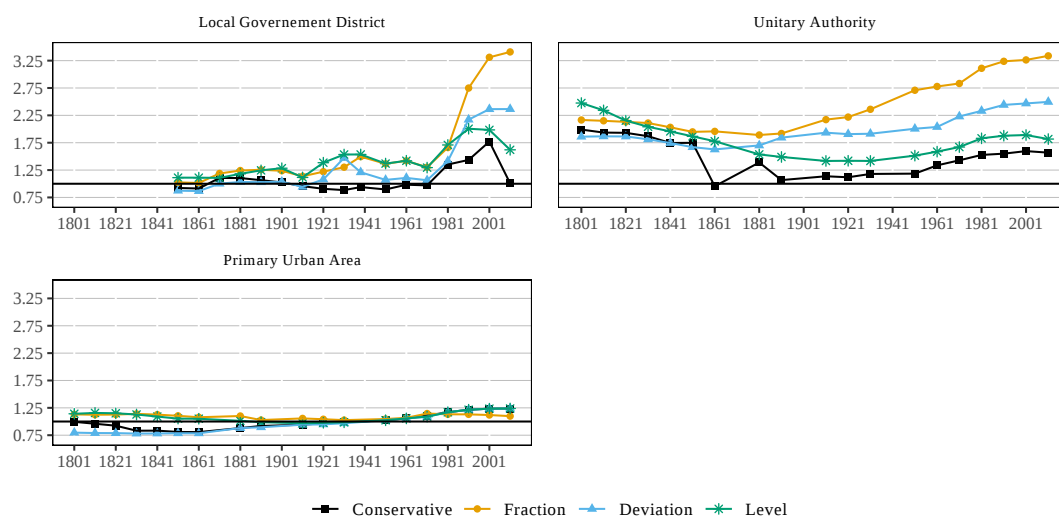
Note: This figure shows, for three definitions of cities, the absolute value of the Pareto coefficient, for each Census year for each cutoff method.

The top two panels of Figure 2.5 show our preferred cutoffs for city size and the black lines in each are the preferred spatial unit (Primary Urban Areas). As confirmed in 2.6, there are striking differences between the estimated Pareto exponent for this unit compared to Unitary Authorities and Local Government Districts, the type of administrative units subject to periodic but wholesale revision. Similarly, the use of more ad-hoc cutoffs (such as the top fraction of urban units or a population cutoff) produces very different

estimates, in particular for administrative urban units.

Three stylised facts emerge from this comparison. Firstly, across all cutoffs and in almost all periods, these smaller administrative units generate a larger Pareto coefficient. In 2001, using the *Fraction* cutoff, the estimated Pareto coefficient in the LGD and UA datasets is roughly three times that from Primary Urban Areas. While a more extreme example, there are similar results from, for example, the *Deviation* and *Level* cutoffs for the most recent Census years.

Figure 2.6: Pareto exponent across cutoff, by spatial units



Note: The four panels in this figure show the comparison of the absolute value of the Pareto coefficient between all cutoff methods, for each Census year and for each unit.

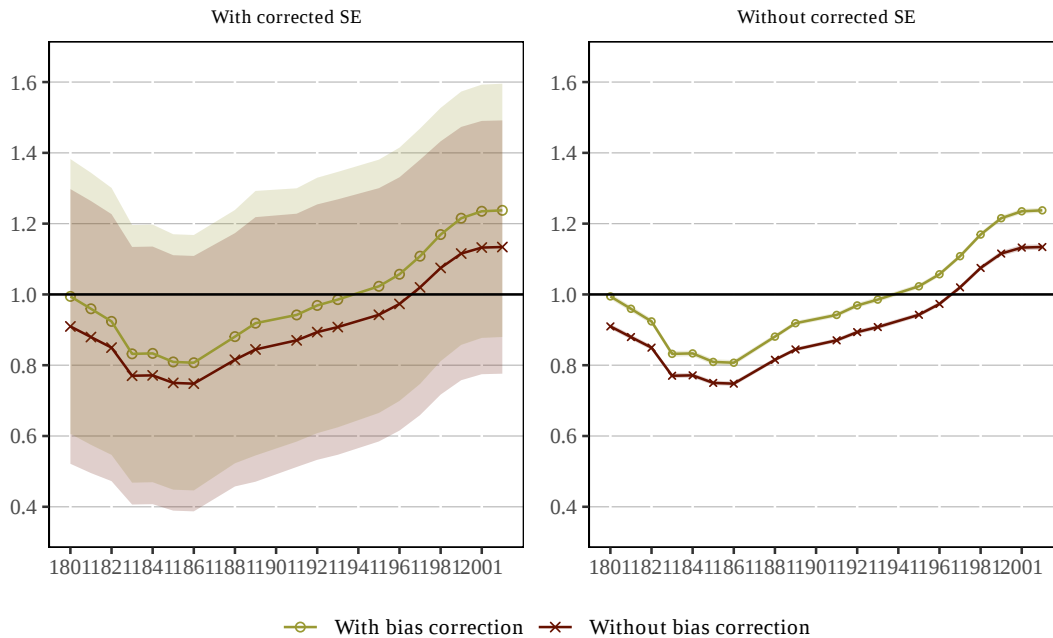
Secondly, using administrative units rather than urban extent, the pattern in the coefficient over time is less consistent within cutoffs. Figure 2.5 shows, in line with expectations given cities' largely slow-moving populations, a general trend towards a higher coefficient and less concentrated spread of population across cities when Primary Urban Areas are used. Using Unitary Authorities, however, the trend is negative (at least until the mid-20th century) using the *Level* cutoff and more erratic using the *Conservative* cutoff. Using LGDs, the trends across the different cutoffs are significantly more erratic, for example, the nearly three-fold increase in the coefficient 1971-2001, using the *Fraction* cutoff.

A related third stylised fact relates to consistency across cutoffs. As mentioned in the paragraph above, where administrative units such as LGDs and UAs are used, the choice of cutoff leads to very different conclusions about the extent of concentration in urban population and its change over time. Focusing just on Local Government Districts in 2011, the *Conservative* cutoff would imply a Pareto coefficient of almost exactly one, the

Level cutoff almost 1.5, the *Deviation* cutoff 2.25 and the *Fraction* cutoff roughly 3.25.

The final element of our analysis of the Pareto exponent concerns estimation methods. Figure 2.7 presents for each Census year the estimated coefficient, for the preferred unit-cutoff pairing – Primary Urban Areas and the Conservative Cutoff – with and without bias correction and, in both cases, using adjusted and naive standard errors. The left-hand panel shows the results using the appropriate standard errors, while the right-hand panel shows naive standard errors. In both panels, series with and without correcting for the bias in estimating the exponent are shown.

Figure 2.7: Pareto exponent, with and without correction for bias and standard errors



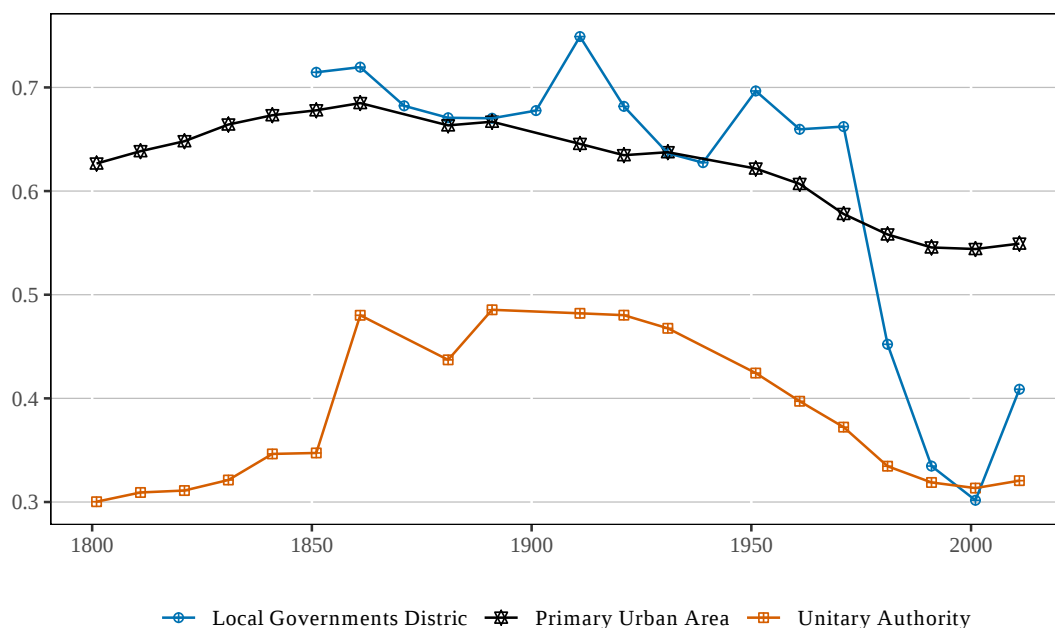
Note: This figure shows, using the preferred unit-method pair (PUAs and Conservative cutoff), the absolute value of the Pareto coefficient, for each Census year. The left-hand panel shows the results using the appropriate standard errors, while the right-hand panel shows naive standard errors. In both panels, two series are presented: with and without correcting for the bias in estimating the coefficient.

Two main stylised facts emerge from this. Firstly, the estimated coefficients, when no correction for bias is applied, are larger in absolute value than with the correction applied. This is in line with its construction but it is important to bear in mind, relative to other findings presented in the existing literature: papers without any adjustment for bias in the estimator will understate the extent of urban concentration. Secondly, while naive standard errors imply very precise estimates of the Pareto exponent, the use of appropriate standard errors indicates far greater uncertainty about its value. Again, this follows directly from the construction of the appropriate standard errors but is relevant when considering, for example, the rejection of Zipf's Law (where the exponent equals one).

2.4.2 Robustness check

We now perform a robustness check using the Gini coefficient of urban population distribution, a measure used by Henderson and Wang (2007) in their analysis of global urban populations. As in its more familiar setting of income distribution, the Gini coefficient ranges between 0 and 1, with 0 representing perfect equality (i.e. all cities have the same population) and 1 representing perfect inequality (i.e. the entire urban population is in one city). As with the Zipf and Gibrat analysis, we compute the Gini coefficient, and associated Lorenz curve, for each of the twelve unit-cutoff pairings, for each Census year for which its available. We present here only the key findings, leaving other analyses for the appendix, in particular Figures A.5 and A.6 and Tables A.6 and A.7 for the Gini coefficient and, for Lorenz curves, Figures A.7 to A.13.

Figure 2.8: Gini coefficient across spatial units, *Conservative* cutoff



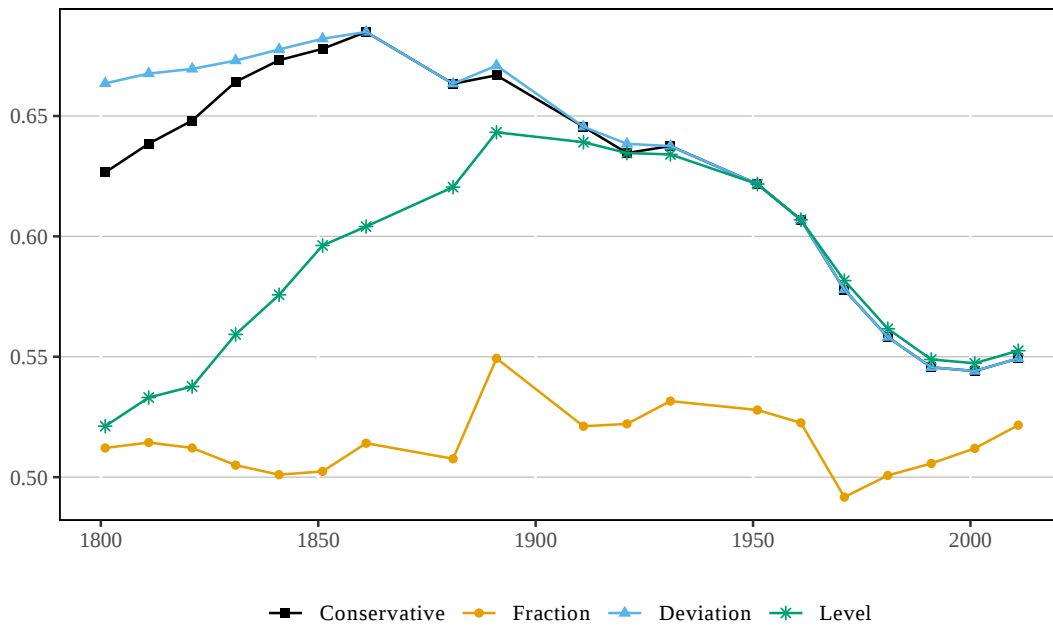
Note: This figure shows the evolution of the Gini coefficient for our preferred cutoff, i.e. the *Conservative* cutoff, for each unit.

An overview across units, using the *Conservative* cutoff is given in Figure 2.8, while an overview across cutoffs, using PUAs is shown in Figure 2.9. Our analysis presents four stylized facts:

1. Using the preferred combination of unit and cutoff reveals a steady rise and then decline in urban concentration in Britain over time. The Gini coefficient for PUAs

risks from 0.63 to 0.69 between 1801 and 1861, a rise of roughly one tenth. The trend then reverses, with that increase undone by 1911, before a pause or significantly reduced fall in concentration between 1911 and 1951. After 1951, the urban population continues to spread, with the Gini coefficient reaching 0.56 in 1991, meaning concentration had fallen by one fifth since 1911. There was little change in concentration between 1991 and 2011. These results point to very different trends in urban concentration in Britain than in those documented by Probst (2017) for Sweden.

Figure 2.9: Gini coefficient by cutoffs, Primary Urban Areas



Note: This figure shows the evolution of the Gini coefficient for our unit, i.e. Primary Urban Area, for each cutoff.

- Figure 2.8 reveals the importance of choice of urban unit is clear: concentration using Primary Urban Areas is nearly twice that observed in Unitary Authorities. Concentration among Local Government Districts is a similar level and trend to PUAs – but is significantly more volatile and jumps downwards sharply after 1971, reflecting municipality reform.
- Similarly, Figure 2.9 underscore the effect of choice of cutoff. Using a set fraction of the PUAs would suggest little meaningful change in urban concentration over time. A set population threshold, typically the most common choice of cut-off in the literature, would present an exaggerated increase in concentration during the 19th century, and with different timing (peaking in 1891).
- The patterns of urban concentration observed through the Gini coefficient are consistent with those seen in the Zipf's Law analysis: a rise in the Pareto exponent

(in absolute value) denotes a fall in urban concentration, similar to a decline in the Gini coefficient. The trends in the Gini coefficient and Pareto exponent point to the same evolution of urban concentration in Britain: rising concentration 1801-1861 and falling thereafter, especially 1861-1911 and 1951-1991.

2.4.3 Gibrat's Law

We turn, finally, to examining whether Gibrat's Law is observed in the data. Gibrat's Law holds that cities follow a growth process that is independent of their size. As per Section 2.2, Gabaix (1999a) proves mathematically that Gibrat's Law implies Zipf's Law and Córdoba (2008) proves that Zipf's Law implies Gibrat's, but as discussed in Section 2.2.3, while the validity of this law has been extensively examined, there are no clear conclusions. The most common approaches in the literature is a non-parametric kernel regression or Nadaraya-Watson estimation. A less well-known method, used by Desmet and Rappaport (2017), is a parametric piece-wise linear spline. We employ both methods here, referring the reader to Desmet and Fafchamps (2006) for more on the detail of these methods. We test Gibrat's Law on three city definitions – LGDs, UAs and PUAs – and for all four cutoffs, but focus our results here on the preferred combination of unit and cutoff: Primary Urban Areas with the Conservative cutoff. Results for other combinations are shown in online Appendix B.

Kernel regression The Kernel regression, or Nadaraya-Watson estimator, gives a continuous nonlinear approximation of growth relative to initial population. Where $L_{i,t}$ refers to the log of population for location i in year t , and with decennial data:

$$(L_{i,t+10} - L_{i,t})/10 = \phi_t L_{i,t} + e_{it}$$

An overview of the results of the Kernel regression, in particular the value of the Nadaraya-Watson estimator, is given in Figure 2.10, for the preferred combination of city-unit and cutoff. For the period as a whole, the coefficient is statistically significant from zero, at odds with Gibrat's Law. In particular, taking start (1801) and end (2011) populations, cities with a population of 20,000 or less (roughly, below log-value 10) grew faster than cities with a larger population in 1801.

The lower four panels in Figure 2.10 show that this is driven, largely, by the later 19th century. For 1801-1861, there is no link between initial size and subsequent growth; arguably the same is true for 1911-1951. However, especially for 1861-1911 – and also for 1951-2011 – there appears to be an inverse link between initial city size and subsequent growth: the smallest cities grew faster on average. This echoes Glaeser et al. (2014), who find that, before 1860 and after 1970, less populous counties grew faster in eastern and

central USA, although the timing is slightly different. The contrast between the 1861-1911 panel and the other three periods, in the scale of the coefficient, suggests a unique set of factors at work then, mostly likely rail infrastructure.

Linear spline The piece-wise linear spline involves mapping (log) population into its vector form, such that the coefficient on each spline segment measures the marginal effect of an increase in population size on growth. If growth is orthogonal, as per Gibrat’s Law, the coefficients of each of the spline segments should be close to zero. It is computed according to the following equation, where the 1-by- k vector $L_{i,t}$ includes a constant and a spline of population with $k - 1$ segments:

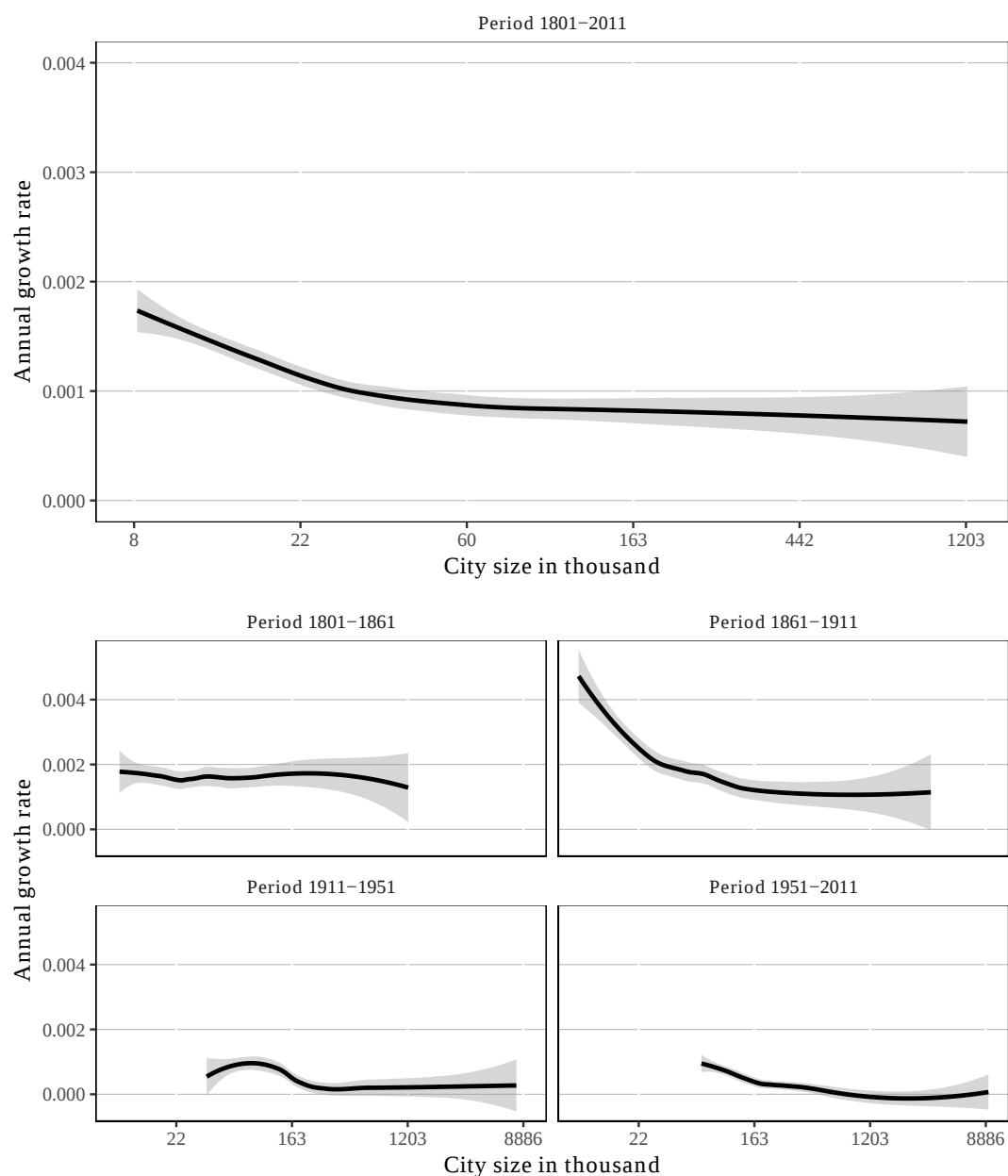
$$(L_{i,t+10} - L_{i,t})/10 = \vec{\beta} \cdot \vec{L}_{i,t} + e_{it}$$

An estimation of the piece-wise linear spline for every unit-cutoff-period combination results in more than 200 regressions, with 819 coefficients. Overall, for 83% (678) of the coefficients, it is not possible to rule out the null hypothesis of no link between initial size and growth, suggesting some support for Gibrat’s Law. Rather than report all these coefficients, we instead present a table summarizing the key results for our preferred unit-cutoff pair (see tables A.8 to A.10 in Appendix).

Tables A.8 to A.10 summarize the results, for our preferred unit-cutoff pairing. For the nineteenth century, the only significant coefficients are on the smallest size bin for the period 1851-1881. Thereafter, there are a greater number of significant coefficients but the pattern is less clear: smaller cities grow more slowly 1931-1961 and again 1981-2001, while mid-tier cities (with populations of between 0.2m and 1.2m) grow more slowly 1971-2001. Combined, these imply faster growth of the largest cities 1981-2001. These provide more detail in relation to patterns of urban growth, especially in the later 20th century, but are largely consistent with the narrative emerging from the analysis presented earlier in this section.

Taking stock of our analysis of the relationship between initial city size and subsequent growth, we highlight two aspects. Firstly, as with tests of Zipf’s Law, conclusions regarding Gibrat’s Law depend pivotally on the choice of urban unit. In particular, use of Unitary Authorities, rather than Primary Urban Areas, would strongly imply a negative relationship between initial city size and subsequent growth (see Figure A.14 – for the whole period 1801-2011 and for three of the four major sub-periods (1861-1911, 1911-1951 and 1951-2011)). Similarly, using LGDs (and the same Conservative cutoff) would lead to a rejection of Gibrat’s Law for the period 1961-1981, and in the same direction: smaller LGDs grew faster. Similarly, we contrast our findings with those of Klein and Leunig (2015), who reject Gibrat’s Law in every setting, looking at England and Wales

Figure 2.10: Kernel regression: Primary Urban Area with Conservative cutoff



Note: Kernel regression: growth rate for the whole period 1801 to 2011 plotted against initial city size in 1801, and intermediate periods 1801–1861, 1861–1911, 1911–1951 and 1951–2011 plotted against initial size.

during the 19th century; this rejection is driven by parishes with very small populations (under 2,000), well below the threshold for cities suggested by the data.

Secondly, we highlight the consistency in the pattern of results between this analysis and those presented earlier. As with the Zipf and Gini analysis, our Gibrat analysis points to the mid-19th century acting as a turning point in the evolution of Britain's urban system. Before this, smaller cities grew no faster than larger cities, but afterwards – especially 1861-1911 – there is strong evidence of compression in the urban structure.

2.5 Conclusion

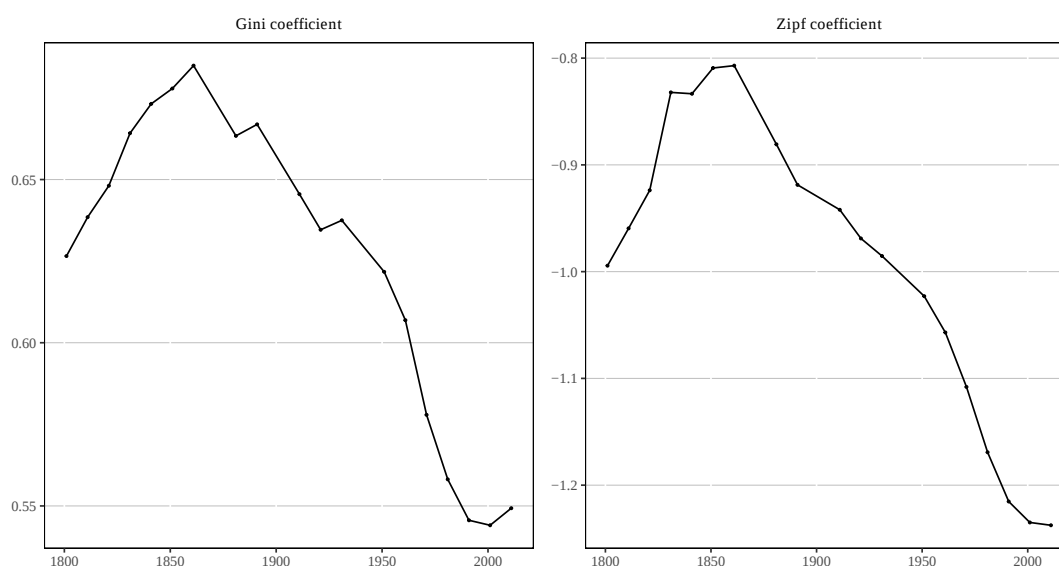
In this paper, we have examined concentration in Britain's urban system over more than two centuries. In particular, we analysed the relationship between the city size-rank relationship (Zipf's Law) and the link between a city size and its subsequent growth (Gibrat's Law), as well the Gini coefficient as a summary measure of urban concentration. Our work builds on a theoretical literature that has established that naive OLS estimation will produce both a biased estimator and artificially precise standard errors. We also place our work within an extensive empirical literature that examines the city size-rank and size-growth relationships. Using detailed Census data for Britain from 1801, we show how the estimated relationship between a city's rank and size varies dramatically across city definitions and sample cutoffs and that rejection of Zipf's Law depends not only on this choice but also on methods employed. We outline 64 possible combinations but choose Primary Urban Areas, and a Conservative sample cutoff, as our preferred sample, as well as using unbiased estimators and adjusted standard errors. The resulting sample is consistent with *a priori* expectations, given the setting: a largely stable sample size, reflecting the lack of new cities emerging during this period, and a steadily rising minimum city size, reflecting growing urban populations.

Under our preferred results, reflecting the strongest combination of city definition, sample cutoff, estimator and standard errors, we are unable to reject Zipf's Law, i.e. that the Pareto exponent is one, in any of the 19 Census years available. This reflects the limits to precision as much as the point estimates themselves: the true sample of cities in Britain over the last two centuries is approximately 60. One implication is that in small urban systems, even less precision will be available to researchers. The large standard errors relate, given their formula for construction, to underlying sample size. This suggests that, if the focus is precision of the estimate of the exponent, studies of the distribution should focus on larger economic units. Regardless, it is unclear that we would expect Zipf's Law to hold in economies with very small populations, such as Andorra (2020 population: 77,000) or Iceland (364,000), even if we expect it to hold in larger geographical units, such as Germany (83 million) or Europe as a whole (750m).

This raises broader questions about the relevant underlying economic structures: for example, with Britain part of the European Union 1973-2020, is that union the relevant unit within which to understand the distribution of British city sizes in those decades? Similarly, London's very large size in the 19th century may stem from strong links to other parts of the British Empire.

This suggests that it may be helpful for researchers to move the debate on from specifically rejecting (or not) Zipf's Law (whether $\beta = 1$ exactly) and instead understanding patterns of urban concentration (whether β is rising or falling over time). We summarize patterns of urban concentration – across both the Pareto exponent and the Gini coefficient – in Figure 2.11. Both series show the same pattern over time: a trend towards more concentration in bigger cities between 1801 and 1861 and then the opposite trend thereafter, especially during 1861-1911 and 1951-1991. Both Zipf and Gini analyses suggest a pause in this fall in urban concentration between 1991-2011, the timing of which coincides with the concept of the consumer city, i.e. one based on centripetal forces relating to consumption, rather than production or employment (Glaeser et al., 2001).

Figure 2.11: Gini coefficient and Zipf coefficient for Primary Urban Area with Conservative cutoff



Note: Gini coefficient and Pareto exponent for Primary Urban Area with Conservative cutoff. Pareto exponent is shown in regular, not absolute, values for consistency across the two panels.

The overall change in Pareto exponent is substantial and, where these results have wider relevance, it has significant implications where policymakers wish to understand the likely patterns of urban concentration over coming decades. For example, in the British urban system, the fifth largest city in the earliest 21st century (Glasgow) has a population of close to one million. Where the Pareto exponent is 0.8 (its peak in the mid-19th century),

this implies the largest city would have a population of over 15 million. Where the exponent is 1.25, similar to the value seen in 2011, the population of the largest city is less than six million. While the drivers of these changes in urban concentration – including transport technology and policies relating to housing and industrial strategy – are beyond the scope of this paper, the implications are substantial as policymakers seek to accommodate population growth and movements in the 21st century.

In addition, our findings have significant implications for researchers looking to understand the patterns and dynamics of city growth. Across the literature, the modal city definition is legal or administrative, rather than functional, and the principal cutoff used is a fixed population cutoff. If this were used in the case of Britain, it would give, as per Figure 2.6, an exponent of close to 2 in the early, implying a largest city of just 3 million where the fifth largest is one million.

More generally, across all cutoffs and in almost all periods, the use of smaller administrative units typically generates an artificially large Pareto coefficient, something compounded by the use of arbitrary sample cutoffs, such as a fixed population threshold. In some of the existing empirical literature, this effect may be offset, in part, by use of a naive OLS estimator, which suffers from an attenuation bias. In the case of Britain, the effect is close to 0.1 throughout the two centuries of data. Thus, much of the existing literature may accidentally benefit from two countervailing errors: the lack of adjusting for bias and the use of administrative units, rather than functional cities. In addition, due to the nature of revisions to administrative boundaries over time, these smaller administrative units imply an unrealistic degree of change in urban concentration over time. These issues are not present with the most appropriate unit-method pairs and highlight the need for researchers to understand the spatial units they are analyzing. Similar to Berry and Okulicz-Kozaryn (2012), we conclude that much of conflict in the literature so far is a consequence of choice of units of observation.

In summary, a growing urban population globally means an understanding of city size and growth patterns is of increasing importance. The experience of Britain over two centuries, the first economy in the world to urbanize, is one of initially rising but then falling urban concentration. In addition, the analysis presented here underscores the importance of meticulous use of data and methods in understanding urban dynamics.

Chapter 3

Skill-biased transportation change: Population, gentrification and Dublin's *Luas* light rail

Abstract

Mass transit is likely to play a key role in cities development, especially as the world continues to urbanize in the coming decades. At the same time, however, there is widespread consensus on that its effects on cities' spacial sorting are non-negligible. In this paper, I use a living-working-commuting model to examine the impact of the *Luas* light rail system introduced in Dublin, Ireland, in the 2000s. I link changes in market access for households and firms across 322 districts in the city to Census outcomes between 2002 and 2016. I find strong evidence, across both OLS and IV set-ups, that the new light rail system redistributed the city's population, with a 10% improvement in market access bringing about a 6% increase in a district's share of population and 4.5% increase in its share of employment. I also find that these impacts vary by skill level, with a district's share of the higher-skilled population rising by 7% (vs 5% for lower-skilled) for a 10% improvement in market access, but lower-skill employment share growing more than higher-skill (5% vs 3%). While improvements in a city's transport system are fundamental to its development and livability, it is also extremely important to account for its effects on spatial distribution.

3.1 Introduction

The majority of the world's population now lives in urban areas, with almost all future growth in the human population expected to take place in cities. Many different factors contribute to shaping cities and notably transport infrastructure investments play a determinant role. Access to reliable and efficient transportation systems in fact allows individuals to commute to work, access services and amenities, and reduces labor market frictions. At the same time, however, these investments can also have spillover effects on spatial distribution and thus on cities' gentrification process. This latter phenomenon, in fact, has garnered significant attention in the field of urban studies in recent decades – while gentrification can bring about improvements in the physical and social fabric of neighborhoods, it can cause ghettoization of certain socio-economics layers of the population. Understanding the drivers and impacts of gentrification is therefore important for policymakers and planners seeking to promote equitable and sustainable urban development.

In this paper, I examine the impact of Public Transport Infrastructure (PTI) investments at the extensive margin,¹ on city dynamics using the case of the *Luas* in Dublin, a light rail system introduced in 2004. The *Luas*, albeit “just” a tram, with its extension of over 40 km, drastically changed the quality of the Public Transit (PT) system in Dublin, which was, until then, only relying on quite a sparse bus network. I use an off-the-shelf work-live-commute model to estimate its effect on market access for households and firms across 322 districts, in cities rent level and gentrification in 2016. I do this by combining detailed spatial information on commuting, including district-to-district patterns, modes and times, with a range of Census outcomes, including a district's share of total city population, and that share broken down by skill level. I also examine the potential impact on residential rents. Within this set-up, PTI investment enters the utility function via a reduction in the disutility of commuting.

I find that improved transportation has an impact both on residential and working population in the areas involved. However the distribution across occupation types is not homogeneous: higher-skill workers tend to relocate to areas close to improved access, while lower-skill workers are more prone to work in those areas, a 10% improvement in market access bringing about a 6% increase in a district's share of population. I also find that this impact was skill-biased, with a district's share of the higher-skilled population rising by 7% – and its lower-skilled population falling by 5% – in response to a 10% improvement in market access. I find no clear effect on rents.

This analysis relates in particular to an emerging body of research that examines the urban economics of living, working and commuting using quantitative spatial equilibrium

¹ This is in contrast to PTI investments at the intensive margin where existing routes or networks are extended. I focus here on investments that are likely to significantly alter commuting patterns within the city.

models. A seminal paper in this regard is Ahlfeldt et al. (2015), who develop a model where agents make choices in respect of these three aspects of their urban lives and use it to estimate the impact of the Berlin Wall on urban agglomeration in the city. This model has since been adapted to answer other questions, including the impact of the *TransMilenio* light-rail system in Bogotá (Tsivanidis, 2019), of the BRT system in Dar es Salaam (Balboni et al., 2020) and the Metrobus network in Buenos Aires (Warnes, 2020), as well as gentrification (Severen, 2019; Heblich et al., 2020), endogenous amenities (Almagro and Dominguez-Iino, 2019) and the welfare-maximising allocation of space (Allen et al., 2016). Of note is that while Balboni et al. (2020) find bigger welfare effects for lower-income families, Tsivanidis (2019) and Warnes (2020) find the opposite.

A second literature to which I contribute concerns the impact of transport infrastructure investments on the spatial distribution of the city and on its economic activity. Earlier works on this topic have given considerable attention to the case of highways, especially in the United States and China, with a consensus on the positive effect of highway extensions on welfare, growth and employment (Allen and Arkolakis, 2019; Duranton and Turner, 2012; Garcia-López et al., 2013) and suburbanization (Baum-Snow, 2007; Baum-Snow et al., 2017; Garcia-López, 2012). Relative to highways, the conclusions about PTIs are more mixed, with milder effects on city growth/suburbanization (Gonzalez-Navarro and Turner, 2018) and links to incomes and density also (Gordon and Willson, 1984; Baum-Snow et al., 2005; Gomez-Ibanez, 1996). There is less clarity on the impact on housing values and on general welfare. While some papers, such as Gibbons and Machin (2005), Billings (2011), Tsivanidis (2019) and Bardaka et al. (2018), find a positive effect on housing values at least within a short radius from a subway extension, Daniele and Romito (2022) observe the opposite.

The third literature with which I connect relates to gentrification and its mechanisms. It is generally accepted that spatial patterns of household income are driven by transport technology (see, for example, Glaeser et al., 2008; Heblich et al., 2020). These investments allow poorer households to relocate to the outskirts (Severen, 2019), where rents are lower (Pathak et al., 2017; Su, 2022). One notable exception to this is Dragan et al. (2019), who finds no correlation between mobility rates and gentrification. But transport is far from the only cause of gentrification; others suggested include, for example, the rising demand for housing (Baum-Snow et al., 2018; Couture et al., 2019) and end of rent control (Autor et al., 2017).

I believe my contribution to these three strands of literature is as follows: firstly, to my knowledge, this is one of the first papers to examine the impact of new light-rail infrastructure on the spread of population in a European city. In particular, I focus on investments at the extensive margin, for which Dublin is a representative case study, but that can be extended to numerous new investment scenarios all over Europe. Secondly, I focus not

only on overall population effects but also on the effect by skill level, again a new feature for the quantitative spatial equilibrium literature concerning higher-income countries.

My results with respect to aggregate increase in population share are broadly in line with those found in Tsivanidis (2019), while those on the distribution of skill are consistent with those in Warnes (2020). The absence of an impact on rents, on the other hand, contrasts with Tsivanidis (2019), who finds a positive effect of increased market access on floorspace prices, Daniele and Romito (2022), who find a negative one, but is in line with Warnes (2020), who also finds no significant results in this respect.² By shedding light on this understudied aspect of gentrification, my work aims to contribute to the development of more equitable and sustainable urban policies.

The rest of this paper is structured as follows. In Section 3.2, I outline the spatial equilibrium model, which gives me equations for commuter market access, both for residents and for firms, and hypotheses on the impact of market access that I can test empirically. In Section 3.3, I outline in brief the evolution of Dublin’s public transport infrastructure and describe my data on outcomes of interest, including commercial and residential populations, commuter flows, and rents. In Section 3.4, I combine the model and data and test for the impact of the *Luas* on the utility of Dublin’s residents, before the paper concludes.

3.2 Theory & Hypotheses

My model draws on live-work-commute (LWC) models in urban economics, in particular starting with Ahlfeldt et al. (2015) and, for its focus on public transport infrastructure, Tsivanidis (2019). In these models, workers decide where to live and work and which mode to use to commute. Locations $i \in I$ differ in their commute times to every other location, their housing floorspace as well as their amenities and productivity level. Firms are located across the city and produce a *numeraire* good using labour and commercial floorspace as inputs. In equilibrium, three variables adjust to clear land and labor markets: the price of land (floorspace), by location; the price of labor (wages) by location; and the share allocated to each of the two uses. I describe below the worker’s maximization problem in detail, giving us measures of commuter market access for firms and residents, as well as that of the firm in brief. I then describe how this model’s theoretical predictions can be tested empirically.

² It has to be noted that both Warnes (2020) and Daniele and Romito (2022) use house values rather than rent level

3.2.1 Workers

A worker o , who lives in location i and works in location j , maximizes the following utility function:

$$\max_{c_{ijo}, l_{ijo}} C_{ijo} = \frac{B_i z_{ijo}}{d_{ijm}} \left(\frac{c_{ijo}}{\beta} \right)^\beta \left(\frac{l_{ijo}}{1-\beta} \right)^{1-\beta} \quad (3.1)$$

subject to $c_{ijo} + l_{ijo} Q_i = w_j$

where:

- c_{ijo} is the final good chosen as numeraire ($p_i = 1$)
- l_{ijo} is residential floorspace
- $d_{ijm} = \exp(\kappa t_{ijm} + v_{ijm})$ is the iceberg commuting costs, or disutility of commuting, where κ is the elasticity of commute cost to commute time, t_{ijm} is the time necessary to commute from i to j with mode m and v is the idiosyncratic preference for a specific mode.
- z_{ijo} is the idiosyncratic preference of each worker to live in i and work in j
- B_i is residential amenity
- Q_i is residential rent
- $1 - \beta$ is the share of residential floor space in consumer expenditure ($0 < \beta < 1$)

I assume workers decide, firstly, where to live and work and secondly which mode to use for commuting. Solving the maximisation problem by backward induction yields the following indirect utility function (dropping subscript o):

$$U_{ij} = \frac{B_i z_{ij} w_j}{\bar{d}_{ij} Q_i^{1-\beta}} \quad (3.2)$$

where the idiosyncratic shock z_{ij} is drawn from a Fréchet distribution, with scale parameter $T_i E_j$ (average utility from residing in i and average utility from working in j respectively, across all workers) and shape parameter ϵ (dispersion of average workplace and residential utility over worker across locations), and $\bar{d}_{ij}$ is the expected value of commute cost between i and j across modes.

Intuitively, all workers have their own distribution of preferences, from which they make a draw and consequently decide their live-work locations. This utility distribution accounts for the average utility from living in i and working in j , T_i , and E_j respectively, across all workers and ϵ , the dispersion of these utilities in space. A bigger $T_i E_j$ implies that high utility draws for any ij pair are more likely. The parameter ϵ , which I treat as common to all workers at any time period, reflects the variation within the distribution. A bigger ϵ

implies less variability.

Workers differ in their idiosyncratic preference for their mode of commuting. Assuming the disutility of commuting for worker o with each transit mode is distributed according to a Generalised Extreme Value distribution of Type I (GEV), or Gumbel distribution, the disutility of commuting of each worker across modes is distributed as a Gumbel distribution for minima allowing for correlation within nests. It follows that, when taking a decision on where to live and commute, a worker will base her choice on the expected disutility of commute, which intuitively is an average of the disutility of commute across all modes $s \in S$.

From McFadden et al. (1973), the probability a worker picks a specific combination of i and j is given by the following equation:

$$P(\max_{i,j} U_{ijo} = i\tilde{j}) = \pi_{ij} = \frac{T_i E_j(d_{ij} Q_i^{1-\beta})^{-\epsilon} (B_i w_j)^\epsilon}{\sum_{r=1}^S \sum_{s=1}^S T_r E_s(d_{rs} Q_r^{1-\beta})^{-\epsilon} (B_r w_s)^\epsilon} = \frac{\Phi_{ij}}{\Phi} \quad (3.3)$$

Summing over i and j gives the following probabilities for each location as a residence (R) and workplace (F):

$$\pi_{Ri} = \sum_{j=1}^S \pi_{ij} = \frac{\sum_{j=1}^S \Phi_{ij}}{\Phi} \quad (3.4)$$

$$\pi_{Fj} = \sum_{i=1}^S \pi_{ij} = \frac{\sum_{i=1}^S \Phi_{ij}}{\Phi} \quad (3.5)$$

Commuting market clearing implies that the total number of workers willing to commute to a location j (H_{Fj}) has to match the sum over all locations i of workers commuting from i (H_{Ri}) times the probability that from i a worker commutes to j :

$$H_{Fj} = \sum_{i=1}^S \pi_{ij|i} H_{Ri} = \sum_{i=1}^S \frac{E_j(w_j/d_{ij})^\epsilon}{\sum_{s=1}^S E_s(w_s/d_{is})^\epsilon} H_{Ri} \quad (3.6)$$

Residential land market clearing is given by:

$$(1 - \theta_i) L_i = \mathbb{E}[l_i] H_{Ri} = \frac{(1 - \beta)}{Q_i} \left[\sum_j \frac{E_j(w_j/d_{ij})^{1/\epsilon}}{\sum_j E_j(w_j/d_{ij})^\epsilon} w_j \right] H_{Ri} \quad (3.7)$$

where $\theta_i \in [0, 1]$ is the block-level share of residential floorspace and L_i is land in block i . This equation imposes equality between the share of residential land in each i and the expected number of units of residential floorspace (l_i) required by each worker in i times

the number of workers residing in that block.³

The expected utility of each worker is defined as the expected value of the maximum of the Fréchet-distributed random variable:

$$\mathbb{E}[u] = \mathbb{E}[\max_{ij} U_{ij}] = \gamma \left[\sum_{r=1}^S \sum_{s=1}^S T_r E_s (d_{rs} Q_r^{1-\beta})^{-\epsilon} (B_r w_s)^\epsilon \right]^{1/\epsilon} = \bar{U} \quad (3.8)$$

where γ is the Gamma function. Finally, keeping in mind that $\pi_{Ri} = H_{Ri}/H$ and rearranging the residential choice probability and expected utility equations, this gives:

$$\frac{B_i T_i^{1/\epsilon}}{\bar{U}/\gamma} = \left(\frac{H_{Ri}}{H} \right)^{1/\epsilon} \frac{Q_i^{1-\beta}}{W_{Ri}^{1/\epsilon}}$$

where W_i is a measure of residential commuter market access (RCMA) and it is defined as:

$$RCMA_i = W_{Ri} = \sum_{s=1}^S E_s (w_s/d_{is})^\epsilon \quad (3.9)$$

By the same reasoning and keeping in mind that $\pi_{Fj} = H_{Fj}/H$ I can also rearrange residential choice probability to obtain a measure for firm commuter market access (FCMA):

$$\frac{w_j E_j^{1/\epsilon}}{\bar{U}/\gamma} = \left(\frac{H_{Fj}}{H} \right)^{1/\epsilon} W_{Fj}^{-1/\epsilon}$$

where

$$FCMA_j = W_{Fj} = \sum_{s=1}^S T_s (B_s/Q_s^{1-\beta} d_{is})^\epsilon \quad (3.10)$$

These measures of RCMA and FCMA are fundamental in the model estimation as they form the explanatory variables I use to capture the introduction of the new mode of transport.

3.2.2 Firms

To simplify the model, in line with other LWC models, I assume that a single final *numeraire* good is produced under conditions of perfect competition, constant returns to scale, and zero trade costs within a larger economy:

$$X_{ij} = A_j H_{Mj}^\alpha L_{Mj}^{1-\alpha}, 0 < \alpha < 1 \quad (3.11)$$

³ Solving the maximisation problem yields $l_{ij} = \frac{w_j}{Q_i} (1 - \beta)$ and wage is defined from the indirect utility function as $w_j = \frac{U_{ij} d_{ij} Q_i^{1-\beta}}{B_i v_{ij}}$ which is random for the presence of the idiosyncratic shock. This is why it is necessary to use the expected value.

Profit maximisation and zero profits conditions yield:

$$q_i = (1 - \alpha) \left(\frac{A_i}{q_i} \right)^{\frac{\alpha}{1-\alpha}} A_j^{\frac{\alpha}{1-\alpha}} \quad (3.12)$$

As was the case with the residential market, I impose a commercial land market clearing condition where I equate the share of commercial floorspace available and the demand of commercial land obtained from the maximisation of the production function:

$$\theta_i L_i = \left(\frac{(1 - \alpha) A_i}{q_i} \right)^{1/\alpha} H_{Mi}. \quad (3.13)$$

3.2.3 Model Predictions

Tsivanidis (2019) shows that Commuter Market Access (CMA) can be used to evaluate the impact of a city transit network on any location. In his model, and in a subset of other urban models, endogenous outcomes, such as population, employment, and floorspace prices, can be written as a log-linear function of CMA, thus allowing for a log-linear reduced form representation. In this section, I use this representation to develop equations that can be tested empirically to evaluate the impact of the PTI on outcomes in Dublin, including rents, but also employment, and gentrification effects. I also discuss threats to the identification strategy.

In the theoretical model developed above, the transit network has an impact on equilibrium outcomes through the RCMA and the FCMA. As per Equations (3.9) and (3.10), they are defined as:

$$RCMA_i = \sum_{s=1}^S E_s (w_s / d_{is})^\epsilon$$

$$FCMA_j = \sum_{s=1}^S T_s (B_s / Q_s^{1-\beta} d_{is})^\epsilon$$

The RCMA of a location i is the sum of wage levels of all locations j reachable from i , discounted by the time cost it takes to get to each j , weighted by average workplace utility and the degree of its dispersion. Intuitively, this measure captures the trade-off between travel cost and wages accessible from a given location i faced by a worker residing in i . The higher the sum of available wage levels reachable from a location, the higher its RCMA; conversely, the further away from locations with good wage levels the lower the RCMA.

The FCMA of a location j is the sum of residential amenity levels of all location i reachable by j , discounted by the cost of commute between ij and the rent level, weighted by the average utility of residing in i and the degree of its dispersion. Intuitively, this

measure captures the trade-off between amenity and price-travel time faced by each worker of location j . The higher the FCMA of a location, the higher the average amenity level reachable by that location for a given rent level and commute cost, or the lower the average rent level and commute cost for a given amenity level.

An Empirical Set-up Tsivanidis (2019) shows that I can use the measures of RCMA and FCMA to estimate the impact of the introduction of new transport infrastructures on the outcomes of interest. This benchmark model has the following reduced form expression:

$$\Delta \log Y_{Ri} = \beta \Delta \log CMA + e_{Ri} \quad (3.14)$$

where $\Delta \log Y_{Ri} = [\Delta \log(H_{Ri}), \Delta \log(H_{Fj}), \Delta \log(Q_i),]$ are log changes in the share of residential populations and employment and changes in log floorspace prices respectively, $\Delta \log CMA$ is the change in either firm or residential commuter market access and e_{Ri} and e_{Fi} are structural residuals that reflect changes in productivity and amenities. To this, I add as outcome variables also the log of share of high- and low-skilled workers, which allows us to inspect the change in the composition of residents and workers as well.

This setup implies I need four principal sets of data to test the predictions of the theoretical model. These include the distribution of the residential population, across i ; the distribution of the workplace population, across j ; residential rents by i ; and the time costs of commuting from i to all other locations. While the original outcome of interest is the impact of commuter market access on rents, I can test for impacts on other outcomes, including the mix of residents or, over sufficiently long time frames, the mix of dwellings in an area.

Identification In the estimation of Equation (3.14), there are two major identification issues. The first is due to the fact that changes in CMA are also driven by changes in population and employment. Thus, since productivity and amenities are in the error term, they will be mechanically correlated with changes in CMA. To overcome this, I can use first-period population and employment data, by location, to instrument for CMA after the change in public transport investment. In my case, the relevant years are 2002 and 2016, which implies that I compute CMA for 2016 using population data for 2002. This allows me to isolate the changes in CMA due only to the introduction of the *Luas*.

A second concern is that the placement of the line is far from random. It is likely in fact that the placement of public transport investments is decided based on factors that may include unobserved neighbourhood characteristics, which may bias the OLS estimation and distort the results. A common approach in the literature is to instrument

the placement of new lanes with that of historical ones (Duranton and Turner, 2012; Warnes, 2020; Tsivanidis, 2019). The identification assumption is that the placement of the historical lanes is not correlated with contemporary changes in unobservable variables that in turn might correlate with the placement of the new lane and affect the outcomes of interest directly. As described below, for one of the two light-rail lines, this is appropriate; for the other, built on land that had been agricultural in the late 19th century, I used instead the inconsequential unit approach described by Redding and Turner (2015).

3.3 Data

To estimate the impact of the Luas on Dublin outcomes in a general equilibrium framework, I use the following datasets: residential population, by level of skill; workplace population, by level of skill; commuting time; and residential rents. I do this at the level of the smallest legally defined administrative area available for both periods, the Census district Electoral Division (ED). In Dublin County, there are 322 EDs, 162 of which are in the contiguous city area. Of EDs within Dublin city, the mean size in square kilometers is 0.72, while EDs in Dublin county but outside the city are on average larger in size (5.01 km²) but significantly less densely populated.

In this section, I describe the various data sources for each of the series required, starting with an overview of the *Luas* lines, whose introduction gives us the change in commuting costs. I then describe the data for both 2002 (before the introduction of the Luas) and 2016 (after), as data series differ across both years. Table 3.1 provides summary statistics for all sets of data in use.

3.3.1 The *Luas* lines

I start with a brief overview of Dublin’s *Luas* light-rail network, including the decision and construction process, as well as an overview of the data I use to estimate its impact on city mobility. Dublin is the biggest city in Ireland, with a metropolitan population of 1.4 million people and an area of 318 square kilometers. It is also the only European capital not to have an underground metro system in place and its bus system is among the slowest in Europe (see Inrix, 2021). While the city had a dense streetcar network in the late 19th and early 20th century, this was disbanded after World War II, with many routes converted into bus routes. Its two-line modern light rail system, the *Luas*, opened in 2004, in several steps and after a long decision process.

The idea for a light rail system for the city of Dublin was first suggested in 1981, in a Dublin Transportation Initiative report (available [here](#)). Following this report, the public

Table 3.1: Summary statistics

2002

	Mean	Median	Min	Max	SD	Sample size
Residential Pop	1109	906	15	9001	909	322
Residential Pop - HS	356	282	13	3174	330	322
Residential Pop - LS	755	624	66	6607	633	321
Workplace Pop	1109	453	0	16592	2050	319
Workplace Pop - HS	374	124	6	6809	842	307
Workplace Pop - LS	760	341	7	9783	1267	319
Commuting time	33	29	1	132	19	103684
Residential Rent	140	138	30	267	40	322

2016

	Mean	Median	Min	Max	SD	Sample size
Residential Pop	1243	915	178	11972	1202	322
Residential Pop - HS	456	330	39	3218	427	322
Residential Pop - LS	931	686	111	9573	912	322
Workplace Pop	1243	442	1	16822	2248	322
Workplace Pop - HS	457	108	1	10810	1123	321
Workplace Pop - LS	931	380	1	10842	1520	322
Commuting time	56	49	0	180	33	103684
Residential Rent	209	210	52	441	73	322

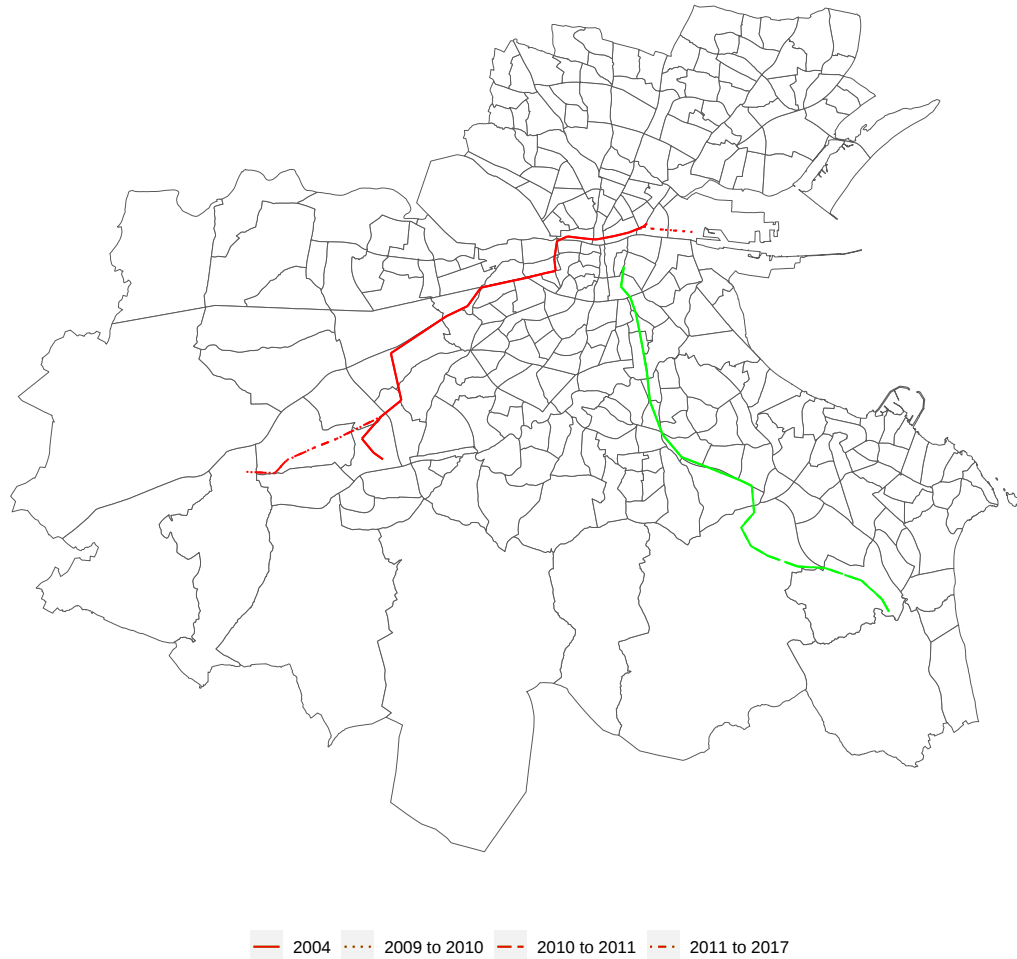
This table shows summary statistics of all data used in the calibration and analysis.

transport operator in Ireland, CIE, recommended implementing the roll-out in two phases (in the Dublin Light Rail Environmental Impact Statement available [here](#)):

- Phase 1: Line Tallaght to Dundrum/Balally via the City Centre (Red Line).
- Phase 2: Line Ballymun to the City Centre and Dundrum/Balally to Sandyford (Green Line).

In May 1997 it applied for permission to start construction; however, the inquiry was put on hold to investigate the possibility of underground sections in the city centre, which were soon abandoned in favour of an entirely over-ground system (Luas history, in Dublin Transportation Initiative Report, available [here](#)). In May 1998, the government finally decided to build two lines, nonetheless, it was not before March 2001 that construction work actually began. The original launch date was to be in 2003 (Annual Report 2003, by Department of Transport, available [here](#)), but there were delays in construction and the Green Line opened in June 2004 while the Red Line opened in September 2004. While a 2017 extension connected the two lines, by extending the Green Line into the north of the city, the timing of this is not relevant to the exercise undertaken here. Instead, there are two extensions that are relevant: in December 2009, the Red Line C1 Connolly to Docklands extension opened while in October 2010, the B1 extension from Sandyford to Cherrywood opened.

Figure 3.1: Luas Extensions



This figure shows the different extensions of the LUAS between 2004 and 2017.

3.3.2 Residential & Workplace Populations

2002 Data on residential population per ED are available from the 2002 census for the full sample; the workplace population in each ED is again available from the census data, from the POWSCAR dataset (Place Of Work School or College - Census of Anonymised Records), but was coded for a 15% sub-sample. To reconcile the two series, a “grossing factor” was used, indicating the percentage of the entire population each record accounts for. The sum of all grossing factors in Dublin county amounts to 478,255, which is in line with the total working population in that year. Additionally, in the model calibration,

we scale the residential population to match the workplace population, in order to satisfy the labour market clearing condition and impose the closed city condition by normalising the total population to 1. We also use the information available on socio-economic groupings to examine the effects on higher- and lower-skilled populations. We use Census Socio-Economic Group (SEG),⁴ available and coherent for both 2002 and 2016, and classify as "higher-skilled" those Census respondents who were coded as "Employers and managers" or "Higher professional", socioeconomic Groups A and B. We classify as "lower-skilled" those coded in all other groups, therefore "Lower professional" (C), "Non-manual" (D), "Manual skilled" (E), "Semi-skilled" (F), "Unskilled" (G).

2016 For residential population per ED, again I use Census data, in this instance from the April 2016 Census. Comprehensive data on workplace population per ED, and ij flows between places of residence and work, are available from the 2016 POWSCAR dataset, available online from the census website. For calibrating the κ parameter, I also used ij flows between places of residence and work at Small Area level, which is a smaller geographical unit, only available from the 2011 census onward. This data is not publicly available and I was granted permission to use it from the CSO for research purposes only.

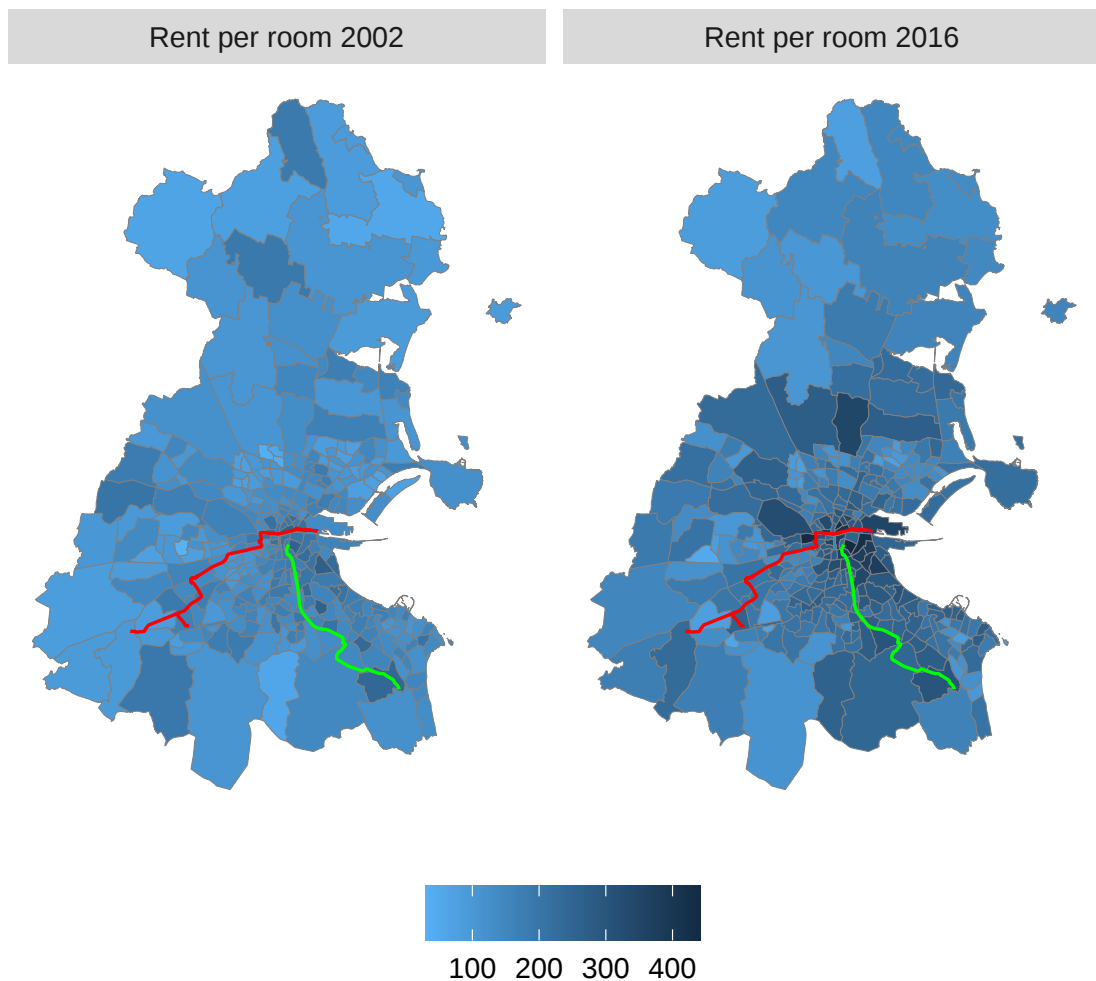
3.3.3 Residential Rents

2002 For residential rents in 2002, two data sources exist. Firstly, the online listings portal *Daft.ie* has listed market rents for 2002-2003; secondly, average residential rents, including from social housing, are also available from the 2002 census. As the former captures market transactions, it is likely a better reflection of the going market rate of accommodation, while the latter reflects rather longer-term renting. There is a total of 88,738 listings from *Daft.ie*, for which unfortunately I do not have the precise address but only centroid coordinates of *Daft.ie* area id. *Daft.ie* id areas are the *Daft.ie* equivalent of Electoral Divisions (EDs) in terms of average size, however unfortunately their boundaries do not correspond. The size of *Daft.ie* id areas is roughly constant over Dublin county, which implies that they are generally greater than EDs within Dublin city, while several *Daft.ie* id can be mapped into one ED in Dublin South and Dublin North, where the average ED size increases substantially. In order to obtain the average ED rent from *Daft.ie*, I have made a spatial interpolation of average rent over area id centroids and then re-aggregated back into EDs.

Average weekly rent by ED from the 2002 census is a more reliable source. Information

⁴ Socio-economic group classifies the population into one of ten categories based on the skill and educational attainment of their current or former occupation.

Figure 3.2: Average rent per room by ED, CSO



This figure shows the rent per room in 2002 and 2016 by electoral division. Source: CSO, census record 2002 and 2016. The LUAS line in 2002 is overlaid for ease of visualisation but was not there yet at the time.

on the number of rooms and house facilities is also available therefore it is possible to obtain rent per room and to create a composite index for house quality. Figure A.16, left panel, plots rents from Daft versus rents from CSO. As it is visible, the two sources are not very coherent with one another, therefore for the main specification I use CSO rents, as they are a more reliable source. More detailed information on the precise location of listings will be necessary to be able to use *Daft.ie* in future research.

2016 Similar sources are available for 2016, including listed residential rents from *Daft.ie*

and from the census. Like for 2002, I compare the two data sources. The average reported rent in Daft is higher than that reported in the census, even when restricting census records to rentals from private landlords only (and not from a non-profit). This is consistent with *Daft.ie* listings capturing market rents, with a faster turnover, while the census figures also include longer-term leases, whose rents may have diverged from market rents, due to a rent increase cap at 4% upon lease renewal. It follows that, at any given moment in time, on *Daft.ie* the proportion of short-term listings will be higher than that of long-term listings, because of their implicit higher frequency, and because of the rent increase cap, their price will also be higher. Census data, on the other hand, report a snapshot of the rent level of all rented houses, regardless of whether they are on the rental market or not at that moment, thus inducing a lower average. Nonetheless, plotting one against the other (the right-hand panel of figure A.16 in the appendix) reveals a positive relation, supported by a correlation coefficient of 0.83. Because of the lack of reliable *Daft.ie* in 2002 in the preferred specification I prefer to use the census data for 2016. A map for the visualisation of rent level in Dublin County is available in figure 3.2.

Housing attributes The 2002 Census also records additional housing attributes such as heating, sewerage system, access to public water, and dwelling age. This information is reported as the number of housing units with access to these facilities for each ED. For example, I have the number of dwellings attached to a public sewerage system, to an individual septic tank, and to other or no sewerage systems for each ED. Similarly, I have the number of dwellings connected to public, Local Authority, private or to no water supply, with and without central heating systems, and the number of housing units built between each decade (see table 3.2 for summary statistics). I define older buildings as those built before 1940, and I have classified an ED as a prevalently older one if the share of older buildings is above 50% in 2002, the mean being 20%. Similarly, if the share of building with access to public water or public sewerage system is below 95% (their mean being 99% and 95% respectively), I have considered them as an ED with limited access to water or sewerage respectively.

Table 3.2: Summary statistics of housing stock in 2002 across EDs

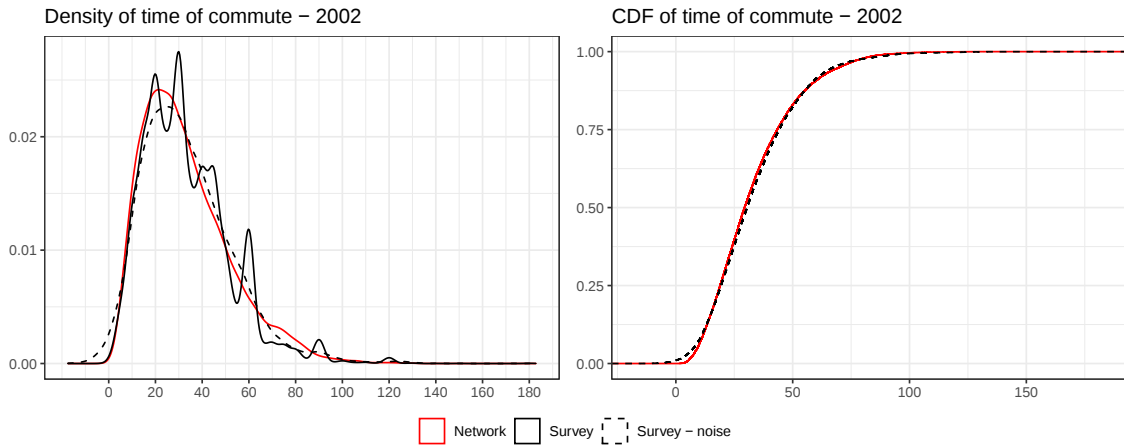
Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Built before 1940	0.198	0.231	0.000	0.020	0.326	0.932
No central heating	0.101	0.082	0.001	0.038	0.142	0.480
Public water supply	0.991	0.044	0.514	0.996	1.000	1.000
Public sewerage	0.950	0.152	0.000	0.983	0.995	1.000

Note: This table reports the housing attributes of the average ED. For example, the average ED has 19.8% of housing units built before 1940, 10% of housing units with no central heating, 99.1% with public water supply, and 95% with public sewerage.

3.3.4 Commute time

2002 For commuting time in 2002, two options are available. The first is to reconstruct the road and public transport network in 2002, use a Dijkstra algorithm to compute the shortest path between ij centroid pairs for each mode (public transport, car, and walk), and then make a weighted average of the time with each mode, where weights are computed from the survey. This data-intensive exercise was made possible by National Transport Authority (NTA), which provided a digitized map of Dublin’s road network and public transport in 2006, from which I have removed the *Luas*.

Figure 3.3: Commute time 2002



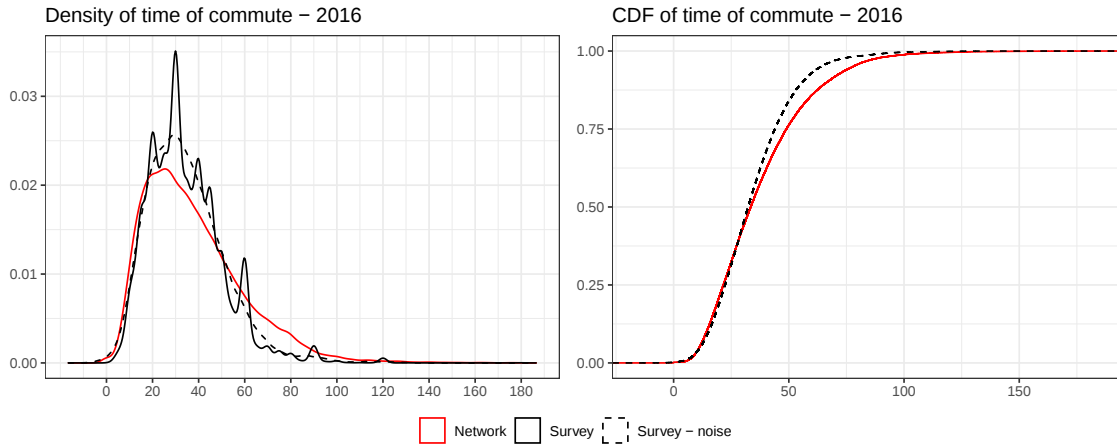
The left panel shows the probability density function of the commute time computed with the constructed network versus that declared in the 2002 mobility survey. Peaks at 10, 15, etc minutes are due to the fact that people tend to give a round number. For a better comparison, I have added Gaussian noise to the graph. The right panel shows the cumulative distribution function of the network versus the survey with noise.

A second option is to use the 2002 POWSCAR dataset, available from CSO, where people were asked their mode and time of commute. Figure 3.3 reports the density and the cumulative distribution function of both measures of commuting time, NTA, and mobility survey, plus a third-density plot based on a modified set of survey results where I add Gaussian noise, to adjust for the fact that respondents tend to round their commuting time. As it is visible, both the density function and the cumulative function are very close, meaning that the network I constructed works well to compute commute times. I, therefore, use the constructed network, based on 2006 NTA maps, in the main specification and leave the survey as a robustness check.

2016 Data on commuting time in 2016 are available from the same two sources: the digitized network and the 2016 mobility survey. For building the PT network in 2016, I use data provided by Kujala et al. (2018), where real-time data on time between stops is provided for 2017. For commutes by car and on foot, I use the Google Maps API.

Once again, I compute the weighted average by mode and find the shortest path between all pairs of centroids with the Dijkstra algorithm. The second option is to use the 2016 mobility survey. Figure 3.4 reports the density and the cumulative distribution function of the two and shows they are very close. As for 2002, I stick to my network for the main specification and use the survey for the robustness checks.

Figure 3.4: Commute time 2016



The left panel shows the probability density function of the commute time computed with the constructed network versus that declared in the 2016 mobility survey. Peaks at 10, 15 etc minutes are due to the fact that people tend to give a round number. For a better comparison, I have added Gaussian noise to the graph. The right panel shows the cumulative distribution function of the network versus the survey with noise.

3.3.5 Instrumental Variables

As described in Section 3.2.3, the placement of lines is not random and thus I include as part of the analysis an Instrument Variables (IV) approach.

- To instrument the *Luas* Green line running North-South an obvious choice is the old Dublin & South Eastern Railway line. The new *Luas* tracks were partly laid on the old rail tracks, which were not removed at the time of the end of its operation.
- For the Red line, which largely runs east-to-west, the majority of its line was agricultural land at the time of development of the rail network. To instrument this line, I use the straight line between the first and last stop, in line with the inconsequential unit approach and subsequently drop start and endpoints. This is a good approximation of the least-cost path between the city center and the terminus of the Red line, ensuring its relevance, while the straight line nature means its validity is plausible. For stop placement, I compute the average distance between stops on the Red line and we “break” the straight line at equally distant points, thus obtaining stops at roughly 1 km distance one from the other.

In both cases, I then calculate the change in CMA between before and after the introduction of the *Luas*, in the hypothetical scenario in which it followed the instrumented

routes. I then use the computed counterfactual change to predict the change in CMA in the current scenario, with the actual placement of the *Luas*.

3.4 Analysis

In this section, I start by outlining commuter market access, first for residents and then for firms. To do this, I need estimates of κ and ε , reflecting the elasticity of commuting flows to shares by mode.

3.4.1 Estimating RCMA

Recall that, where $\omega = E^{1/\varepsilon} w_j$, the level of wages adjusted for the shape parameter ε , RCMA is given by the following equation:

$$RCMA_i = \sum_{s=1}^S \omega_j / d_{is}^\varepsilon$$

This means that, in addition to district-level data on living and working populations, rents, and commuting costs, I also require data on two series in order to calculate RCMA: firstly, the disutility of commuting from i to j $d_{ij}^\varepsilon = e^{\kappa \varepsilon t_{ij}}$ (so the parameter $\nu = \kappa \varepsilon$); and secondly, adjusted wages at ED level, ω_j . I calibrate these, using the steps described below.

Disutility of commuting Where $d_{ij} = e^{\kappa \bar{t}_{ij}}$, and $\bar{t}_{ij}$ is the average commute time across modes, it follows that $d_{ij}^\varepsilon = e^{\varepsilon \kappa \bar{t}_{ij}}$. From the model, the probability of commuting between i and j is defined as:

$$\pi_{ij} = \frac{\Phi_{ij}}{\Phi} = \frac{d_{ij}^{-\varepsilon} \omega_j Q_i^{(\beta-1)\varepsilon} \tilde{B}_i}{\sum_i \sum_j d_{ij}^{-\varepsilon} \omega_j Q_i^{(\beta-1)\varepsilon} \tilde{B}_i}$$

Taking the natural log, and using the definition for d_{ij} , I obtain a gravity form of this equation where I can estimate $\varepsilon \kappa$ from a pairwise regression:

$$\log(\pi_{ij}) = \gamma_j + \delta_i - \varepsilon \kappa \tau_{ij} \quad (3.15)$$

where I have defined $\gamma_j = \log(\omega_j)$ and $\delta_i = \log(Q_i^{(\beta-1)\varepsilon} \tilde{B}_i)$, which are absorbed in workplace and residential FE respectively.

The number of all possible ED ij pairs is 103,684, which produces in most cases an estimated probability of commuting between pairs that is very small and an estimation that is noisy and unstable. For this reason, I aggregate to the level of the Dublin postal

district, of which there are 22. Aggregating at the postcode level shows a visible linear relation between log probabilities and time while also producing more stable results across samples. I run this regression for two model specifications on both datasets: log-log OLS and Poisson Maximum Likelihood on mobility survey data and estimated commute time from my network model. Results are reported in table 3.3 and figure A.17 in Appendix.

Table 3.3: Elasticity of commute flows to commute shares

	<i>OLS</i>		<i>Poisson</i>	
	Model	Survey	Model	Survey
$\kappa\epsilon$	−0.078*** (0.001)	−0.057*** (0.001)	−0.087*** (0.0002)	−0.065*** (0.0001)
Constant	0.493*** (0.114)	−0.433*** (0.136)	0.325*** (0.011)	−0.072*** (0.011)
Residential FE	Yes	Yes	Yes	Yes
Workplace FE	Yes	Yes	Yes	Yes
Observations	601	601	601	601
R ²	0.931	0.892		
Adjusted R ²	0.925	0.882		
Log Likelihood			−19,279.230	−27,584.040
Akaike Inf. Crit.			38,658.470	55,268.080
Residual Std. Error	0.360	0.453		
F Statistic	152.289***	92.406***		

This table shows the estimation gravity equation 3.15 to compute $\kappa\epsilon$, i.e. commute time to commute share. It is run using log-log OLS regression and Poisson Maximum Likelihood estimation on the probability of commuting obtained from the POWSCAR Mobility Survey and the time of commuting taken from the POWSCAR Mobility Survey or my network simulation.

Note:

*p<0.1; **p<0.05; ***p<0.01

The sign and significance of both sets of estimates match expectations. Further, the estimated values are similar across all four specifications, which gives support to the necessary assumption that my estimated time, based on network data, matches well the reported time in the survey. Overall I get an estimate of $\kappa\epsilon$ between 0.06 and 0.09, which is in line with that from Ahlfeldt et al. (2015). In the rest of the calibration, I assume $\kappa\epsilon$ to be 0.07.⁵

Adjusted wages Ahlfeldt et al. (2015) in their paper show that there exists a unique ad-

⁵ Note that I need the shape parameter ϵ to be greater than zero for the model to converge. As a consequence, since I get a value of κ which is also around 0.07, it is reasonable to continue the calibration considering ϵ to be slightly less than 0.07.

justed wage vector ω_j that solves Equation 3.6:

$$H_{Fj} = \sum_{i=1}^S \frac{\omega_j / d_{ij}^\varepsilon}{\sum_{s=1}^S \omega_s / d_{is}^\varepsilon} H_{Ri}$$

It is also clear from the same equation that there exists no closed-form solution for wages and thus it needs to be solved numerically. I do so using the same method as in Ahlfeldt et al. (2015) for three settings: 2002, 2016, and the instrument, where employment and residential data from 2002 are used with commuting costs from 2016. I adjust workplace employment H_{Fj} to match residential employment H_{Ri} , to comply with labour market clearing conditions, and rescale wages to have a geometric mean of 1. A visual representation of the results can be seen in figure A.18 in the Appendix.

3.4.2 Estimating FCMA

Analogous to estimating RCMA, I require two further sets of information in order to calculate FCMA. Where $\tilde{B}_s = T_s B_s^\varepsilon$ represents the adjusted residential amenity and Q_i is the price of residential floorspace:

$$FCMA_j = \sum_{s=1}^S \tilde{B}_s (Q_s^{\beta-1} / d_{is})^\varepsilon$$

Information required includes firstly the adjusted residential amenities of district i , $\tilde{B}_i = T_i B_i^\varepsilon$, and secondly the parameter ε , in order to be able to estimate $Q_i^{(1-\beta)\varepsilon}$. Below, I detail the estimation of both these parameters.

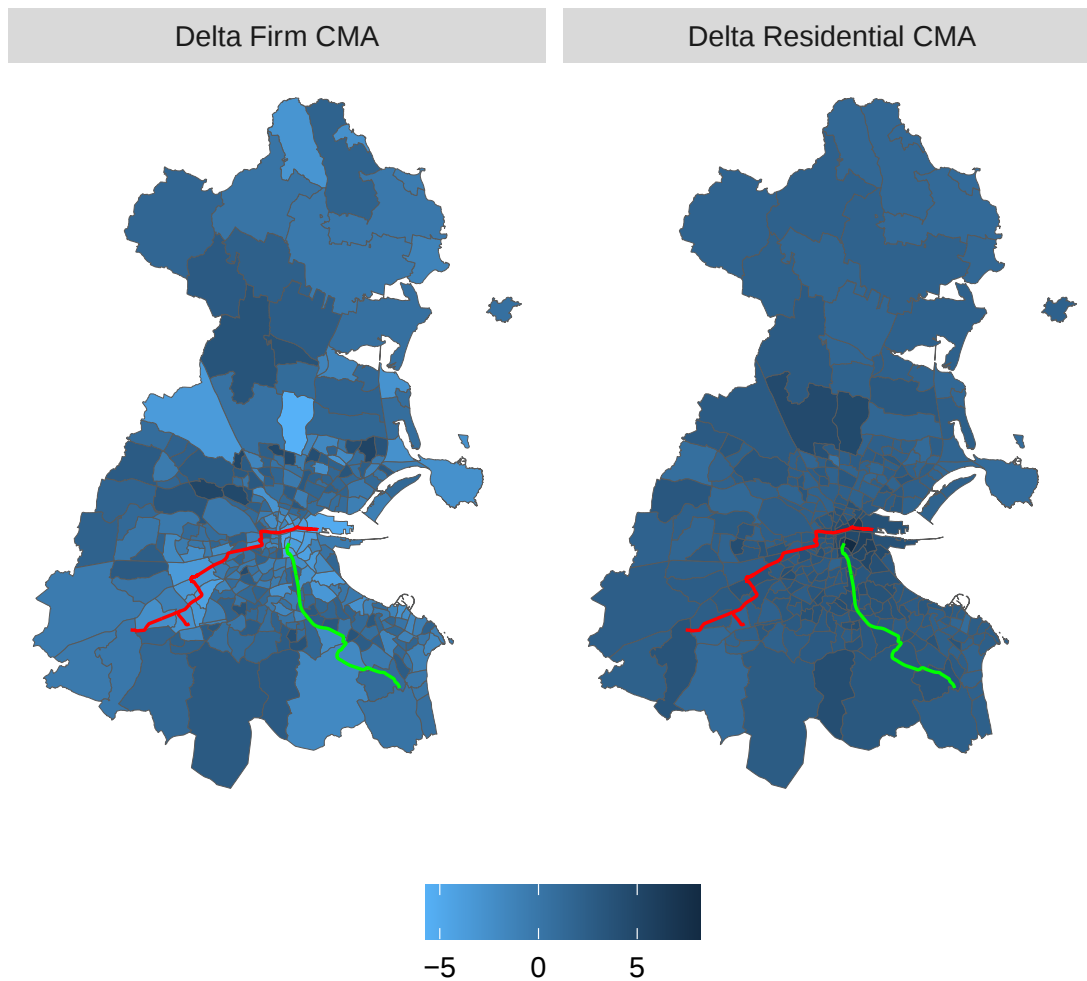
Adjusted residential amenities Once I have estimated transformed wages and disutility of commuting, I can once again use residential choice probability and expected utility equations, rearranged, divided by their geometric mean and in logs, yielding an expression for the residential amenities divided by its geometric mean:

$$\log \left(\frac{\tilde{B}}{\bar{\tilde{B}}} \right) = \frac{1}{\varepsilon} \log \left(\frac{H_{Ri}}{\bar{H}_{Ri}} \right) (1 - \beta) \log \left(\frac{Q_i}{\bar{Q}_i} \right) + \left(\frac{W_i}{\bar{W}_i} \right)$$

Since I have already calibrated the RCMA, I can easily obtain residential amenities, by plugging in each element in the equation.

Semi-elasticity of commute shares to commute cost The cost of commuting from i to j using mode m , d_{ijm} , is given by $d_{ijm} = e^{\kappa \tau_{ijm} + v_{ijm}}$ where τ_{ijm} is the time to commute from i to j with mode m , $m \in \{car, walk, PT\}$, κ is the semi-elasticity of commute time to commute cost and v_{ijm} is the idiosyncratic preference of each individual for a given mode. Following the literature on transportation, I treat this as a discrete choice problem

Figure 3.5: Commuting Market Access calibration



The left panel shows the change in calibrated RCMA per ED, while the right panel shows the change in calibrated FCMA. The darker the shade of blue the higher the CMA level in a given ED.

(McFadden et al., 1973). Having made a decision on where to live and work, an individual o has to decide which mode to use for commuting. As earlier, they maximise the following utility function:

$$U_{ijo} = \frac{B_i \bar{z}_{ijo} w_j}{d_{ijo} Q_i^{1-\beta}}$$

This means that they minimize the cost of commuting d_{ijm} over m , so the realisation of mode choices is the set of draws from the random variable $\min \{d_{ijm}\}$.

Table 3.4: Estimation of kappa

<i>Dependent variable:</i>	
Utility of Commuting	
b_{PT}	-1.801*** (0.649)
b_{car}	-7.369*** (1.977)
κ	-0.069*** (0.002)
λ_{Public}	0.442*** (0.025)
Sex	Yes
Age	Yes
Departure Time	Yes
$\lambda_{Private}$	Yes
Observations	8,003
R^2	0.202
Log Likelihood	-6,045.937
LR Test	3,060.402*** (df = 47)

This table shows the estimation of the multinomial logit model following McFadden et al. (1973). κ is the elasticity of commute cost to commute time and it is the parameter of interest for the calibration of the model.

Note:

*p<0.1; **p<0.05; ***p<0.01

Assuming that v_{ijm} has a GEV distribution and that individuals are independent, and allowing for a nested structure for public and private modes, I can write the probability that an individual commuting from i to j makes choice $\bar{m}$ as:

$$P(\min \{d_{ijm}\} = \bar{m} | ij) = \pi_{m|ij} = \pi_{k|ij} \times \pi_{m|ijk}$$

$$= \frac{\sum_{n \in B_k} \exp \left(b_n - \frac{\kappa}{\lambda_k} \tau_{ijn} \right)^{\lambda_k}}{\sum_{k'} \left(\sum_{n \in B_k} \exp \left(b_n - \frac{\kappa}{\lambda_{k'}} \tau_{ijn} \right) \right)^{\lambda_{k'}}} \times \frac{\exp \left(b_m - \frac{\kappa}{\lambda_k} \tau_{ijm} \right)}{\sum_{n \in B_k} \exp \left(b_n - \frac{\kappa}{\lambda_k} \tau_{ijn} \right)}$$

where b_n is the mean preference for mode n , λ_k is the degree of correlation within nests and κ is the semi-elasticity of commute cost to commute time.

I estimate this probability using a nested multinomial logit on a 10 percent subsample of the 2016 mobility survey, as the information on the respondent location is at the Small Area level, which is more precise than at the ED level, as provided for 2002. I code the respondent's residence area to its respective Small Area centroid and workplace area to its workplace zone centroid. I use my network to estimate the commute time with all three modes (car, public transit, and walking), and I take from the survey the chosen mode from each respondent. Results are reported in Table 3.4. I get an estimate for κ of -0.069. This is both statistically significant and, as expected, negative – as utility decreases as commute time increases. However, the value is substantially higher than what is found in Tsivanidis (2019), who finds a semi-elasticity of 0.012. Above, in Section 3.4.1, the value for $\kappa\epsilon$ was estimated to be 0.07. This gives a value for ϵ of 1.01, which is sufficient condition for the moment of degree one of the distribution to converge.

3.4.3 Results

In Tables 3.5-3.6, I report regression results where the regressor of interest is $\Delta \log(CMA)$, using both OLS and the IV set-up described in Section 3.2.3. Table 3.5 focuses on changes in residential commuter market access and the three outcomes of interest are the change in the share of the overall working population in Dublin residing in an ED; and the change in a district's share of the city's higher- and lower-skilled population (HS and LS, respectively). Table 3.6 presents equivalent results but where the treatment of interest is firm market access and the outcomes reflect populations of workers, rather than residents.

In the first two columns of Table 3.5, the impact of changes in residential commuting market access (RCMA) on overall population share are shown. The estimated elasticity is close to 0.6 in OLS and slightly smaller in magnitude for IV (0.5). In other words, a 10% increase in market access for residents, due to the Luas light rail, brought about a 6% increase in a district's share of the city's population. The second and third pair of columns show the estimated coefficients where only higher-skilled or lower-skilled populations are considered. These results indicate a link between changes in population density and skill level: the elasticity for higher-skilled workers to the change in market access is greater than for lower-skilled workers. A 10% increase in market access brings about a 7% increase in a district's share of higher-skilled workers but only a 5% increase in lower-skill share. The gap between the two shares is statistically significant: where the outcome of interest is instead the (log) within-district share that is higher-skilled, the coefficient is positive and statistically significant, while the coefficient for lower-skilled is negative and statistically significant.

Table 3.5: Effect of introduction of LUAS on probability of commuting

	<i>Dependent variable:</i>					
	log(share of pop)		log(share of HS)		log(share of LS)	
	OLS	IV	OLS	IV	OLS	IV
$\Delta \log(\text{RCMA})$	0.587*** (0.061)	0.487*** (0.089)	0.693*** (0.069)	0.600*** (0.098)	0.574*** (0.065)	0.472*** (0.094)
PC FE	Yes	Yes	Yes	Yes	Yes	Yes
ED area	Yes	No	Yes	No	Yes	No
Intercept	No	No	No	No	No	No
F 1 st stage		281.424***		281.424***		281.424***
Observations	322	322	322	322	322	322
R ²	0.436	0.403	0.400	0.387	0.438	0.399
Adjusted R ²	0.386	0.353	0.347	0.335	0.388	0.349
Residual Std. Error	0.293	0.301	0.330	0.333	0.309	0.319
F Statistic	8.759***		7.567***		8.834***	

This table shows the estimation of the main specification. Log share of population of is the log of the share of resident population of each ED over the total population resident in Dublin, Log share of HS is the log of the share of higher-skilled workers resident in each ED over the total workers resident in each ED, Log share of LS of is the log of the share of lower-skilled workers resident in each ED over the total worker in each ED and log(RCMA) is log of calibrated residential market access defined as per equation 3.9

Note:

*p<0.05; **p<0.01; ***p<0.001

In Table 3.6, I repeat the same analysis using firm commuter market access as the regressor and, as the outcome of interest, a district's share of the city's workplace population. As with changes in RCMA and where people live, the introduction of the *Luas* brought a change in the distribution of where people work. This is shown in the first pair of columns. A 10% increase in FCMA brought about a 4.5% increase in a district's share of the city's employment. The second and third pair of columns again break this down by skill level. For employment, the results are the opposite of residence: changes in market access had a greater impact on lower-skilled workers than higher-skilled ones. The impact of a 10% increase in FCMA was above 5% for lower-skilled groups compared to below 4% for higher-skilled groups. Again, these differences are statistically significant: the within-district share of lower-skilled workers is positively associated with changes in FCMA.

Combining both sets of results suggests that, after the *Luas* was introduced, generally more people work in areas well-connected by it, because the number of high-amenity areas reachable is now higher. The workplace areas connected, however, were marginally skewed to lower-skilled jobs: areas where the *Luas* was introduced could host lower-skilled workers. For residents, however, the impact went the other way: higher-skilled residents responded more to the changes in market access.

This latter finding is suggestive of higher-skilled groups pricing out lower-skilled ones in areas with the best transport facilities. I investigate this in Tables 3.7 and A.12, where the outcome of interest is the district-level change in residential rents between 2002 and

2016, as reported in the Census. However, the analysis presented consistently finds no systematic link between changes in market access and rents recorded in the Census.

Table 3.6: Effect of introduction of LUAS on share on probability of commuting

	<i>Dependent variable:</i>					
	log(share of workers)		log(share of HS)		log(share of LS)	
	OLS	IV	OLS	IV	OLS	IV
$\Delta \log(\text{FCMA})$	0.442*** (0.045)	0.469*** (0.049)	0.399*** (0.071)	0.365*** (0.081)	0.519*** (0.051)	0.552*** (0.057)
PC FE	Yes	Yes	Yes	Yes	Yes	Yes
ED area	Yes	No	Yes	No	Yes	No
Intercept	No	No	No	No	No	No
F 1 st stage		1097.974***		1142.985***		1142.985***
Observations	319	319	307	307	307	307
R ²	0.383	0.363	0.228	0.201	0.429	0.419
Adjusted R ²	0.328	0.308	0.156	0.130	0.376	0.368
Residual Std. Error	0.426	0.432	0.587	0.596	0.420	0.422
F Statistic	6.961***		3.179***		8.082***	

This table shows the estimation of the main specification. Log share of workers of is the log of the share of workers of each ED over the total workers in Dublin, Log share of HS of is the log of the share of higher-skilled workers working in each ED over the total workers working in each ED, Log share of LS is the log of the share of lower-skilled workers working in each ED over the total workers working in each ED and log(FCMA) is log of calibrated residential market access defined as per equation 3.10

Note:

*p<0.05; **p<0.01; ***p<0.001

Market conditions may be relevant here: overall, as indicated by the constant, rents rose substantially between 2002 and 2016, as the city's population grew steadily but the housing stock for most of the period did not. Alternatively, it may be the case that Census rents reflect market conditions with a lag. This is due to the fact that the census rent data also capture all long-term rents, which, in a market with a rent increase cap, are mechanically lower as they are renewed at a lower frequency. As a result, the census rent data may not fully capture the true rental market dynamics and trends, although arguably any effects of a system opened in 2004 should be visible over a decade later. Regardless, a suggestion for future research is to supplement the analysis here with data on listed/market rents at the district level as, as mentioned before, census rents also encompass long-term rents, whose increase is capped and therefore not indicative of rent market dynamics.

In Table A.12, I separate the improvement in market access due to each of the two lines separately and also include measures of housing quality provided in the Census. Doing this is equivalent to a counterfactual where only either the red or the green line was introduced and then the model is calibrated accordingly. Again, I find no significant impact of market access on rents. Overall, the results suggest that there was a crowding-out effect of low-skilled labour due to the introduction of the *Luas*. This effect, however, was at the extensive margin – the number living in each district, by skill level – rather than at the intensive one, as revealed by rents by district.

Table 3.7: Effect of introduction of LUAS on ED rents

	log(residential rent)			
	All buildings		Older buildings	
	OLS	IV	OLS	IV
$\Delta \log(\text{RCMA})$	0.105 (0.058)	0.046 (0.081)	-0.045 (0.198)	-0.341 (0.299)
PC FE	Yes	Yes	Yes	Yes
ED area	Yes	No	No	No
Intercept	No	No	No	No
F 1 st stage		281.424***		30.565***
Observations	322	322	49	49
R ²	0.156	0.150	0.271	0.224
Adjusted R ²	0.081	0.079	-0.0004	-0.064
Residual Std. Error	0.276	0.276	0.172	0.178
F Statistic	2.094**		0.998	

This table shows the estimation of the main specification. Log residential rent for all buildings is the change in rent per square meter of all buildings reported in the census, log residential rent of older buildings is the change in rent for buildings built before 1940 and and log(RCMA) is change in log of calibrated residential market access defined as per equation 3.9

Note:

*p<0.05; **p<0.01; ***p<0.001

Finally, I perform a placebo test, using the Dublin & Lucan Electric Tramway Line as a counterfactual scenario. Another option would be to use the extensions plans of the LUAS made in 2017, however these extensions were really short and in the immediate surrounding of the line and it is therefore very likely that those areas did benefit from the LUAS even before its extension. The Dublin & Lucan Electric Tramway, on the other hand, was a tramway system that operated from 1881 until 1925 and served the western part of Dublin, connecting the city with the town of Lucan, in the northeast, therefore nowhere close to the actual placement of any of the two lines. I build a counterfactual scenario where the Luas was built along this line instead of its actual placement and re-calibrate CMA access in this setting. If there were unobservable effects not strictly linked to the Luas, I would observe a significant coefficient. Tables A.13 and A.14 show the results of the regression with the CMA computed using the placebo line. The coefficient is in both cases insignificant, meaning that the actual placement of the Luas is what had an effect on the changed population and workers' dispersion.

3.5 Conclusion

In this paper, I have analysed the impact of a new infrastructural investment at the extensive margin, in the context of a European capital city like Dublin. I have calibrated a general equilibrium model to obtain a measure of market access, which captures the

effects of the introduction of the *Luas* on commute cost and run a difference in difference instrumental regression to inspect its impact on changes in population share, gentrification patterns, and rent levels. My results are partly in line with the existing literature, however, some differences stand out. Like Warnes (2020), I find a positive impact of the introduction of new infrastructure on resident and workplace population shares, meaning that more people moved their residence or their workplace close to where the improvement of market access took place. However, I find that this migration was not homogeneous across higher- and lower-skilled workers, in fact, the elasticity for the former to the change in market access is greater than that for the latter. Also, unlike Tsivanidis (2019), I do not find an effect on long-term rents.

Overall, I believe this paper brings three main contributions to the existing literature. First, it calibrates a general equilibrium model on Dublin and it retrieves its city-specific parameters. Second, to my knowledge, this is the first paper that quantifies the impact of infrastructural transport investment on the spread of population in a European capital city. The majority of existing literature concentrates on cities in developing countries, while I believe the comparison is interesting in a more mature urban context. Third, it makes a further link between gentrification and transport. I observe in fact that higher-skilled workers relocate to places where market access has improved, while lower-skilled workers rather work in better-connected areas.

I am aware that this work suffers from a few limitations. Incomplete information on the precise localisation of market rental listings does not allow for a reliable analysis of the effects on market rent level of improved market access. Also, census records at the Small Areas level are not available for 2002 and I am thus unable to exploit information on changes at a more granular level. This implies the sample size in my analysis is quite limited - 322 EDs versus 4990 potential Small Areas - and therefore the power of my estimates is also limited.

Nonetheless, results are significant and solid and I believe this research shows the importance of policymakers being aware that major public transport investments can have distributional effects, in particular in the case of Dublin, concentrating higher-skilled cohorts, in residential terms, in areas that are better served by the new transport system, even as – from an employment perspective – lower-skilled groups benefited more.

More research is needed to uncover the effects on residential rents, including market rents, rather than average long-term rents. Similarly, the impact on commercial rent is also a topic worthy of analysis, however, so far I did not find suitable data. Finally, this same setup can be extended to analyse how extensive and intensive margins relate: the elasticity of housing supply is likely to be pivotal in determining the relative impact of new transport

infrastructure over the skill and income distribution. I believe this is of great relevance in an urbanising world because it not only helps understand the pattern of city dynamics but it also poses questions in terms of additional policies necessary alongside infrastructural investments. While improvements in a city's transport system are fundamental, it is also important to fully understand all its implications.

Chapter 4

Market segmentation and long-run competition: the case of the Italian rail market¹

¹ This paper is co-authored with Paolo Beria.

Abstract

Rail liberalisation in Europe is an ongoing process, with a high degree of heterogeneity across countries. Its long-run effects on the rail market however are still debated. This paper focuses on the specific case of the Italian rail market and explores the link between the presence of competition, price level and frequency. It first identifies statistically two distinct market segments, which allows for analysing routes with and without competition within the appropriate context. It then exploits this segmentation to explore the evolution of prices and frequency. Results are not always in the expected direction, in fact, they vary considerably deepening on the segment. Within one segment, in fact, there is a decrease in early booking tariffs, within one year of entry of the newcomer, while, in the other segment, entry is linked to an increase in prices on competitive routes. Frequency on the other hand is affected positively until before the pandemic in one segment, whereas no effect is visible in the other segment.

4.1 Introduction

The idea that the rail market should be a protected state monopoly has become increasingly obsolete in recent years. In fact, since 1992, transport market liberalisation has been central in the European Agenda (EC, 1992). The rationale is that competition decreases prices, and increases passenger volume and service quality, thus increasing overall consumer benefits. However, whether these are the true effects of liberalisation, especially in the long run, is still unclear.

In this paper, we aim to shed light on the true effects of rail market liberalisation on prices and service quality, nearly 35 years after the start of the European liberalisation process. By exploiting a unique dataset on daily prices on a set of 33 representative Italian routes, for the two competing companies, we observe the differences in their pricing strategies, the difference between routes in long-run competition versus those where competition started only recently and between those in competition since fairly recently and those not in competition at all for the incumbent player. Previous literature (Cascetta and Coppola (2015), Giuricin (2018), Beria et al. (2022b) and Beria et al. (2022c)) observed that market liberalisation has driven prices down and service quality up in the short-run; in our paper, we point out that effects are substantially different in the long-run. We observe a constant increase in prices for both companies and we do not find any effect of entrance on new routes when compared to routes where competition was already in place. We only observe a generalised price increase after entry, slightly offset by a one-year decrease in average prices of early booking discounted tickets. When analysing the impact of entry compared to routes where no competition is in place, the result is even more unexpected, as the prices of the treated group after entry actually increase.

This issue is at the intersection of the industrial organisation literature and the transport literature and has been therefore tackled by both with two different approaches; the former more in a perspective way while the latter rather in terms of retrospective case studies. As per what concerns the former, several game-theoretical models have been designed, especially back in the early 2000s, to forecast the effects of liberalisation in the transport sector and then subsequently in the rail sector more specifically. It is interesting to notice how many theoretical papers to some extent were able to “foresee” what I observe in this work. Notably Preston et al. (1999) was the first to make an attempt in this sense, applying a game theory model in the rail sector, followed by Whelan and Johnson (2003), who developed an ad-hoc model, PRAISE, to simulate competition on rail networks. Preston et al. (1999) found that in the presence of a minimum efficient scale often there is no way two competitors can profitably survive on a market with high entrance cost unless they enter an agreement of some sort, and therefore did not see much potential for improvements

in economic efficiency unless it could induce cost reductions. At the same time, Bitzan (2003) claims that railroads are natural monopolies and that multiple-firm competition would result in increased resource costs, such that the price decreases necessary for the introduction of such competition to be beneficial would be large. More recently, Johnson and Nash (2012) pointed out that it is likely that competition generally leads to excessive levels of service and costs, for which the benefits to consumers are more than offset by the losses of profitability to the existing operator. Ruiz-Rúa and Palacín (2013) and Adler et al. (2021) provide a more technical and extensive review covering the models, which we leave out for sake of brevity, however the general point, on which the theoretical literature agrees, is that no major price decrease is possible in the context of liberalisation.

As per what concerns the transport literature, the consequences of market liberalisation have been studied for each of the three main transport sectors separately. In the context of the airline industry, Gerardi and Shapiro (2009) find that price dispersion declines upon entrance of a new player on a route, while in a monopoly prices tend to be very clustered around a higher mean during peak periods, thus reducing dispersion (Gaggero and Piga, 2011). In the coach sector Blayac and Bougette (2017) show that benefits are as expected - quality, quantity, new routes, etc. – except for fares. In fact, after an initial period of lower fares, prices tend to raise again. Also Beria et al. (2018) document this same phenomenon; after a short launch period when the incumbent prices below average market level and the newcomer is more aggressive, prices go up again for both players.

Full on-track competition in the rail sector is more recent. This implies that this literature is not as rich, and it mostly focuses on short-term equilibrium, however some evidence for the long-term exists as well. Using a panel of 69 routes across Europe, Beria et al. (2022b) provides an extensive summary of all possible outcomes of inter- and intra-modal competition: increased quality, increased supply, increased network and decreased prices. Vigren (2017) and Tomeš and Jandová (2018) find prices are decreased as a consequence of entrance of a new player in the Swedish and central European rail market respectively. In the Italian setting, early studies find a decrease around 30-40% with respect to initial incumbent levels (Cascetta and Coppola (2015) and Giuricin (2018)), and strategic pricing on both sides, although to a different degree on different routes. In this setting, both companies appear to take into account the dynamics of rival's pricing behaviour (Bergantino et al. (2015), Beria and Grimaldi (2016) and (Bergantino et al., 2018)), but no clear evidence of price-wars exists (Beria and Bertolin, 2019). Again in the Italian setting, looking at the market at an equilibrium stage, later research suggests reductions are smaller. Beria and Bertolin (2019) and Beria et al. (2016), for example, compare routes with and without competition, finding that, the newcomer, NTV/Italo, is usually 10% to 20% cheaper than the incumbent, Trenitalia, but that Trenitalia's prices are not significantly different among

routes with and without the competitor. Along the same line, but exploiting before and after Italo's entry effects on the Milan-Venice route, Beria et al. (2022c) find that there is an immediate effect on prices. This effect vanishes within six months for prices one day before departure but is still visible for advanced bookings, which however is reduced within one year. Our paper aims at extending their analysis to 33 OD pairs, grouped into market segments that are more easily comparable.

We believe this paper allows to generalise in a more statistically rigorous manner the thesis expressed by Beria et al. (2019), on the short-lived effects on prices of a competitive market. By exploiting a difference in difference estimation, we analyse price trends over five summer periods and find a constant increase, which is independent of the treatment status of the routes. Frequency, on the other hand, after an initial increase, remains higher, however impacted by the COVID crisis. What's more, it represents an empirical validation of what predicted by the theory more than 20 years ago from now.

The rest of the paper is structured as follows. Section 4.2 provides an overview of the Italian rail market structure, section 4.3 describes the data, sections 4.5.1 and 4.5.2 provide an analysis of effects of competition within each identified segment, section 4.6 provides an explanation of the mechanisms of price increase and finally section 4.7 concludes.

4.2 Italian Rail Market

The Italian rail market provides an interesting setting for analysis both for geographic and structural reasons. From the point of view of the first, Italy is characterised by a peculiar urban geography with several large and small cities aligned along a limited number of corridors (Beria et al., 2020), and a major North-South divide. On the structural side, there is a reasonably well-performing rail network, where the publicly owned operator Trenitalia runs regional and intercity trains (Public Service Obligation, PSO) at capped prices and faster long-distance trains at market price in an on-track competition with NTV-Italo, a private company.

On-track competition is common also to Czech Republic and Slovakia (Tomeš and Jan-dová, 2018), Sweden, (Fröidh and Nelldal (2015) Vigren (2017)) Austria (Tomeš and Jan-dová (2018)) and more recently Germany (Guihéry, 2020) and France, however it happens mostly on intercity services or regular fast services.² The Italian case is an *unicum* in that competition happens on high-speed tracks and, as of now, it is been in place for nearly 10 years (Beria et al., 2022a).

As early as 2003 in fact, the Italian government issued a *decreto legislativo* - a legislative

² With the exception of France where competition on HS tracks started in December 2021 and Spain.

decree - which implemented the European Directives on rail competition (2001/12/CE, 2001/13/CE, and 2001/14/CE). According to the decree, any licensed rail company can have access to the national infrastructure manager's (RFI) rail network and operate both open access or Public Service Obligation (PSO) contracted services (regional and intercity). While the PSO segment remained under the management of Trenitalia, the market segment was soon entered by a new player.

No later than 2006 a new rail company, ArenaWays, was founded and by 2010 it started operation. This first experience, however, was short-lived and ArenaWays went bankrupt two years later. In 2007 another company was founded, NTV, which started operating in 2012 under the commercial name of *Italo*. At least initially, NTV entered the market on the main high-speed routes and was operating at a loss. It was only in 2015 that it started making profits, which continued increasing since³ and only halted with the COVID-19 crisis, even though they were still positive thanks to the substantial government financial aid. In response to this, Trenitalia – the incumbent company - increased supply on the high-speed segment already in 2013 with respect to its 2009 levels (Bergantino et al., 2015). Italo generally operated the same routes as Trenitalia however both companies increased frequency and stops to meet increased demand (see figure 4.1).

Total passengers per km went from just below 15 billion to nearly 25 in 2018, allowing absolute figures to increase for both companies, despite Trenitalia's loss in share. Press releases for data in 2019, suggest a further increase for both companies, placing Italo at 20 billion passengers per km, and Trenitalia at 40 (Legambiente, 2021).⁴ In general, on pairs where both companies operate, Italo's share of supplied train-km ranges between 30% and 50% or more (Beria and Bertolin, 2019), making it a competitor far from marginal.

4.3 Data

In order to undertake our analysis, we need data on prices and daily frequency of trains managed at market price both by the incumbent (Trenitalia) and the newcomer (Italo). As mentioned before, there exists a PSO segment as well, whose prices are regulated⁵ and whose frequency off-peak is controlled by the government, and which is, in principle, also open to competition, however so far only Trenitalia has won the one-candidate tender for it. We, therefore, disregard it to focus only on the segment not under PSO.

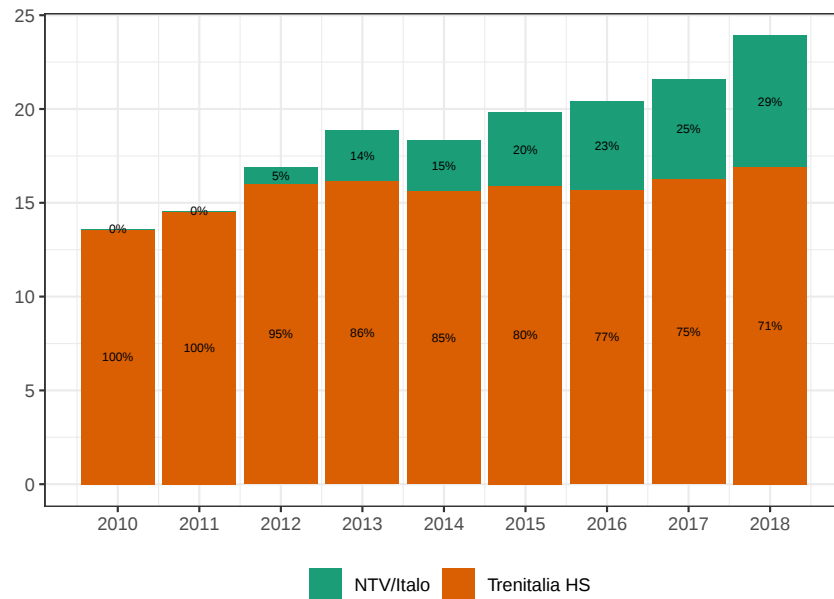
Before describing the data in what follows we provide a quick overview of the fare system in Italy. For each train, there is a fixed number of full-flexibility, partial-flexibility and

³ Source: NTV and Trenitalia public end-of-year balance sheets

⁴ No official data is released by either company after 2018.

⁵ The regulation of prices on the PSO segment consists of a maximum mean price over all routes on a yearly basis, not on a single route cap on price.

Figure 4.1: Passenger per km in billion

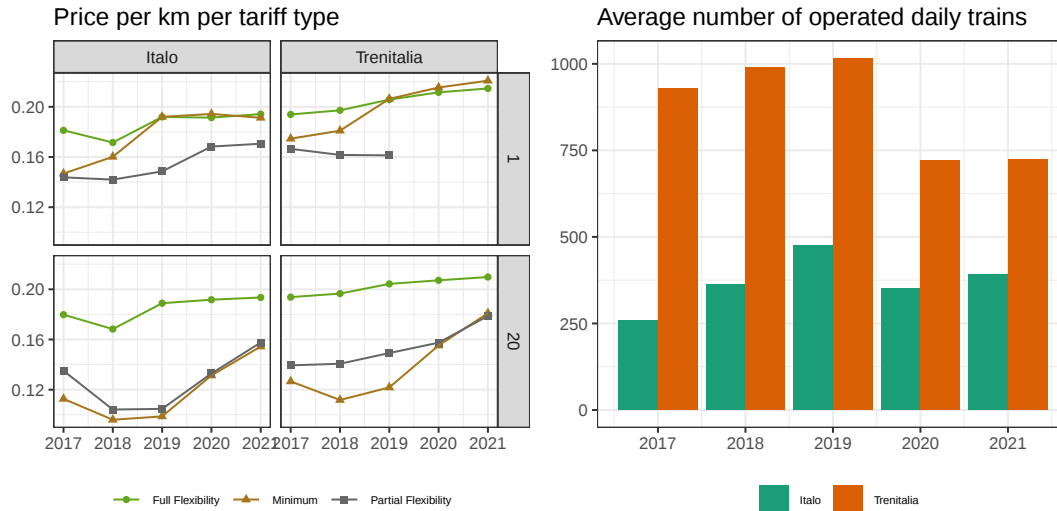


*This graph shows the total passenger*km on the market segment of both companies in billions.
Source: NTV (2019) for Italo (no data is published for the years 2019 and later), CNIT (2020) for Trenitalia*

no-flexibility tickets, where clearly the sum of the three is the total number of available seats on a train. Full-flexibility tickets are more expensive and allow for the possibility to change or cancel, followed by partial flexibility, which only allows for change and finally by no-flexibility tickets, whose name speaks for itself and which is clearly the cheapest option. So for example, if a train has 100 seats available, 80 are sold at full-flexibility price, 15 at partial-flexibility and 5 at no-flexibility. Like in other countries, within every fare type, one can choose between first and second class, which differ for the number of seats per row (three for the first and four for the second). This implies that there are 6 possible fare type-class combinations travellers can choose from (provided none is sold out at the time of purchase).

The full dataset we are drawing from is composed of observations collected directly from Trenitalia booking websites, on a selection of 33 Origin-Destination (OD) pairs representing relevant situations in the Italian domestic market between 2017 and 2021. The price for each fare-class combination - if available - for every train operated on each route for each day in the period of observation is registered. All six price combinations are observed 1, 2, 10 and 20 days before departure. For example, if in one day there are 40 trains on the route Milan-Rome, all six fare-class combinations are available 20 days before departure and only 3 one day before departure, we have 240 data points for the Milan-Rome 20 days before departure and 120 for 1 day before departure. Besides the six tariff-class combinations, the dataset comprises also the minimum price on each route

Figure 4.2: Summary statistics



Note: The left panel shows the evolution of price per km for every tariff type (Full Flexibility, Partial Flexibility and Minimum tariff) and every simulation of anticipated booking (-1 and -20) for summer 2017 to 2021, on the 33 routes considered. The right panel shows the average number of total operated trains daily for summer 2017 to 2021 for the two companies.

for each day for every simulation of anticipated booking. This price heavily depends on the load factor of trains - it can also be a full-flexibility first-class fare if everything else is sold out - and can have big variations, however, it provides a good overview of prices the average consumer faces.

In this paper we use a subset of this dataset, concentrating on comparable summer periods - June, July, and August between 2017 and 2021 - for two simulations of anticipated booking - 1 and 20 days before the date of travel - for second class service and for partial- and full-flexibility tariffs, plus minimum price, for a total of 6 possible price combinations for each company. We disregard no-flexibility tariffs because the sample size is very small and they disappear altogether one day before departure by 2019.

Table 4.1 reports summary statistics on average price per km and sample size for each fare type, time of booking and company. It has to be noted here that the average minimum price is sometimes higher than the partial flexibility tariff, which can be explained by the fact that partial flexibility tariffs might not always be available at the time of booking and therefore the minimum price coincides with the base tariff. Therefore, we can think of it as an average between partial, full and no-flexibility tariffs, weighted by their availability. Figure 4.2, left panel, shows their behaviour overtime, while the right panel shows the evolution over the five summer periods of the average number of daily trains for each company.

Table 4.1: Summary Statistics

Price per km per tariff type												
Year	Trenitalia						Italo					
	-1			-20			-1			-20		
	Min	PF	FF	Min	PF	FF	Min	PF	FF	Min	PF	FF
2017	0.175	0.167	0.194	0.127	0.139	0.194	0.147	0.144	0.181	0.113	0.135	0.180
2018	0.181	0.162	0.197	0.112	0.141	0.197	0.160	0.142	0.172	0.096	0.104	0.168
2019	0.206	0.161	0.206	0.122	0.149	0.204	0.192	0.149	0.192	0.099	0.105	0.189
2020	0.215		0.212	0.155	0.157	0.207	0.194	0.168	0.191	0.131	0.133	0.192
2021	0.221		0.215	0.181	0.179	0.210	0.191	0.171	0.194	0.154	0.158	0.193

Average daily sample size												
Year	Trenitalia						Italo					
	-1			-20			-1			-20		
	Min	PF	FF	Min	PF	FF	Min	PF	FF	Min	PF	FF
2017	31	25	31	32	30	32	23	13	23	24	22	24
2018	34	21	33	34	33	34	27	13	27	31	30	31
2019	33	13	32	33	31	33	23	8	23	23	22	23
2020	24		23	21	20	21	19	15	17	19	18	18
2021	27		25	27	22	26	19	17	19	19	18	19

Note: This table shows the mean price per km computed on all 33 routes available, for each tariff (Minimum available tariff (Min), Partial Flexibility (PF), Full Flexibility (FF)) and for each simulation of anticipated booking (1 and 20) and the average daily sample size for each tariff type. Minimum tariff can be higher than PF tariff because when PF are no longer available, the minimum price is base tariff.

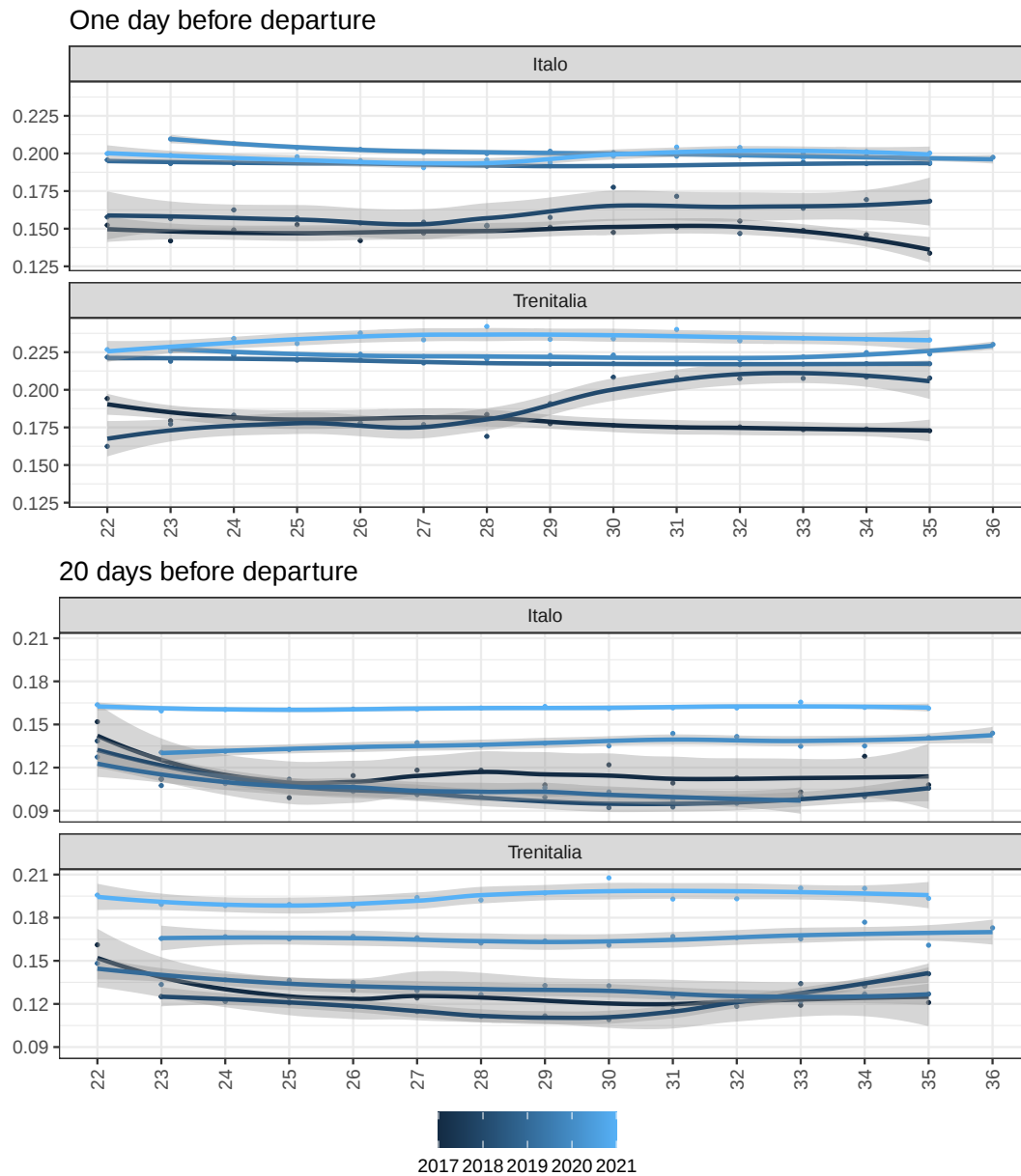
4.4 Trenitalia and Italo: a comparison

In this section, we perform a comparison between prices and frequency on routes run by the two companies. Figure 4.3 helps us in this task. The top panel reports minimum daily prices per km one day before departure and the bottom panel reports minimum daily prices 20 days before departure, for the period 2017 to 2021 for each company. For coherence, we focus here only on routes where Italo has been competing since before our observation period starts, as to observe the average over the same routes at each point in time. This implies that we are comparing prices of Bologna-Firenze, Bologna-Venezia, Milano-Bologna, Milano-Firenze, Milano-Roma, Milano-Torino, Roma-Bologna, Roma-Firenze, Roma-Torino and Roma-Venezia only. We leave the analysis of routes where competition started later in section 4.5.1.

For prices one day before departure (top panel), there are two distinct periods of pricing for both companies, with the first period (2017-2018) exhibiting relatively low prices and the second period (2019-2021) showing significantly higher prices after an abrupt increase in August 2018 (see section 4.6 for the mechanisms). The increase in prices after the health crisis in 2020 appears to be relatively small in comparison to the earlier increase, suggesting that the current higher prices are not necessarily a result of the health crisis, but rather a continuation of a general increase that occurred prior to the crisis.

The bottom panel shows the prices of the same tickets 20 days before departure, which exhibits a different pattern compared to the prices one day before departure. This pattern can be divided into two periods: 2017-2018-2019 and 2020-2021. Prior to the COVID-19

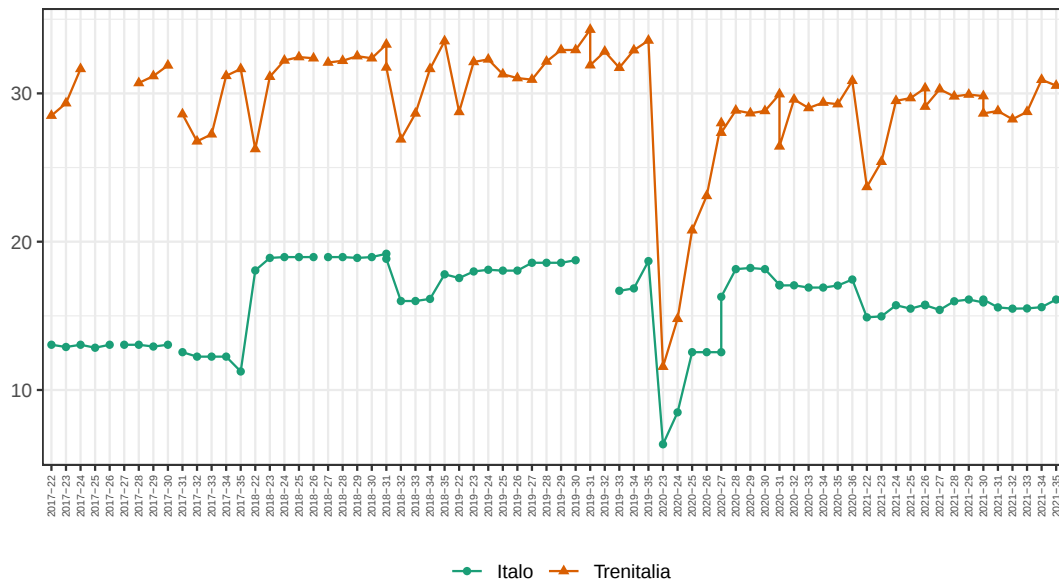
Figure 4.3: Minimum Tariff - Core segment



Note: This figure shows the evolution of average prices per km on routes which were in competition since before 2017 for both companies. Routes included are: Bologna-Firenze, Bologna-Venezia, Milano-Bologna, Milano-Firenze, Milano-Roma, Milano-Torino, Roma-Bologna, Roma-Firenze, Roma-Torino and Roma-Venezia.

pandemic, the prices for both companies remained at a lower level with a constant offset between them. In fact, during the summers of 2018 and 2019, the prices for Trenitalia were even lower than in 2017. However, after the COVID-19 crisis, there was a significant increase in prices, especially during the summer months, making travel more expensive even for those who book in advance.

Figure 4.4: Average daily frequency - Core segment



Note: This graph shows the average daily frequency over routes where competition is in place since before 2017. Routes included are: Bologna-Firenze, Bologna-Venezia, Milano-Bologna, Milano-Firenze, Milano-Roma, Milano-Torino, Roma-Bologna, Roma-Firenze, Roma-Torino and Roma-Venezia.

Excluding the impact of the COVID-19 pandemic, the data suggests that competition does not significantly affect the prices of last-minute bookings for either the incumbent or the newcomer. However, it does provide a cheaper option for those who book in advance. This can be attributed to the fact that Italo operates with a lower frequency on all routes compared to Trenitalia, which means that travellers who are willing to pay a premium for the convenience of taking the first available train will likely choose Trenitalia over Italo. As a result, Trenitalia can maintain its market share even with higher prices, while Italo has no incentive to lower its prices. On the other hand, for early bookers who are more organized and have more time to plan, the profit margin is smaller, and as a result, both companies must compete more on price to attract these customers.

Figure 4.4 shows that in 2017 when prices were generally lower, frequency was also lower, and it increased until 2019 for both Trenitalia and Italo. The COVID-19 pandemic caused a significant decrease in frequency in 2020, but there was a rebound in 2021. When calculating the correlation coefficient between price and frequency, the resulting value was 0.09, which indicates a positive but weak correlation. This suggests that com-

petition has a positive effect on frequency for both the incumbent and the newcomer, but its relationship with prices is not as strong.

4.5 Competition effects

In this section, we want to estimate the impact of competition on prices. In order to do so we split our sample into two segments, a core and a tourist segment. We do so because the price behaviour in these two segments is structurally different and one would not be a sound control for the other. Details of routes assigned to each segments are reported in table 4.3, along with the presence of high-speed tracks (HS) on the route, the average speed, the date of entry of Italy and the presence of stated regulated competition (Intercity competition). We observe that routes belonging to the core segment are all partially or completely on HS rails, they are more frequent, more expensive, and they link the main “business” centres in the country. Tourist routes, on the other hand, are generally slower, less frequent, less expensive, and they link mainly cities to main holiday destinations. What’s more, in the core segment, competition started very early whereas, in the tourist segment, some routes are still only operated by Trenitalia. Differences between segments are summarised in table 4.2.

Table 4.2: Market segments characteristics and differences

	Daily trains		Average km		Average price per km		Speed	
Year	Core	Tourist	Core	Tourist	Core	Tourist	Core	Tourist
2017	19	8	330	423	0.203	0.169	140	116
2018	20	8	330	423	0.207	0.171	140	117
2019	21	8	330	423	0.216	0.176	138	116
2020	15	8	330	423	0.219	0.190	131	114
2021	17	8	330	423	0.221	0.191	131	118

Note: This tables reports the five main differences between Trenitalia minimum prices of the two market segments, obtained performing the cluster analyses on Trenitalia minimum prices for summer periods between 2017 and 2021.

4.5.1 Competition effects on core routes

We start by analysing the effects of competition in the core segment. In this segment, we include the following routes: Bologna-Firenze, Bologna-Venezia, Milano-Bologna, Milano-Firenze, Milano-Napoli, Milano-Roma, Milano-Torino, Roma-Bologna, Roma-Ferrara, Roma-Firenze, Roma-Torino, Roma-Venezia and Venezia-Firenze - where competition started in 2012 and which we use as control - and Bologna-Trieste, Milano-Brescia, Milano-Venezia, Roma-Bari, Torino-Brescia and Torino-Venezia where competition started in May 2018, which we use as treatment, whereby pre-treatment is the period 2017 and post-treatment are the years 2018 and 2019.⁶ For simplicity, we refer to routes in the

⁶ To avoid comparing pre- and post- pandemic years, we do not use 2020 and 2021 as control years.

Table 4.3: Market segmentation

Route	Segment	HS	Av. Speed	Italo comp	Intercity comp
Bari-Ancona	Tourist	No	102	No	Yes
Bologna-Ancona	Tourist	No	106	02/07/2020	Yes
Milano-Ancona	Tourist	Partly	104	02/07/2020	Yes
Milano-Pisa	Tourist	No	57	No	Yes
Milano-Rimini	Tourist	No	105	02/07/2020	Yes
Roma-Genova	Tourist	No	85	12/12/2021	Yes
Roma-ReggioCalabria	Tourist	Partly	86	14/06/2020	Yes
Roma-Verona	Tourist	Partly	142	Before 2017	Yes
Bologna-Bolzano	Core	No	87	01/08/2018	No
Bologna-Firenze	Core	Yes	137	Before 2017	Yes
Bologna-Trieste	Core	No	76	27/05/2018	Yes
Bologna-Venezia	Core	Partly	103	Before 2017	No
Milano-Bologna	Core	Yes	151	Before 2017	Yes
Milano-Brescia	Core	Yes	130	01/05/2018	No
Milano-Firenze	Core	Yes	141	Before 2017	Yes
Milano-Genova	Core	No	76	No	Yes
Milano-Napoli	Core	Yes	174	Before 2017	Yes
Milano-Roma	Core	Yes	173	Before 2017	No
Milano-Torino	Core	Yes	180	Before 2017	No
Milano-Udine	Core	No	80	No	No
Milano-Venezia	Core	Partly	110	01/05/2018	No
Roma-Bari	Core	Partly	95	27/05/2018	Yes
Roma-Bologna	Core	Yes	169	Before 2017	Yes
Roma-Ferrara	Core	Partly	123	Before 2017	Yes
Roma-Firenze	Core	Yes	169	Before 2017	No
Roma-Torino	Core	Yes	169	Before 2017	Yes
Roma-Venezia	Core	Partly	144	Before 2017	Yes
Torino-Brescia	Core	Yes	127	01/05/2018	No
Torino-Venezia	Core	Partly	130	01/05/2018	No
Venezia-Firenze	Core	Partly	107	Before 2017	Yes

Note: This table shows the assignment of each route to its segment, whether it is on high speed track entirely, partly or not at all (HS), whether there is competition from the newcomer player Italo and whether on the same PSO trains are also operated.

former group as long-term competitive routes, while to routes in the latter as short-term or newly competitive routes. In order to estimate the impact of entry we use two different methods: an averaged before-after difference in difference with bootstrapped standard error and a pooled difference in difference with clustered standard error. Both specifications are described by the following equation:

$$Price_{rt} = \alpha + \beta_1 TripLength_r + \beta_2 Treated_r + \beta_3 PostEntry_r + \beta_4 DID_{ri} + \varepsilon_{rt} \quad (4.1)$$

where $TripLength_r$ is a variable controlling for the length of route r ; $Treated_r$ is a binary variable indicating whether route r is part of the treated or the control group; $PostEntry_r$ is a dummy variable which takes value one in the post-treated period; DID_{ri} is a dummy

variable which takes value one if route r is treated at time t . The dependent variable $Price_{rt}$ varies according to the specifications. In the averaged difference in difference case, it is the average price of route r over time t , where t indicates the period before ($t = 0$) or after ($t = 1$) treatment. In the pooling difference in difference case, on the other hand, $Price_{rt}$ is the price of route r for time t , where t takes value zero for all day-route combinations before the treatment period and one after (not their average over the entire pre or post-treatment period). We run each estimate for 1 and 20 days before departure and for the difference in prices between the year before treatment and the year of treatment (2017 vs 2018) and for the year before treatment and the year after treatment (2017 vs 2019).

Table 4.4: Averaged Diff in diff estimation on core segment

	log(price)			
	1 day before		20 days before	
	2017 vs 2018	2017 vs 2019	2017 vs 2018	2017 vs 2019
Trip length	−0.001*** (0.0001)	−0.001*** (0.0001)	−0.001*** (0.0002)	−0.001*** (0.0002)
Treated	0.018 (0.060)	0.014 (0.059)	0.040 (0.088)	0.042 (0.084)
Post-entry	0.075 (0.045)	0.203*** (0.044)	−0.064 (0.067)	0.032 (0.065)
DID	−0.025 (0.083)	−0.030 (0.082)	−0.091 (0.122)	−0.106 (0.117)
Constant	−1.375*** (0.050)	−1.356*** (0.049)	−1.704*** (0.074)	−1.710*** (0.071)
Observations	38	38	38	38
R ²	0.772	0.820	0.628	0.631
Adjusted R ²	0.744	0.799	0.583	0.586
Residual Std. Error (df = 33)	0.125	0.121	0.186	0.178
F Statistic (df = 4; 33)	27.882***	37.701***	13.942***	14.099***

This table shows the averaged before and after diff-in-diff estimation with bootstrapped standard error.
Note: *p<0.05; **p<0.01; ***p<0.001

Results for the averaged difference in difference estimation are reported in table 4.4. In this regression, we treat the long-term competitive routes as control group and the newly competitive routes as treated group, whereby treatment we mean Italo's entry (June 2018). To avoid time autocorrelation issues - the price of a route on a given day is unlikely to be independent of its price the next day (see Bertrand et al., 2004) - the unit of observation is the average price of each route over the entire period before and after treatment. This results in a total of 18 routes * 2 (one pre-treatment and one post-treatment average) observations, and it is assumed that the prices of these routes are independent of each other. Because we are only left with 33 degrees of freedom, we use bootstrapped standard

errors.

Table 4.4 shows the results for the difference in average prices with respect to the year before entry (2017 vs 2018) and to the year of entry (2017 vs 2019) for one (columns 1 and 2) and twenty days before departure (column 3 and 4), respectively. The first coefficient in all the specifications, representing the "Trip length", is negative and statistically significant, indicating that a route that is 10 km longer is priced on average 1% less per km. This result is consistent with reality and holds for all four specifications. The constant term suggests that the unconditional expected mean of prices is 0.25 € per km ($\exp(-1.375)$) one day before departure and 0.18 € per km ($\exp(-1.704)$) 20 days before departure, which is consistent with the fact that advanced booking leads to lower prices. The only other significant coefficient is the after-treatment dummy for one day before departure in the year following treatment (2017 vs 2019), which is positive and of non-negligible magnitude, around 20%, meaning prices increased on average on all routes, the year after competition started. None of the other coefficients is significant, indicating that there is no significant difference in pricing between long-term and short-term competitive routes ("Treated"), between before and after treatment ("Post-entry"), or any specific effects for treated routes after treatment ("DID"). This suggests that, as a result of entry, there is no reduction in prices visible 1 or 20 days before departure.

Table 4.5: Pooled difference in difference on Core segment

	log(price)			
	1 day before		20 days before	
	2017 vs 2018	2017 vs 2019	2017 vs 2018	2017 vs 2019
Trip length	-0.002*** (0.000)	-0.002*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)
Treated	0.032** (0.011)	-0.099*** (0.014)	-0.082* (0.042)	-0.427*** (0.042)
Post-entry	0.072*** (0.005)	0.204*** (0.013)	-0.066* (0.028)	0.034 (0.023)
DID	-0.022 (0.019)	-0.030 (0.022)	-0.092 (0.066)	-0.107 (0.059)
Constant	-1.028*** (0.003)	-1.063*** (0.009)	-1.218*** (0.017)	-1.362*** (0.017)
Route FE	Yes	Yes	Yes	Yes
Observations	1,956	2,411	1,634	2,202
R ²	0.915	0.980	0.831	0.877
Adjusted R ²	0.914	0.980	0.829	0.876
Residual Std. Error	0.075	0.038	0.130	0.104
F Statistic	1,043.612***	5,875.202***	397.013***	776.794***

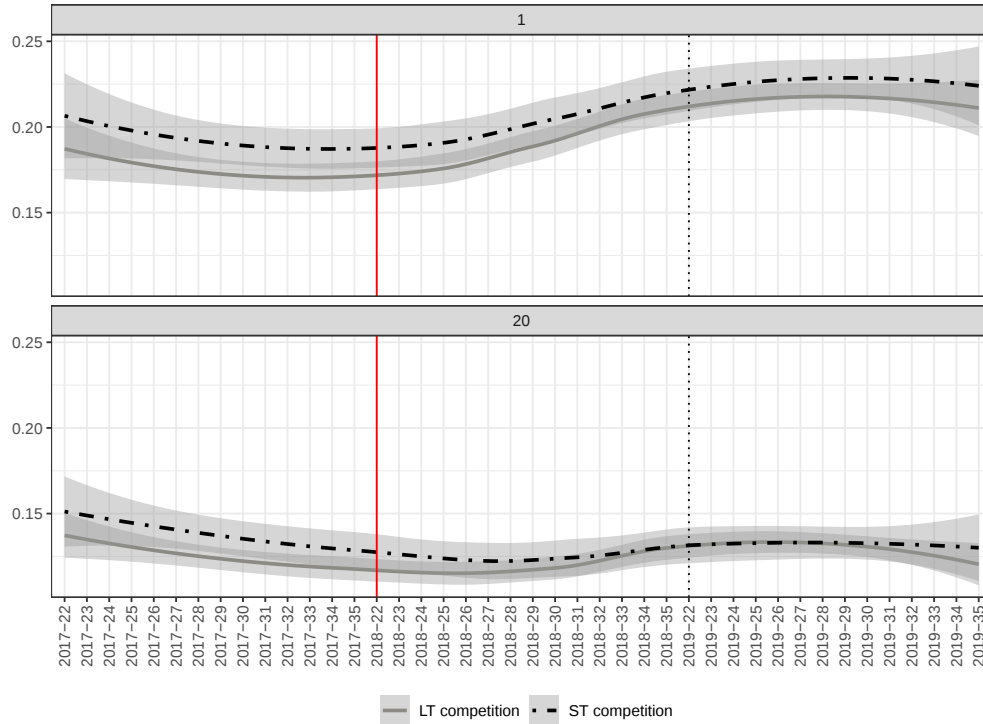
This table shows the pooling OLS with clustered standard error at route level.

Note:

*p<0.05; **p<0.01; ***p<0.001

Figure 4.5 illustrates the results discussed above. The solid vertical red line shows the time of entry while the black dotted line delimits the control years (end of 2018). This figure shows that there is a decrease in the difference between prices for competitive and newly competitive routes, but this difference is always within the error bands, reflecting the non-significance of the coefficients of the difference in difference regression.

Figure 4.5: LR competition vs routes in competition since 2018



Note: This figure shows the difference in prices between routes in competition since before 2017 and those where competition started in 2018. Routes in competition since before 2017 are: Bologna-Firenze, Bologna-Venezia, Milano-Bologna, Milano-Firenze, Milano-Napoli, Milano-Roma, Milano-Torino, Roma-Bologna, Roma-Ferrara, Roma-Firenze, Roma-Torino, Roma-Venezia and Venezia-Firenze. Routes where competition started in 2018 are: Bologna-Trieste, Milano-Brescia, Milano-Venezia, Roma-Bari, Torino-Brescia and Torino-Venezia. The solid line indicates the time when treatment starts (Italo's entrance), the dotted delimits the one year after treatment, the dotted line two years and the dashed the three years period after treatment.

We repeat the same analysis using a pooled difference in difference model with route fixed effects and clustered standard error at route level. The unit of observation in this analysis is the average price of each route for each day in the period before and after treatment, resulting in 18 routes * 180 days * 2 (before and after treatment). As before, we consider the post-treatment period to be the same year (2017 vs 2018) and one after Italo's entry (2017 vs 2019), using long-term competitive routes as the control group for newly competitive routes. The results of the pooled model with cluster standard error and routes fixed effects are reported in table 4.5. These results are broadly consistent with the averaged difference in difference estimation in terms of coefficient size, however, the increased number of degrees of freedom makes some coefficients significant. With the exception of trip length and the constant, which are mostly unchanged, we observe that after

treatment (“Post-entry” coefficient) there is a significant generalised price increase in both models for one day before departure, which is however counterbalanced by a generalised decrease in prices twenty days before departure, in the year of entry, with respect to the year before. In other words, prices one day before departure experienced a first 7% increase in 2018 with respect to their 2017 level, then a further 12% increase in 2019 (+20% with respect to 2017), however, this increase was partly counterbalanced by a 7% decrease in prices twenty days before departure, between 2017 and 2018. As per what concerns the treated group coefficient (“Treated”) it is of more difficult interpretation because of the presence of route fixed effects. In all but the first specification, this coefficient is negative, meaning that, net of all other route-specific factors and of treatment, the control group has generally lower prices than the treated group. The first specification, however, shows a positive coefficient for the treatment group, meaning that, on the other hand, the treated group is generally cheaper than the control. In this setting, we can only infer that some collinearity exists between the route fixed effects and the treatment coefficient, possibly of opposite sign, and we will therefore disregard the interpretation of this coefficient.

In summary, this section points out that the entrance of the newcomer on new routes does not contribute to lower prices of the incumbent on the treated routes nor to keep prices lower in the long term. It does however, seem to contribute to a counterbalancing mechanism whereby for one year after entrance an increase in prices one day before departure is offset by a decrease in anticipated discounted fares (twenty days before departure). These results contrast what found by Beria et al. (2022c), who found a significant decrease for treated routes, however, they did not cluster the standard error, thus having autocorrelation decreasing their size.

4.5.2 Competition on tourist routes

We now replicate the analysis on the tourist segment. In this segment, we include Bari-Ancona, Milano-Pisa, Roma-Genova – where Italo entered in July 2020 and which we use as treated – and Bologna-Ancona, Milano-Ancona and Milano-Rimini – which are, as of now, only operated by Trenitalia and are therefore our control. Once again, we use the model specified in equation 4.1, we consider prices both 1 and 20 days before departure and we observe differences one year before and after entry and all years before and after, meaning pre-treatment years are 2019 and 2020, until July) and post-treatment is 2020 after July and 2021. Unlike for the core segment, however, we only use pooled OLS with clustered standard error, as the degrees of freedom for the averaged difference in difference would be as low as 8 in this case.

Table 4.6: Pooled difference in difference on Tourist segment

	log(price)					
	2019 to 2021	1 day before 2020 vs 2021	2019 vs 2020	2019 to 2021	20 days before 2020 vs 2021	2019 vs 2020
Trip length	0.001*** (0.00000)	0.0003*** (0.00000)	0.001*** (0.00000)	-0.004*** (0.00000)	-0.003*** (0.00004)	-0.004*** (0.00001)
Treated	0.148*** (0.010)	0.145*** (0.021)	0.175*** (0.005)	-0.371*** (0.013)	-0.342*** (0.077)	-0.361*** (0.018)
Post-entry	0.083*** (0.011)	0.055** (0.018)	0.081*** (0.011)	0.271*** (0.024)	0.125 (0.092)	0.240*** (0.063)
DID	0.067*** (0.018)	0.031 (0.025)	0.040* (0.017)	0.114*** (0.025)	0.188* (0.096)	0.038 (0.063)
Constant	-2.115*** (0.006)	-1.947*** (0.015)	-2.196*** (0.003)	-0.568*** (0.012)	-0.746*** (0.061)	-0.575*** (0.015)
Route FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,350	860	915	1,304	813	893
R ²	0.836	0.831	0.852	0.787	0.768	0.711
Adjusted R ²	0.835	0.829	0.851	0.786	0.766	0.709
Residual Std. Error	0.068	0.067	0.063	0.126	0.115	0.133
F Statistic	974.022***	597.566***	745.988***	684.935***	380.848***	311.195***

This table reports the pooled difference in difference estimation on the tourist segment with clustered standard error at route level.
Note: *p<0.05; **p<0.01; ***p<0.001

Table 4.6 shows the result for the pooled OLS. Once again, “Trip length” is negative and significant - a 10 km longer route is priced roughly 1 to 4% less per km - and the constant term suggests that the unconditional expected mean is around 0.35 € and 0.22 € per km for one and twenty days before departure respectively. For prices one day before departure, we also observe that the treated group is generally 14% more expensive than the control group, while it is up to 37% cheaper for fares twenty days before departure. Once again, however, these coefficients have to be interpreted carefully because of the presence of route fixed effects. The time effect - coefficient “Post-entry” - is positive and significant in all regressions, showing a constant year-to-year increase in average prices, both last minute and advanced booking.

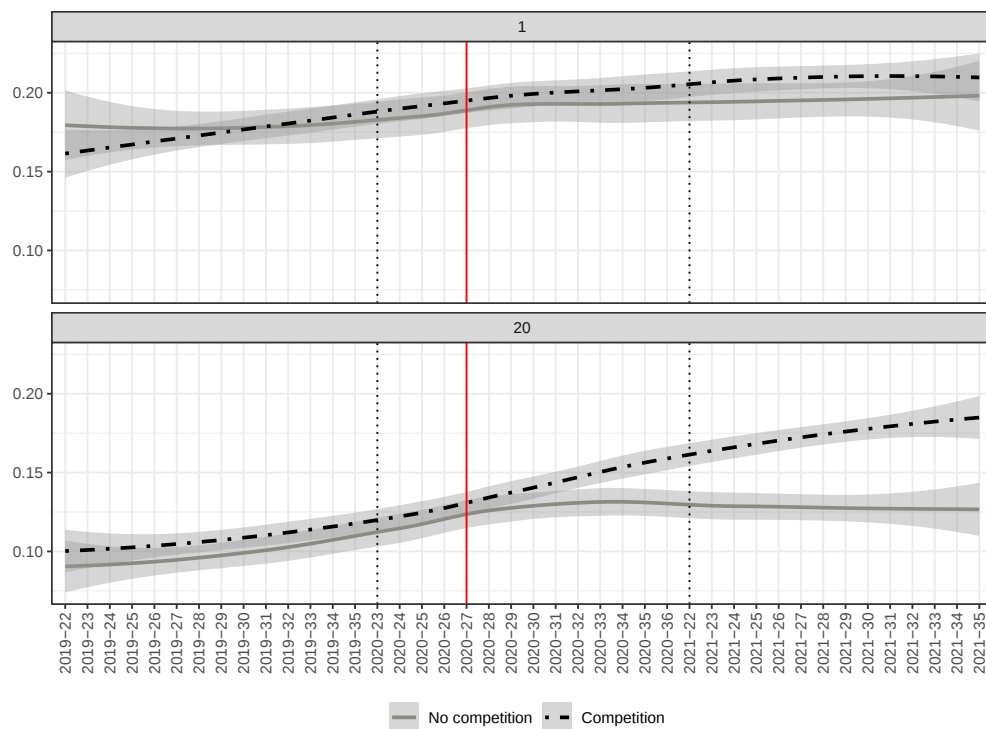
Unlike for the core segment, in these models, some treated-after-treatment coefficients (“DID”) are significant. For one day before departure, comparing the years 2019 to 2021 and 2019 to 2020 we find that routes, where competition started, were 7% and 4% more expensive after treatment than the untreated. No significant effect is to be found comparing 2020 and 2021, however the coefficient remains positive. For prices twenty days before the departure, the effect is also positive and significant however much higher in magnitude, around 11% and 18% for 2019 to 2021 and 2020 to 2021. The last column, 2019 vs 2020, shows an insignificant coefficient of smaller magnitude, 3%, however always positive.

Much as it cannot explain the causality implied in the regression, figure 4.6 helps us visu-

alise the price trends of the two groups. In the top panel, we see that both groups become more expensive over time but slightly more so the treated group. In the bottom panel, this pattern on the other hand is very visible and reflects also well the magnitude of the coefficient, with the treated group becoming increasingly more expensive after competition starts.

Within this segment therefore the effects of competition on prices are at best inexistent. In fact, prices are higher for both last-minute and early-booking for the treated group after entry, besides being anyway increasing year by year. The only improvement can be found in increased speed on these routes - the routes going North-South were moved on the HS tracks until Salerno - however this was well before Italo entered this segment. The findings in this segment in particular can be linked to the prediction made by the theoretical IO literature when liberalisation started. A possible explanation for the positive DID coefficient, in fact, is that where Italo entered, costs increased – possibly because of increased frequency and train service level – and thus also prices were pushed up, in a somehow unexpected direction for a competitive setting.

Figure 4.6: Routes in competition since July 2020 vs routes only operated by Trenitalia



Note: This figure shows the evolution of the price of routes where competition started in July 2020 versus those operated only by Trenitalia. Routes not in competition are Bari-Ancona, Milano-Pisa, Roma-Genova, routes in competition are Bologna-Ancona, Milano-Ancona, Milano-Rimini.

4.6 Mechanism of price increase

Focusing on summer 2018, we see that, in mid-July, there was a sudden increase in fares one day before departure, led primarily by Trenitalia but followed by Italo short after (see figure 4.7, top panel). Using data on the number of available partial flexibility tariffs (figure 4.7, middle panel) we can see that the increase is driven by a sudden decrease in the share of partial flexibility tariffs available one day before departure by both companies, which led to a generalised increase of the minimum tariff on the market. A possible explanation could be that all partial-flexibility combinations were sold out on that date, for a sudden increase in passengers for both companies, however the stability of the decrease suggests that this might not be the right explanation. By the end of summer 2019, in fact, neither Trenitalia nor Italo was offering discounted tariffs one day before departure anymore, which implies minimum tariff one day before departure coincides with full-flexibility tariffs and is therefore higher. At the same time, partial flexibility tariffs themselves increased. The bottom panel of figure 4.7 shows a decrease in the first half of July followed by an increase, mainly by Trenitalia but also by Italo. This also avails the hypothesis of an intentional increase in prices rather than a sudden and long-lasting increase in the number of passengers. A possible explanation is the one provided by Button (2003) in the context of the aviation industry. In his paper, he points out that the liberalisation of the aviation market, and therefore the presence of competition, induces overcapacity, which then in turn induces an unsustainable increase in investment to stay on the market. This could relate to the case of the Italian rail market, which indeed experienced an increase in capacity after liberalisation, alongside an increase in prices. Because of the lack of desegregated data on profits and passenger volume, however, it is not possible to draw a conclusion in this respect.

4.7 Conclusion

In this paper, we have examined the effects of competition in the rail market in the long run, in order to shed light on the true impact of the European liberalisation process, started in the early 2000s. We have done so by splitting our sample into two market segments and running a difference in difference model on each of them. We have done so to compare routes in competition since 2012 with those in competition since 2018, for the first segment, and to compare routes in competition since 2020 with those not in competition at all, for the second segment.

In the first case, we find that the presence of a competitor, for neither incumbent nor newcomer, contributes to keeping prices low. There is evidence of a generalised price decrease for discounted fares, to counterbalance the year-by-year increase in standard fares, however it is short-lived. In the second case, the effects are the opposite of what one

Figure 4.7: Share of partial flexibility tickets one day before departure



Note: This figure shows the evolution of the share of partial flexibility tickets available one day before departure in summer 2018.

would expect: prices are higher on routes where there is competition and frequency is unchanged. Also, prices are higher for both last-minute and early bookings. A pooled regression of the two segments (see table A.11 in appendix) reveals an insignificant difference between segments for one day before departure – we cannot reject the hypothesis the two DID coefficients are the same – while a significant difference 20 days before departure – the coefficient of the DID tourist segment remains positive and significant. Much as we do not have data on cost trends, these results point in the direction of what previously predicted in the literature: in the long-run, competition in the rail sector tends to raise prices.

Finally, we point out that there was a simultaneous increase in prices by both companies, which happened around the same time, i.e. in July 2018. Once again, this can be seen as something foreseen in the theory (Preston et al., 1999) however lack of transparency on data such as train load factor and desegregated profits do not allow us to draw a final conclusion on the reason for this joint behaviour.

We are aware that our work suffers from some limitations. The lack of accurate data on train load factors does not allow us to distinguish between pricing strategy and capacity effect on tariffs. Also, the lack of reliable desegregated data on companies' profits does not allow us to investigate the simultaneous price increase in 2018 more in-depth. Nonetheless, we believe our work shows the importance of policymakers being aware of the true effects of the policies they enact. Market liberalisation in the transport sector in fact has been at the top of the European agenda for more than a decade now and it is more and more common in several EU member states. However, the idea that this phenomenon would have reduced prices and, possibly, also increased modal share in favour of public transport has proven wrong. This paper shows that in the long run, a war on prices did not happen, but on the contrary prices – and probably costs – kept rising, and more so on routes in competition. This implies that liberalisation alone is not sufficient to enhance consumer benefits and ultimately shift people away from more polluting private transport modes. More ad-hoc policies accompanying the transition towards a more sustainable transport sector are necessary.

Chapter 5

General Conclusion

This dissertation presents three essays at the intersection of urban and transport economics. In the first two chapters, it shows that transport investments have an impact on within-city migration and gentrification, and on long-run dynamics of cities size. The third chapter tackles how can competition in the transport sector affect prices and service quality. More broadly, this dissertation presents relevant results on how transport investments have impacted and still affect our lives.

Chapter 2 examines concentration in Britain's urban system over more than two centuries. In particular, it analyses the city size-rank relationship (Zipf's Law) and the link between a city size and its subsequent growth (Gibrat's Law), as well the Gini coefficient as a summary measure of urban concentration. The estimation of the Zipf coefficient across 64 different combinations of city definitions and cutoff methods reveals that the empirical evidence on Zipf's law is heavily dependent on these factors, however under our preferred cutoff-definition combination we cannot reject Zipf's law. Both Gini and Gibrat's Law on the other hand show the same pattern over time: a trend towards more concentration in bigger cities between 1801 and 1861 and then the opposite trend thereafter, especially during 1861-1911 and 1951-1991. Both Zipf and Gini analyses suggest a pause in this fall in urban concentration between 1991-2011, the timing of which coincides with the concept of the consumer city, i.e. one based on centripetal forces relating to consumption, rather than production or employment.

The overall change in Pareto exponent is substantial and, where these results have wider relevance, it has significant implications where policymakers wish to understand the likely patterns of urban concentration over coming decades. While the drivers of these changes in urban concentration – including transport technology and policies relating to housing and industrial strategy – are beyond the scope of this chapter, the implications are substantial as policymakers seek to accommodate population growth and movements in the 21st century.

Chapter 3 investigates the impact of a new infrastructural investment at the extensive margin, in the context of a European capital city like Dublin. It calibrates a general equi-

librium model to obtain a measure of commuter market access, which captures the effect of the introduction of the *Luas* in Dublin. It then exploits an instrumental difference in difference specification to evaluate its impact on changes in the level of gentrification, population and rent. It finds a positive impact of the introduction of a new infrastructure on resident and workplace population shares, however it also points out that this migration is not homogeneous across higher and lower skilled workers. In fact the elasticity for the former to the change in market access is greater than that for the latter. It does not find any effect on long-term rent but future research is needed to analyse the effects on short-term market rent.

This chapter points out that investments in transport infrastructure have relevant externalities on cities dynamics and spacial sorting. In particular, in the case of Dublin, it shows there are distributional effects, concentrating higher-skilled cohorts, in residential terms, in areas that are better served by the new transport system, even as – from an employment perspective – lower-skilled groups benefited more. As a consequence, it is of great relevance to policymakers to understand these effects in details, upon decisions over major investments aimed at restructuring mobility. In a continuously urbanising world, this does not only helps understanding the pattern of city dynamics but it also poses questions in terms of additional policies necessary alongside infrastructural investments, to avoid exacerbating urban inequality. While a city's transport system is improved it is also extremely important to account for the consequences on cities spacial distribution.

Chapter 4 examines the particular case of the Italian rail market. It first uses a statistically solid method to identify two well-distinct market segments, with a set of identifying characteristics. It finds that there exists a core segment, where willingness to pay, speed, frequency and average prices are higher, and a tourist segment, where they are lower. Also, in the former segment, competition by Italo started very early on most routes while on the latter Italo entered the market only very recently. It then uses the market segments to inspect the effects of long and short run competition, both for last-minute tariffs and early booking fares. Results are not what one would expect from a liberalised market. The analysis in fact reveals a substantial generalised price increase on both segments, of both companies, over the five years. Also, while upon Italo's entry on the market in the core segment there is a reduction in early-booking fares, in the tourist segment entrance coincides with an increase in prices of routes in competition. Finally this chapter points out that there was a simultaneous increase in prices by both companies in July 2018 however lack of data on train load factor and desegregated profits do not allow to identify its drivers.

Overall, this chapter aims at resizing the narrative on the positive effects of competition in the rail sector. Its conclusion in fact is that, while prices increased substantially, the level of service remained the same, when it did not decrease. Clearly, this does not imply

that competition is useless or harmful altogether however it points out how the expected effects are not always the true ones and how important it is to monitor constantly the outcomes of policy choices. Also, it stresses that while liberalisation might be positive per-se, it is not sufficient to keep prices low, enhance consumer welfare and ultimately, increase passenger volume off other private, more polluting modes. If at the core of European Agenda is to decarbonise transport, mere rail liberalisation will not be enough to achieve this goal.

Bibliography

- Nicole Adler, Amir Brudner, and Stef Proost. A review of transport market modeling using game-theoretic principles. *European Journal of Operational Research*, 291(3): 808–829, 2021.
- Gabriel M Ahlfeldt, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf. The economics of density: Evidence from the Berlin wall. *Econometrica*, 83(6):2127–2189, 2015.
- Treb Allen and Costas Arkolakis. The welfare effects of transportation infrastructure improvements. Technical report, National Bureau of Economic Research, 2019.
- Treb Allen, Costas Arkolakis, and Xiangliang Li. Optimal city structure. *Working Paper*, 2016.
- Milena Almagro and Tomás Domínguez-Iino. Location sorting and endogenous amenities: Evidence from Amsterdam. *Working Paper*, 2019.
- Gordon Anderson and Ying Ge. The size distribution of chinese cities. *Regional Science and Urban Economics*, 35(6):756–776, 2005.
- Sidra Arshad, Shougeng Hu, and Badar Nadeem Ashraf. Zipf’s law, the coherence of the urban system and city size distribution: Evidence from Pakistan. *Physica A: Statistical Mechanics and its Applications*, 513:87–103, 2019.
- Seyed Peyman Asadi. Empirical investigation of city size distribution in Iran. *Working Paper*, 2019.
- J. Auerbach. Das gesetz der bevoelkerungskonzentration. *Petermanns Geographische Mitteilungen*, LIX:73–76, 1913.
- David H. Autor, Christopher J Palmer, and Parag A Pathak. Gentrification and the amenity value of crime reductions: Evidence from rent deregulation. Technical report, National Bureau of Economic Research, 2017.
- P. Bairoch, J. Batou, and P. Chevre. The population of European cities. Data bank and short summary of results: 800-1850. 1988.
- Pankaj Bajracharya and Salima Sultana. Rank-size distribution of cities and municipalities in Bangladesh. *Sustainability*, 12(11):4643, 2020.
- Clare Balboni, Gharad Bryan, Melanie Morten, and Bilal Siddiqi. Transportation, gentrification, and urban mobility: The inequality effects of place-based policies. *Working Paper*, 2020.

- Eleni Bardaka, Michael S Delgado, and Raymond JGM Florax. Causal identification of transit-induced gentrification and spatial spillover effects: The case of the Denver light rail. *Journal of Transport Geography*, 71:15–31, 2018.
- M. Batty. Rank clocks. *Nature*, 444(7119):592–596, 2006.
- Nathaniel Baum-Snow. Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805, 2007.
- Nathaniel Baum-Snow, Matthew E Kahn, and Richard Voith. Effects of urban rail transit expansions: Evidence from sixteen cities, 1970-2000. *Brookings-Wharton papers on urban affairs*, pages 147–206, 2005.
- Nathaniel Baum-Snow, Loren Brandt, J Vernon Henderson, Matthew A Turner, and Qinghua Zhang. Roads, railroads, and decentralization of chinese cities. *Review of Economics and Statistics*, 99(3):435–448, 2017.
- Nathaniel Baum-Snow, Matthew Freedman, and Ronni Pavan. Why has urban inequality increased? *American Economic Journal: Applied Economics*, 10(4):1–42, 2018.
- Angela S Bergantino, Claudia Capozza, and Mauro Capurso. The impact of open access on intra-and inter-modal rail competition. A national level analysis in Italy. *Transport Policy*, 39:77–86, 2015.
- Angela S Bergantino, Claudia Capozza, and Mauro Capurso. Pricing strategies: who leads and who follows in the air and rail passenger markets in Italy. *Applied Economics*, 50(46):4937–4953, 2018.
- P Beria, Y Crozet, L Guihéry, et al. Transport ferroviaire de passagers: la concurrence «sur le marché» s’installe en Europe. *Transports, Infrastructures & Mobilité (TI&M)*, 533:37–46, 2022a.
- Paolo Beria and Alberto Bertolin. Evolving long-distance passenger services. market concentration, fares and specialisation patterns in Italy. *Research in Transportation Economics*, 74:77–92, 2019.
- Paolo Beria and Raffaele Grimaldi. Reality and opportunities for on-track competition in HSR. In *Evaluating High-Speed Rail*, pages 156–174. Routledge, 2016.
- Paolo Beria, Renato Redondi, and Paolo Malighetti. The effect of open access competition on average rail prices. The case of Milan–Ancona. *Journal of Rail Transport Planning & Management*, 6(3):271–283, 2016.
- Paolo Beria, Dario Nistri, and Antonio Laurino. Intercity coach liberalisation in Italy: Fares determinants in an evolving market. *Research in Transportation Economics*, 69: 260–269, 2018.
- Paolo Beria, Samuel Tolentino, Alberto Bertolin, and Gabriele Filippini. Long-distance rail prices in a competitive market. Evidence from head-on competition in Italy. *Journal of Rail Transport Planning & Management*, 12:100144, 2019.
- Paolo Beria, Samuel Tolentino, and Gabriele Filippini. Are prices reduced from direct competition in high-speed rail? some unexpected evidences from Italy. *RE.PUBLIC@POLIMI pubblicazioni di ricerca del Politecnico di Milano*, 2020.

- Paolo Beria, Vardhman Lunkar, Samuel Tolentino, Vilém Pařil, and Michal Kvasnička. The seeds of competition in long-distance rail transport. comparing the early open-access countries. *Working Paper*, 2022b.
- Paolo Beria, Samuel Tolentino, Evgeniia Shtele, and Vardhman Lunkar. A difference-in-difference approach to estimate the price effect of market entry in high-speed rail. *Competition and Regulation in Network Industries*, 2022c.
- Brian JL Berry and Adam Okulicz-Kozaryn. The city size distribution debate: Resolution for US urban regions and megalopolitan areas. *Cities*, 29:S17–S23, 2012.
- Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1): 249–275, 2004.
- Luís MA Bettencourt, José Lobo, and Geoffrey B West. Why are large cities faster? Universal scaling and self-similarity in urban organization and dynamics. *The European Physical Journal B*, 63(3):285–293, 2008.
- Stephen B Billings. Estimating the value of a new transit option. *Regional Science and Urban Economics*, 41(6):525–536, 2011.
- John D Bitzan. Railroad costs and competition: The implications of introducing competition to railroad networks. *Journal of Transport Economics and Policy (JTEP)*, 37(2): 201–225, 2003.
- Duncan Black and Vernon Henderson. Urban evolution in the USA. *Journal of Economic Geography*, 3(4):343–372, 2003.
- Thierry Blayac and Patrice Bougette. Should I go by bus? The liberalization of the long-distance bus industry in France. *Transport Policy*, 56:50–62, 2017.
- Maarten Bosker, Steven Brakman, Harry Garretsen, and Marc Schramm. A century of shocks: the evolution of the German city size distribution 1925–1999. *Regional Science and Urban Economics*, 38(4):330–347, 2008.
- Kenneth Button. Does the theory of the “core” explain why airlines fail to cover their long-run costs of capital? *Journal of Air Transport Management*, 9(1):5–14, 2003.
- Ennio Cascetta and Pierluigi Coppola. New high-speed rail lines and market competition: Short-term effects on services and demand in Italy. *Transportation Research Record*, 2475(1):8–15, 2015.
- Centre for Cities. Primary Urban Areas: City definition, 2021. URL <https://www.centreforcities.org/city-by-city/puas/>.
- Juan Pablo Chauvin, Edward Glaeser, Yueran Ma, and Kristina Tobio. What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics*, 98:17–49, 2017.
- P. Cheshire. Trends in sizes and structures of urban areas. *Handbook of Regional and Urban Economics*, 3:1339–1373, 1999.
- Panayotis Christidis and Juan Nicolas Ibanez Rivas. Measuring road congestion. *JRC Scientific and Policy Reports*, 2012.

- A. Clauset, C. R. Shalizi, and M. E. Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
- CNIT. Introduction to Italo. Company profile. October 2019. *Conto Nazionale delle Infrastrutture e della Mobilità Sostenibili*, 2020.
- Victor Couture, Cecile Gaubert, Jessie Handbury, and Erik Hurst. Income growth and the distributional effects of urban spatial sorting. Technical report, National Bureau of Economic Research, 2019.
- J. C. Córdoba. A generalized Gibrat’s law. *International Economic Review*, 49, 2008.
- Federica Daniele and Elena Romito. When building one more metro line is not enough: evidence from the “Metro C” in Rome. *Working Paper*, 2022.
- D. R. Davis and D. E. Weinstein. Bones, bombs, and break points: the geography of economic activity. *The American Economic Review*, 92(5):1269–1289, 2002.
- Ethan H Decker, Andrew J Kerkhoff, and Melanie E Moses. Global patterns of city size distributions and their fundamental drivers. *PLoS One*, 2(9):e934, 2007.
- Klaus Desmet and Marcel Fafchamps. Employment concentration across US counties. *Regional Science and Urban Economics*, 36(4):482–509, 2006.
- Klaus Desmet and Jordan Rappaport. The settlement of the United States, 1800–2000: the long transition towards gibrat’s law. *Journal of Urban Economics*, 98:50–68, 2017.
- J. I. Dingel, A. Miscio, and D. R. Davis. Cities, lights, and skills in developing economies. *Journal of Urban Economics*, 31(6):103–174, 2019.
- J. Dittmar. Cities, markets, and growth: The emergence of Zipf’s law. *Explorations in Economic History*, 2020.
- L. Dobkins and Y. Ioannides. Spatial interactions among us cities: 1900-1990. *Regional science and urban Economics*, 31(6):701–731, 2001.
- L. Dobkins, Y. Ioannides, et al. Dynamic evolution of the US city size distribution. *The Economics of Cities*, pages 217–260, 2000.
- Kacie Dragan, Ingrid Ellen, and Sherry A Glied. Does gentrification displace poor children? New evidence from New York City Medicaid data. Technical report, National Bureau of Economic Research, 2019.
- Gilles Duranton. Urban evolutions: The fast, the slow, and the still. *American Economic Review*, 97(1):197–221, 2007.
- Gilles Duranton and Matthew A Turner. Urban growth and transportation. *Review of Economic Studies*, 79(4):1407–1440, 2012.
- J. Eaton and Z. Eckstein. Cities and growth: Theory and evidence from France and Japan. *Regional science and urban Economics*, 27(4):443–474, 1997.
- EC. The future development of the common transport policy: A global approach to the construction of a community framework for sustainable mobility. *Commission of the European Communities.*, 1992.

- J. Eeckhout. Gibrat's law for (all) cities. *American Economic Review*, 94(5):1429–1451, 2004.
- Michail Fragkias and Karen C Seto. Evolving rank-size distributions of intra-metropolitan urban clusters in South China. *Computers, Environment and Urban Systems*, 33(3): 189–199, 2009.
- Oskar Fröidh and Bo-Lennart Nelldal. The impact of market opening on the supply of interregional train services. *Journal of Transport Geography*, 46:189–200, 2015.
- X. Gabaix. Zipf's law for cities: an explanation. *The Quarterly Journal of Economics*, 114(3):739–767, 1999a.
- X. Gabaix. Zipf's law and the growth of cities. *The American Economic Review*, 89(2): 129–132, 1999b.
- X. Gabaix. Power laws in economics and finance. *Annual Review of Economics*, 1(1): 255–294, 2009.
- X. Gabaix and R. Ibragimov. Rank- $1/2$: a simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics*, 29(1):24–39, 2011.
- X. Gabaix and Y. M. Ioannides. The evolution of city size distributions. *Handbook of Regional and Urban Economics*, 4:2341–2378, 2004.
- Alberto A Gaggero and Claudio A Piga. Airline market power and intertemporal price dispersion. *The Journal of Industrial Economics*, 59(4):552–577, 2011.
- Miquel-Àngel Garcia-López. Urban spatial structure, suburbanization and transportation in barcelona. *Journal of Urban Economics*, 72(2-3):176–190, 2012.
- Miquel-Àngel Garcia-López, Elisabet Viladecans-Marsal, and Adelheid Holl. Suburbanization and highways: When the romans, the bourbons and the first cars still shape Spanish cities. 2013.
- Kristopher S Gerardi and Adam Hale Shapiro. Does competition reduce price dispersion? New evidence from the airline industry. *Journal of Political Economy*, 117(1):1–37, 2009.
- Stephen Gibbons and Stephen Machin. Valuing rail access using transport innovations. *Journal of Urban Economics*, 57(1):148–169, 2005.
- Robert Gibrat. *Les inégalités économiques*. Paris : Recueil Sirey, 1931.
- Kristian Giesen and Jens Südekum. Zipf's law for cities in the regions and the country. *Journal of Economic Geography*, 11(4):667–686, 2011.
- Kristian Giesen, Arndt Zimmermann, and Jens Suedekum. The size distribution across all cities—double pareto lognormal strikes. *Journal of Urban Economics*, 68(2):129–137, 2010.
- Francisco J Goerlich Gisbert and Matilde Mas. La distribución empírica del tamaño de las ciudades en españa, 1900-2001.¿ quién verifica la ley de Zipf? *Revista de Economía Aplicada*, 18(54):133–159, 2010.

- Andrea Giuricin. Ownership change heralds expansion at Italo-NTV. *Railway Gazette International*, 2018.
- Edward L Glaeser, Jed Kolko, and Albert Saiz. Consumer city. *Journal of Economic Geography*, 1(1):27–50, 2001.
- Edward L Glaeser, Matthew E Kahn, and Jordan Rappaport. Why do the poor live in cities? the role of public transportation. *Journal of Urban Economics*, 63(1):1–24, 2008.
- Edward L Glaeser, Giacomo AM Ponzetto, and Kristina Tobio. Cities, skills and regional change. *Regional Studies*, 48(1):7–43, 2014.
- Jose A Gomez-Ibanez. Big-city transit rider snip, deficits, and politics: Avoiding reality in Boston. *Journal of the American Planning Association*, 62(1):30–50, 1996.
- Marco Gonzalez-Navarro and Matthew A Turner. Subways and urban growth: Evidence from earth. *Journal of Urban Economics*, 108:85–106, 2018.
- Rafael González-Val and Luis Lanaspá. Patterns in US urban growth, 1790–2000. *Regional Studies*, 50(2):289–309, 2016.
- Rafael González-Val and Javier Silvestre. An annual estimate of spatially disaggregated populations: Spain, 1900–2011. *The Annals of Regional Science*, pages 1–18, 2020.
- Rafael González-Val, Luis Lanaspá, and Fernando Sanz-Gracia. New evidence on gibrat’s law for cities. *Urban Studies*, 51(1):93–115, 2014.
- Peter Gordon and Richard Willson. The determinants of light-rail transit demand—an international cross-sectional comparison. *Transportation Research Part A: General*, 18(2):135–140, 1984.
- Laurent Guihéry. Autocars interurbains en Allemagne: pour Flixbus, le temps des interrogations. *TI&M, Transports, Infrastructures & Mobilité*, (521):p–33, 2020.
- Stephan Heblich, Stephen J Redding, and Daniel M Sturm. The making of the modern metropolis: evidence from London. *The Quarterly Journal of Economics*, 135(4): 2059–2133, 2020.
- J Vernon Henderson and Hyoung Gun Wang. Urbanization and city growth: The role of institutions. *Regional Science and Urban Economics*, 37(3):283–313, 2007.
- Inrix. Report here. 2021.
- Yannis Ioannides and Henry G. Overman. Zipf’s law for cities: an empirical examination. *Regional Science and Urban Economics*, 33(2):127–137, 2003.
- Yannis Ioannides and Junfu Zhang. Walled cities in late imperial China. *Journal of Urban Economics*, 97:71–88, 2017.
- Bin Jiang, Junjun Yin, and Qingling Liu. Zipf’s law for all the natural cities around the world. *International Journal of Geographical Information Science*, 29(3):498–522, 2015.

- Daniel Johnson and Chris Nash. Competition and the provision of rail passenger services: A simulation exercise. *Journal of Rail Transport Planning & Management*, 2(1-2):14–22, 2012.
- Alexander Klein and Tim Leunig. Gibrat’s law and the British industrial revolution, 2015.
- Paul Krugman. Confronting the urban mystery. *Journal of the Japanese and International Economies*, 10(4):399–418, 1996.
- Rainer Kujala, Christoffer Weckström, Richard K Darst, Miloš N Mladenović, and Jari Saramäki. A collection of public transport network data sets for 25 cities. *Scientific data*, 5(1):1–14, 2018.
- Luis Lanaspa, Fernando Pueyo, and Fernando Sanz. The evolution of Spanish urban structure during the twentieth century. *Urban Studies*, 40(3):567–580, 2003.
- Julie Le Gallo and Coro Chasco. Spatial analysis of urban growth in Spain, 1900–2001. *Empirical economics*, 34(1):59–80, 2008.
- Legambiente. Rapporto pendolaria 2021. 2021.
- M. Levy. Gibrat’s law for (all) cities: Comment. *American Economic Review*, 99(4):1672–75, 2009.
- Valente J Matlaba, Mark J Holmes, Philip McCann, and Jacques Poot. A century of the evolution of the urban system in Brazil. In *Review of Urban & Regional Development Studies: Journal of the Applied Regional Science Conference*, volume 25, pages 129–151. Wiley Online Library, 2013.
- Daniel McFadden et al. Conditional logit analysis of qualitative choice behavior. 1973.
- Guy Michaels, Ferdinand Rauch, and Stephen J Redding. Urbanization and structural transformation. *The Quarterly Journal of Economics*, 127(2):535–586, 2012.
- N. Z. Muller. Power laws and air pollution. *Environmental Modeling & Assessment*, 21(1):31–52, 2016.
- M.E. Newman. Power laws, pareto distributions and Zipf’s law. *Contemporary Physics*, 2005.
- Volker Nitsch. Zipf zipped. *Journal of Urban Economics*, 57(1):86–100, 2005.
- NTA. National transport authority. URL <https://www.nationaltransport.ie/>.
- NTV. Introduction to Italo. Company Profile. October 2019. 2019.
- Rahul Pathak, Christopher K Wyczalkowski, and Xi Huang. Public transit access and the changing spatial distribution of poverty. *Regional Science and Urban Economics*, 66:198–212, 2017.
- John Preston, Gerard Whelan, and Mark Wardman. An analysis of the potential for on-track competition in the British passenger rail industry. *Journal of transport economics and policy*, pages 77–94, 1999.
- Julius Probst. Agglomeration forces: A study of Swedish cities and municipalities since 1800, 2017.

- Stephen J Redding and Matthew A Turner. Transportation costs and the spatial organization of economic activity. *Handbook of Regional and Urban Economics*, 5:1339–1398, 2015.
- Stijn Ronsse and Samuel Standaert. Combining growth and level data: An estimation of the population of belgian municipalities between 1880 and 1970. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 50(4):218–226, 2017.
- K. T. Rosen and M. Resnick. The size distribution of cities: an examination of the pareto law and primacy. *Journal of Urban Economics*, 8(2):165–186, 1980.
- E. Rossi-Hansberg and M. L. Wright. Establishment size dynamics in the aggregate economy. *The American Economic Review*, 97(5):1639–1666, 2007.
- H. D. Rozenfeld, D. Rybski, X. Gabaix, and H. A. Makse. The area and population of cities: New insights from a different perspective on cities. *American Economic Review*, 101(5):2205–25, 2011.
- Aurora Ruiz-Rúa and Roberto Palacín. Towards a liberalised European high speed railway sector: Analysis and modelling of competition using game theory. *European Transport Research Review*, 5:53–63, 2013.
- Alexandra Schaffar and Michel Dimou. Rank-size city dynamics in China and India, 1981–2004. *Regional Studies*, 46(6):707–721, 2012.
- Christopher Severen. Commuting, labor, and housing market effects of mass transportation: Welfare and identification. *The Review of Economics and Statistics*, pages 1–99, 2019.
- Shalini Sharma. Persistence and stability in city growth. *Journal of Urban Economics*, 53(2):300–320, 2003.
- Christopher Small and Christopher D Elvidge. Night on earth: Mapping decadal changes of anthropogenic night light in asia. *International Journal of Applied Earth Observation and Geoinformation*, 22:40–52, 2013.
- Christopher Small, Christopher D Elvidge, Deborah Balk, and Mark Montgomery. Spatial scaling of stable night lights. *Remote Sensing of Environment*, 115(2):269–280, 2011.
- K. T. Soo. Zipf’s law for cities: a cross-country investigation. *Regional Science and Urban Economics*, 35(3):239–263, 2005.
- Kwok Tong Soo. Zipf’s law and urban growth in malaysia. *Urban Studies*, 44(1):1–14, 2007.
- Kwok Tong Soo. Zipf, Gibrat and geography: Evidence from China, India and Brazil. *Papers in Regional Science*, 93(1):159–181, 2014.
- Humphrey Southall. A vision of Britain through time, 2017. URL <https://www.visionofbritain.org.uk/>.
- Yichen Su. The rising value of time and the origin of urban gentrification. *American Economic Journal: Economic Policy*, 14(1):402–39, 2022.
- J. Sutton. Gibrat’s legacy. *Journal of Economic Literature*, XXXV:40–59, 1997.

- Zdeněk Tomeš and Monika Jandová. Open access passenger rail services in central Europe. *Research in Transportation Economics*, 72:74–81, 2018.
- Nick Tsivanidis. Evaluating the impact of urban transit infrastructure: Evidence from Bogota’s transmilenio. *Accepted American Economic Review*, 2019.
- UK Houses of Parliament. The Bournemouth, Dorset and Poole (Structural Changes) Order 2018. UK Statutory Instruments, 2018 No. 648, PART 1, Article 2, 2018. URL <https://www.legislation.gov.uk/ukSI/2018/648/article/2/made>.
- UN-DESA. World urbanization prospects: The 2018 revision, 2018.
- Gerson Javier Pérez Valbuena and Adolfo Meisel Roca. Ley de Zipf y de Gibrat para Colombia y sus regiones: 1835-2005. *Revista de Historia Económica-Journal of Iberian and Latin American Economic History*, 32(2):247–286, 2014.
- Andreas Vigren. Competition in Swedish passenger railway: Entry in an open access market and its effect on prices. *Economics of Transportation*, 11:49–59, 2017.
- Pablo Ernesto Warnes. Transport infrastructure improvements and spatial sorting: Evidence from Buenos Aires. *Working Paper*, 2020.
- G Whelan and D Johnson. Using praise to model competition on passenger rail networks. *First Conference on Railroad Industry Structure, Competition and Investment. Institut D’Economie Industrielle, Toulouse, France*, 2003.
- George K Zipf. Human behavior and the principle of least effort. Cambridge, Mass., 1949. *MK Petrov’s non-linear thinking model as social sciences gnoseology crisis overcoming project*, 1965.

Appendices

A.1 The Rise & Fall of Urban Concentration in Britain: Zipf, Gibrat and Gini across two centuries Appendix

A.1 Literature

In this section, we outline the findings of the literature review summarized briefly in Section 2.2.3 above. In total, we review almost fifty published empirical analyses and to aid exposition, we structure our review in two parts – focusing first on shorter-run studies (either cross-sectional or less than a half-century), which typically test Zipf’s Law, before turning to longer-run studies, which are more likely to focus on Gibrat’s Law. In the former, we pay particular attention to the definition of city used, as this has implications for the pattern of findings, and in the latter, we group papers by region. Ahead of our review, we highlight that papers published prior to Gabaix and Ioannides (2004) will use naive standard errors, while papers published prior to Gabaix and Ibragimov (2011) will not correct for the bias in the OLS estimator, although some will use the Hill estimator.

A.1.1 Short-run analyses

We define the “short-run” literature to be empirical analyses using less than 50 years of data on urban populations and review a range papers that meet these criteria, distinguishing where possible between three sets of analyses, reflecting the choice of urban unit. The first uses officially-defined municipalities to delineate cities, the second uses metropolitan areas, while the third uses definitions of cities built from satellite imagery. We start with two seminal papers on the topic and a 2005 meta-analysis.

Two seminal cross-countries studies are Rosen and Resnick (1980) and Soo (2005). Both assemble datasets that cover a larger number of countries (44 and 73 respectively), using data from the late 20th century. In both cases, they find that, while estimates vary substantially by country and definition of city, Zipf’s Law is far less likely to be rejected when urban agglomerations are used instead of administrative units. Rosen and Resnick (1980) do not employ any bias or error correction; Soo (2005) corrects for bias but not for standard errors, which may affect his results more generally.

The literature to 2002 is reviewed comprehensively in a meta-analysis by Nitsch (2005). They review 515 estimates of the Zipf coefficient from 29 different studies published between 1925 and 2002. Overall, they find that the Zipf coefficient turns out to be significantly larger than 1 (around 1.1), on average, implying that cities are more evenly distributed than Zipf’s Law would predict. Estimates closer to 1 are mostly found in studies for the period post-1900 (average is 1.03 for the period after 1901 - 1950) and for metropolitan areas rather than municipalities. In particular, in the US, one of the largest urban systems, Zipf’s Law holds for metropolitan areas. These three papers together already highlight the importance of city definition: where metro areas, not municipalities, are used, it is hard to reject Zipf’s Law.

Nonetheless, additions to the literature continue to use municipalities as unit of analysis. This includes Giesen et al. (2010) and Giesen and Südekum (2011). In the former, the authors examine a sample of eight countries and find evidence that the double Pareto lognormal distribution provides a better fit than the simple lognormal and Pareto distributions, at odds with Zipf’s Law. However, in the latter, the authors use administratively-defined cities within German regions 1975-1997. Applying a cutoff of 100,000 and a corrected OLS estimation, they find that Zipf’s law holds *within regions*. Two other papers using municipalities reject Zipf’s Law. One is Schaffar and Dimou (2012), who study the evolution of Chinese and Indian cities above 100,000 during the period 1981-2004. While

they find the city size-rank distribution is Pareto, and cannot reject Gibrat's Law, they conclude that Zipf's Law is systematically violated for both countries, with the exponent greater than one (and varying over time). Using Iranian data 2006-2016 and cutoffs of between 20,000 and 100,000, Asadi (2019) finds a coefficient higher than predicted by Zipf's Law in the truncated sample.

Eeckhout (2004) employs a cross-section of over 25,000 "places" in the US, officially designated so either by state law or by the federal Census in 2000. He uses an OLS estimator on different cutoff levels, from 42 to 155,000, and shows that the Pareto exponent is very sensitive to the choice of the cutoff level. He then shows that a log-normal distribution, rather than a power law, best fits. In a responding paper, Levy (2009) establishes that, for the largest 150 cities – home to almost one quarter of the US population – the relationship is unequivocally described by a power law. This is relevant where the objects of interest are the largest urban units. We take the key findings from both papers – the importance of cut-offs and the power-law distribution of larger urban units – to our analysis.

Five papers in our review use cities as defined by functional metropolitan areas, rather than administrative municipalities. Using 1970s-1990s data for the USA and France, Duranton (2007) links city size distribution with industrial turnover and finds that second-nature industries prevent small cities from disappearing. His model suggests that the steady-state Zipf's curve is concave, with coefficient below 1 in lower tail and above 1 in upper tail. Bettencourt et al. (2008) analyze labor markets in China, US and Europe and find that the processes relating urbanization to economic development and knowledge creation are common to all big cities belonging to the same urban system. They observe that wealth creation and innovation are associated with an exponent greater than one, while infrastructure is associated with an exponent below one.

Chauvin et al. (2017) compare the US, Brazil, India and China 2000-2010. They find that both Gibrat's Law and Zipf's Law hold in Brazil and the USA, but not in China and India, a finding that is suggestive of spatial equilibrium emerging with economic development. Decker et al. (2007) examine global data on cities, using a combination of metropolitan areas and cities defined as night-light clusters, a method discussed further below. They conclude, in line with Eeckhout (2004), that the full distribution of cities is best fit by a log-normal distribution, while the Pareto distribution only emerges in the upper tail. Similarly, Bajracharya and Sultana (2020) combine both official data on metro areas with cities defined by disaggregated spatial data, in their case the street network. Using the case of Bangladesh 1991-2019, and applying the corrected OLS method, they find that Zipf's Law does not hold: instead, it is smaller when all municipalities are considered and concave at the upper end of the distribution.

We turn, lastly, to the newer literature using cities defined by satellite imagery. An early contribution is by Fragkias and Seto (2009), who use contiguous urban built-up areas in three parts of South China's Pearl River Delta, 1988-1999. They find that urban clusters in metropolitan areas do follow a power law distribution but its parameters oscillate over-time. However, their definition of city is likely very small, as their analyses includes over 5,000 built-up areas across three regions with a population of 21 million.

Rozenfeld et al. (2011) define cities as "maximally connected cluster of populated sites defined at high resolution". They use this new measure on US and British populations, for 2001 and 1981 respectively, and find that the size-rank relationship is well described by Zipf's Law. Small et al. (2011) employ a similar approach at a global scale; specifically, they use spatially contiguous patches of stable night light over a range of brightnesses corresponding to different intensities of anthropogenic development. Using both OLS and MLE and different brightness cutoff levels, they find Power law exponents in the range 0.95 to 1.11, with the estimated slope varying by brightness cutoff. Overall, they

also conclude that Zipf's Law holds for a wide range of developed land areas at both continental and global scales.

In related research, Small and Elvidge (2013) examine night-light patterns in China and India 1992-2005. They find that the size distributions of "lighted cities" are consistent with power laws with exponents near -1. The larger lighted segments are closer to spatial networks of contiguous development than individual cities. This finding is consistent with two subsequent analyses using similar data. Using "natural cities" globally extracted from satellite imagery for 1992, 2001 and 2010, Jiang et al. (2015) test for the presence of Zipf's Law. Using an OLS estimator, they are largely unable to reject Zipf's Law, especially at continental and global levels. Exceptions include Africa, in certain periods, and at country level, where Zipf's Law is violated in certain countries and periods. Dingel et al. (2019) construct lights-based metropolitan areas for US, Brazil, China and India and, similarly, are unable to reject Zipf's Law – but would if they had used administratively defined cities. In the case of US and Brazil, their distribution mirrors the distribution of commuting-based definitions of cities.

A.1.2 Long-run analyses

The second part of our literature review cover empirical analyses of city growth and size distributions in the long-run, which we define to be over a period of at least four decades. We review these below, separating into four broad categories based on their (principal) region of analysis: global studies; the Americas; Asia; and Europe.

Global Henderson and Wang (2007) assemble data on 1,644 cities (usually metropolitan areas, although especially earlier these definitions can vary) in 142 countries over the period 1960-2000. They employ a cutoff of 100,000, which they say reflects national definitions, and eschew the estimation of a Zipf component, preferring instead to analyse the "spatial Gini" coefficients of population inequality at national level. They highlight the importance of institutional variables, with planned economies associated with a smaller Gini coefficient, as are federal and democratic political systems. They also document the importance of "new" cities (i.e. those that grow above the cut-off of 100,000) in driving urban population growth. Related, they find no evidence of concentration into so-called mega-cities in the period under analysis.

Soo (2014) examines the Zipf coefficient in three of the world's most populous countries – Brazil, China and India – between 1950 and 2000. He uses population by sub-national unit (such as region or province) rather than city (however defined), but does employ the GI corrections for estimator and standard error. None of the three countries has more than 30 sub-national units, meaning that standard errors on the Zipf coefficient are large. As a result, he is unsurprisingly unable to reject Zipf's Law in any year for either Brazil or China – in India, Zipf's Law is rejected where at least 20 sub-national units (of 27) are included, with the coefficient being closer to zero.

The Americas Ten papers in our review examine long-run city growth dynamics in the USA, including Rose (2006), who combines both global and USA perspectives. Specifically, he compares the rank-size relationship between cities (MSAs) within the USA and between the 50 largest countries, over the period 1900-2004 and to 2050 using population projections. Using adjusted standard errors but unadjusted OLS estimators, he finds that Zipf's Law holds both within the USA and across countries - with the coefficient for the 50 largest countries rising from -0.78 to -0.99 between 1900 and 2050.

Five papers focus on US city growth during the 20th century, typically 1900-1990 and using Metropolitan Statistical Areas (MSAs). Dobkins et al. (2000) reject the null hypothesis of parallel growth of cities and find evidence that the Pareto exponent has been decreasing (in absolute value) over time, implying increased concentration towards the upper end of the distribution; Dobkins and Ioannides (2001) add to this that in cities with neighbours, growth rates are closely interdependent. However, using a non-parametric approach, Ioannides and Overman (2003) find evidence in favor of both Zipf's Law and Gibrat's law: for MSAs, it is not possible to reject the hypotheses that the first two moments of MSA growth, as well as the Pareto exponent, are invariant to city size. Similarly, Black and Henderson (2003) document a stable size distribution and transition process, with bigger cities exhibiting minimal downward mobility. Using a later end date (2010), a higher cutoff (500,000 rather than 50,000), and BEA "Economic Areas", Berry and Okulicz-Kozaryn (2012) find that, when urban regions are properly defined, US urban growth obeys both Gibrat's and Zipf's Law. They conclude that conflict in the literature is a consequence of choice of units of observation.

Four papers analyze city growth in the US over longer horizons, in most instances using Census data that start in 1790. Batty (2006) examines the population of the 100 largest US cities since 1790, together with data for the UK (1901-2001) and globally, using the Chandler dataset from 430BC. By documenting evidence of deviations from growth by proportionate effect, he concludes that rank-size scaling is far from universal, with "micro-level" dynamics of cities rising and falling over time an important aspect to consider. This is consistent with Glaeser et al. (2014), who explore the dynamics of county growth in eastern and central USA over the period 1860-2000, with some data extending back to 1790. While they find evidence in favour of Gibrat's Law for the sample as a whole, this does not hold for long sub-periods. Before 1860 and after 1970, less populous counties grew more quickly, while between, population growth was regularly faster in more populated areas. While one interpretation is that Gibrat's law is universal only over sufficiently long time periods, another is that Gibrat's law is an artefact of the accidental balancing of centripetal forces, which dominated during the industrial era, and centrifugal forces before and after.

González-Val and Lanaspá (2016) analyze the populations of 190 incorporated places in the USA, 1790-2000, their sample reflecting a population cut-off of 100,000 in the year 2000. They find mixed evidence regarding long-run city growth. On the one hand, the unit root hypothesis underpinning random growth cannot be rejected in most specifications: growth does not depend on initial size. However, there is strong evidence in favour of conditional convergence in growth rates within "clubs", suggestive of "local" mean-reversion within size bins. Lastly, Michaels et al. (2012) investigate the nature of population growth and extend their analysis of the USA, from 1800 to 2000, to include rural areas; they compare this with Brazil 1970-2000. They establish six stylized facts about the dynamics of population and employment, when rural areas are added to the picture, one of which is that Gibrat's Law fails when rural areas are included: agricultural employment growth appears to be decreasing in the initial population density, while in urban settings, employment growth is uncorrelated to initial conditions.

Matlaba et al. (2013) also study Brazil, using a dataset of 185 functionally defined urban areas 1907-2008 and the GI method of estimation. Their principal finding is that the power parameter of the size distribution of the 100 largest urban areas grows over time, approaching unity: Zipf's and Gibrat's Laws became steadily more appropriate descriptions of Brazil's city size distribution during the 20th century. Valbuena and Roca (2014) examine Colombian municipalities that are home to 50% of the country's population, 1835-2005. Using the adjusted rank-size relationship and non-parametric

techniques, they are unable to reject both Zipf's law and, from the mid-20th century, Gibrat's law. Their results are consistent with changes in the drivers of Columbia's population growth at both national and regional levels from the 1950s.

Asia In addition to Soo (2014) mentioned above, other researchers have examined the city size-rank relationship in China and India over the long run. Anderson and Ge (2005) analyze China during the period 1949-1999, employing maximum likelihood estimation (MLE) on municipalities and prefecture-level cities with a population of greater than 100,000. They find evidence of a stable city size distribution before the reforms of 1980 but of convergence in growth thereafter. They also suggest that the best-fitting distribution in China is log-normal. Using data on the area of population of walled cities during Ming and Qing era China (1368-1911), Ioannides and Zhang (2017) find the size-rank relationship is well described by Zipf's Law.

Two papers examine the case of Japan over the very long run. Eaton and Eckstein (1997) examine both Japan (925-1985) and France (1876-1990), in the case of Japan focusing on urban areas above 250,000. They find that the relative populations of the top 40 urban areas of France and Japan remained constant during periods of industrialization and urbanization and are described quite well by Zipf's Law. This is at odds with Davis and Weinstein (2002), who analyze Japan during the period 600BC-1998AD. They find that long-run city size is very robust even to large temporary shocks, such as the Allied bombing on Japan during WWII, something inconsistent with random growth rates for cities (Gibrat's Law). They suggest instead that the evolution of Japanese cities over the long run is consistent instead with a hybrid theory of locational fundamentals and increasing returns.

Sharma (2003) examines Census-defined cities in India over the period 1901-1991 and finds that urban population is non-stationary. While the population of cities may be parallel in the long-run, reflecting common long-run growth rates, in the short-run deviations occur, typically reflecting exogenous shocks that take less than a decade to dissipate. Arshad et al. (2019) examine the case of Pakistan, across five Census years 1951-1998, using administrative boundaries, including metropolitan and municipal corporations, and OLS methods. They observe that Zipf's law does not hold in any of the five census years at national level but that it is more likely to hold for the city-size distribution at province-level, of which there are four. Soo (2007) examines the case of Malaysia across five Censuses between 1957 and 2000, using urban areas of at least 10,000 people and the OLS and Hill estimators. For the full sample, Zipf's law is rejected for all periods except 1957, in favor of a more unequal distribution, while in the upper tail, the results better fit Zipf's Law.

Europe European city growth dynamics have also been the subject of a substantial literature, including Eaton and Eckstein (1997) mentioned above, who found evidence in favour of Zipf's Law in France and Japan. Using the Bairoch et al. (1988) dataset of European cities with a population of more than 5,000 over the course of the second millennium, Dittmar (2020) establishes the emergence of a power-law distribution of city size and rank, first in Western Europe (by 1500) and later in Eastern Europe. This is consistent with technological improvements relaxing the land constraint.

Lanaspa et al. (2003) examine the evolution of Spanish urban structure during the twentieth century. They find divergent growth before 1970 and convergent growth 1970-1999, with significant intra-distribution movements. Their OLS analysis includes the top 100-

700 cities, legally defined, in Spain. Le Gallo and Chasco (2008) undertake an analysis of Spanish towns for the period 1900-2001, with cut-offs of 10,000 and 50,000. They find two main phases, one of divergence (1900–1980) and latterly convergence (1980–2001), and also evidence of the influence of the geographical environment on urban population dynamics. Gisbert and Mas (2010) also employ OLS analysis, in their case on Spain’s municipal populations 1900-2001. They find that rejection of Zipf’s law depends on the concept of cities used.

González-Val et al. (2014) compare Italy, Spain and the USA over the course of the 20th century, using municipality-level data for both Italy and Spain (and incorporated places in the USA). They use the Nadaraya-Watson method and employ a population cutoff of 200. They observe divergent city growth; however, they also find that data are well fitted by a log-normal distribution and that Gibrat’s law holds, at least for certain samples. Lastly, in the case of Spain, González-Val and Silvestre (2020) present annual estimates of population for provinces and provincial capital cities in Spain, 1900–2011. Unlike when data from decennial censuses are used, an analysis of their annual series cannot reject Zipf’s Law after the 1940s.

Research has also been undertaken on the long-run dynamics of city growth in Belgium, Germany and Sweden, with results largely at odds with Gibrat’s Law. Ronsse and Standaert (2017) estimate the population of Belgian municipalities at annual frequency for three sub-periods within the overall period 1880-1970. They reject Gibrat’s Law for Belgium in this period, using the Nadaraya-Watson method. Bosker et al. (2008) examines the urban population in Germany, 1925-1999, using administrative city definitions and cut-offs of 50,000 and 100,000. In addition to World War 2 having a major and lasting impact on city size distributions, he finds that city growth is trend-stationary, consistent with increasing returns to scale but at odds with Gibrat’s Law of proportional effect. A working paper by Probst (2017) examines municipalities in Sweden 1800-2010, using Census records. They find that Gibrat’s Law is rejected in the sample, earlier because of the growth of smaller locations and later because of city agglomeration. The “Zipf” coefficient reaches its peak at 1.15 in 1900 then falls to 0.89 in 2010.

Finally, the paper most closely related to ours in setting is Klein and Leunig (2015). They examine the dynamics of urban growth in England (a large subset of Britain) during the Industrial Revolution period, 1761-1891. They combine data at the parish level to form over 10,000 ‘recognisable towns’ for the Census years 1801-1891 and use data on nearly 600 administrative units known as “hundreds” prior to this. However, they do not appear to employ a cutoff, with the result that their dataset is weighted heavily towards smaller municipalities; in 1895, their mean location has a population of less than 2,500. With this dataset and using the Nadaraya-Watson non-parametric method, they find that Gibrat’s Law is violated consistently, although violations of Gibrat’s Law are driven by areas with a population of less than 2,000. They also find evidence that large places grew too quickly to be consistent with Gibrat’s Law before 1841, especially in locations where the Industrial Revolution took places. The authors do not present any estimates of the Zipf coefficient.

A.2 Tables

Table A.1: Review of Key Papers in the Literature

Author	Method	Cutoff	Country	Time Period	Unit of analysis
Rosen and Resnick (1980)	OLS	Top 50 or 100,000	Cross section of 44 countries	1970	Urban agglomeration and cities
Eaton and Eckstein (1997)	OLS	France: 50,000 Japan: 250,000	France and Japan	France: 1876-1990; Japan: 925 and 1985	Urban areas
Dobkins et al. (2000)	OLS	50,000	US	1900-1990	Metropolitan area
Dobkins and Ioannides (2001)	OLS	50,000	US	1900-1990	Metropolitan area
Davis and Weinstein (2002)	OLS	NA	Japan	-600-1998	Regional density and city size
Ioannides and Overman (2003)	Local Zipf exponent	50,000	US	1900-1990	Metropolitan area
Eeckhout (2004)	OLS	Different cutoffs between 42 and 155,000	US	2000	Legally incorporated "places" or Census Designated Places
Soo (2005)	OLS; Hill	10,000	Cross section of 73 countries	1971-2001	Urban agglomeration and cities
Batty (2006)	OLS	NA	US and GB	USA: 1790-2000; GB: 1901-2001; World: 430BD-2000AD	US: "cities and other urban places", GB: "urban places".
Levy (2009)	OLS	NA	US	2000	Legally incorporated "places" or Census Designated Places.
Rozenfeld et al. (2011)	OLS	NA	US and GB	1981 for GB and 2001 for US	"city" defined as a maximally connected cluster of populated sites defined at high resolution.
Dittmar (2020)	Theil estimator	5,000	Europe	1300-1800	Urban agglomerations
Ioannides and Zhang (2017)	OLS	1000, 800, 400, 200	China	1368-1911	Walled cities area and population
Dingel et al. (2019)	OLS	NA	China India and Brazil	2017	Metropolitan areas, reconstructed with satellite images
Giesen et al. (2010)	Hill, OLS		DE, US, FR, Brazil, CZE, Hungary, IT, CH	2000	"Places" for US, "cities" for DE, "communes" for FR
Giesen and Südekum (2011)	OLS	100,000	Germany	1975-1997	City proper data within regions
Small and Elvidge (2013)	OLS		Asia	1992-2009	Aggregation from satellite images of lighted areas
Le Gallo and Chasco (2008)	OLS, SUR	10,000 and 50,000	Spain	1900-2001	Towns and municipalities
Berry and Okulicz-Kozaryn (2012)	OLS	500,000	US	1900-2010	Economic Areas (EAs) defined by the Bureau of Economic Analysis of the US Department of Commerce
Fragkias and Seto (2009)	OLS	One isolated urban pixel	China	1988-1999	Contiguous urban built-up areas
Soo (2007)	OLS, Hill	10,000	Malaysia	1957, 1970, 1980, 1991, 2000	Urban areas
Jiang et al. (2015)	Hill		World	1992, 2001, 2010	Natural cities extracted from satellite imagery

Table A.2: Summary statistics for the four units

	Local Government District			Unitary Authority			Primary Urban Area		
Year	N	Mean	Median	N	Mean	Median	N	Mean	Median
1801				379	26,777	19,894	61	57,273	24,334
1811				379	30,390	22,860	61	68,253	27,972
1821				379	36,038	26,294	61	83,687	32,779
1831				379	41,698	30,373	61	103,386	40,476
1841				379	47,597	34,897	61	124,253	46,980
1851	527	20,910	5,109	379	53,520	38,117	61	148,331	52,238
1861	581	22,679	5,350	380	60,864	42,451	61	176,064	68,048
1871	930	14,864	4,698						
1881	955	18,086	6,017	380	78,109	49,224	61	269,204	111,689
1891	998	20,903	6,546	348	83,301	51,926	57	307,997	131,258
1901	1,110	22,429	6,969						
1911	1,128	32,217	7,936	380	107,451	62,546	61	409,624	177,354
1921	1,143	30,135	8,595	380	112,550	65,388	61	429,480	183,906
1931	1,077	28,424	9,520	380	117,883	69,578	61	459,634	205,284
1941	986	36,927	14,914						
1951	987	47,328	15,549	380	128,564	81,923	61	490,654	220,001
1961	987	48,473	17,462	380	134,810	90,364	61	507,679	244,559
1971	895	53,917	18,698	380	152,092	101,168	61	565,390	289,065
1981	461	139,827	93,292	380	138,840	105,428	61	479,426	268,653
1991	459	114,087	93,146	380	144,473	114,052	61	487,358	283,121
2001	376	138,409	112,797	380	150,273	119,207	61	504,148	280,756
2011	347	131,776	97,099	380	161,503	125,499	61	546,837	276,786

Note: This table shows the summary statistics for every spacial unit used in the paper

Table A.3: Local Government District

Minimum city size					Sample size				
Year	Conservative	Fraction	Deviation	Level	Year	Conservative	Fraction	Deviation	Level
1851	2,418	13,050	635	50,000	1851	462	106	523	34
1861	2,621	14,712	844	50,000	1861	483	117	574	38
1871	3,641	11,446	1,516	50,000	1871	574	187	837	33
1881	3,840	14,941	2,178	50,000	1881	670	191	829	46
1891	4,322	18,026	3,364	50,000	1891	676	200	768	62
1901	3,955	18,929	4,268	50,000	1901	797	223	763	75
1911	3,939	22,551	3,503	50,000	1911	842	226	875	99
1921	3,867	27,118	13,490	50,000	1921	878	228	425	128
1931	3,549	30,754	44,416	50,000	1931	857	216	147	133
1939	4,374	41,621	13,067	50,000	1939	840	198	536	160
1951	4,262	46,270	9,715	50,000	1951	846	198	632	182
1961	6,051	51,336	10,552	50,000	1961	772	198	641	205
1971	6,932	52,175	12,071	50,000	1971	707	179	578	187
1981	10,119	150,086	14,241	50,000	1981	459	93	456	407
1991	9,986	145,884	66,557	50,000	1991	457	93	353	403
2001	24,457	187,908	89,542	50,000	2001	374	75	266	360
2011	7,375	209,122	117,237	50,000	2011	347	70	144	285

Note: This table shows the size of the population of the smallest city and the sample size according to each cutoff

Table A.4: Unitary Authority

Minimum city size					Sample size				
Year	Conservative	Fraction	Deviation	Level	Year	Conservative	Fraction	Deviation	Level
1801	24,379	36,078	17,904	50,000	1801	155	76	207	45
1811	23,427	41,915	20,795	50,000	1811	184	76	208	55
1821	28,318	48,955	25,040	50,000	1821	184	76	205	72
1831	30,373	57,994	27,298	50,000	1831	190	76	210	97
1841	27,620	64,650	27,620	50,000	1841	226	76	226	118
1851	34,956	70,781	28,217	50,000	1851	208	76	238	133
1861	4,160	81,887	32,235	50,000	1861	366	76	238	157
1881	24,903	100,677	70,725	50,000	1881	309	76	112	188
1891	3,961	105,924	94,701	50,000	1891	344	70	76	185
1911	8,748	152,242	116,538	50,000	1911	373	76	100	245
1921	9,551	166,729	112,113	50,000	1921	375	76	112	255
1931	15,728	179,652	105,014	50,000	1931	375	76	118	278
1951	4,152	209,713	104,299	50,000	1951	379	76	135	321
1961	5,860	210,098	108,208	50,000	1961	379	76	144	338
1971	8,450	213,711	124,330	50,000	1971	379	76	138	355
1981	5,095	195,659	115,288	50,000	1981	379	76	164	359
1991	4,129	202,193	118,089	50,000	1991	379	76	178	364
2001	7,276	210,134	119,126	50,000	2001	379	76	191	366
2011	7,375	231,997	137,835	50,000	2011	379	76	168	370

Note: This table shows the size of the population of the smallest city and the sample size according to each cutoff

Table A.5: Primary Urban Area

Minimum city size					Sample size				
Year	Conservative	Fraction	Deviation	Level	Year	Conservative	Fraction	Deviation	Level
1801	1,701	58,324	2,055	50,000	1801	61	13	60	14
1811	2,235	65,695	2,503	50,000	1811	61	13	60	17
1821	2,717	82,816	3,281	50,000	1821	61	13	60	19
1831	3,073	108,566	3,784	50,000	1831	61	13	60	25
1841	3,382	124,548	4,629	50,000	1841	61	13	60	28
1851	3,614	137,176	6,413	50,000	1851	61	13	60	34
1861	3,656	162,451	7,793	50,000	1861	61	13	60	38
1881	13,893	257,805	16,667	50,000	1881	61	13	60	46
1891	19,026	306,445	20,503	50,000	1891	57	11	56	47
1911	31,137	345,479	37,246	50,000	1911	61	13	60	58
1921	36,254	350,494	36,789	50,000	1921	61	13	60	58
1931	35,368	358,630	45,396	50,000	1931	61	13	60	59
1951	41,886	381,132	65,139	50,000	1951	61	13	60	60
1961	48,434	429,253	65,776	50,000	1961	61	13	60	60
1971	65,925	470,256	72,023	50,000	1971	61	13	60	60
1981	70,836	444,240	85,154	50,000	1981	61	13	60	60
1991	80,799	457,568	91,909	50,000	1991	61	13	60	60
2001	85,056	452,404	97,568	50,000	2001	61	13	60	60
2011	90,254	479,924	104,640	50,000	2011	61	13	60	60

Note: This table shows the size of the population of the smallest city and the sample size according to each cutoff

Table A.6: Gini Coefficient

Local Government District					Unitary Authority				
Year	Conservative	Fraction	Deviation	Level	Year	Conservative	Fraction	Deviation	Level
1851	0.715	0.625	0.728	0.603	1851	0.347	0.292	0.363	0.319
1861	0.72	0.632	0.736	0.601	1861	0.48	0.29	0.372	0.338
1871	0.682	0.639	0.712	0.62	1871				
1881	0.671	0.618	0.69	0.606	1881	0.437	0.301	0.341	0.39
1891	0.67	0.6	0.682	0.58	1891	0.485	0.293	0.304	0.4
1901	0.678	0.594	0.673	0.561	1901				
1911	0.749	0.681	0.753	0.667	1911	0.482	0.266	0.298	0.414
1921	0.682	0.568	0.611	0.538	1921	0.48	0.262	0.304	0.413
1931	0.637	0.466	0.422	0.41	1931	0.468	0.248	0.297	0.414
1939	0.627	0.495	0.555	0.489	1939				
1951	0.697	0.598	0.654	0.598	1951	0.424	0.209	0.276	0.385
1961	0.66	0.577	0.631	0.576	1961	0.397	0.203	0.274	0.37
1971	0.662	0.541	0.634	0.543	1971	0.372	0.203	0.256	0.352
1981	0.452	0.45	0.448	0.425	1981	0.335	0.184	0.248	0.319
1991	0.335	0.21	0.278	0.296	1991	0.319	0.176	0.237	0.306
2001	0.302	0.169	0.249	0.287	2001	0.313	0.174	0.236	0.302
2011	0.409	0.161	0.235	0.34	2011	0.321	0.168	0.23	0.311

Note: This table shows the Gini coefficient for LGD and UA according to each cutoff

Table A.7: Gini Coefficient

Primary Urban Area				
Year	Conservative	Fraction	Deviation	Level
1801	0.627	0.512	0.664	0.521
1811	0.638	0.514	0.668	0.533
1821	0.648	0.512	0.67	0.538
1831	0.664	0.505	0.673	0.559
1841	0.673	0.501	0.678	0.576
1851	0.678	0.502	0.682	0.596
1861	0.685	0.514	0.685	0.604
1881	0.663	0.508	0.663	0.62
1891	0.667	0.549	0.671	0.643
1911	0.646	0.521	0.646	0.639
1921	0.635	0.522	0.638	0.635
1931	0.638	0.532	0.638	0.634
1951	0.622	0.528	0.622	0.622
1961	0.607	0.523	0.607	0.607
1971	0.578	0.492	0.578	0.582
1981	0.558	0.501	0.558	0.562
1991	0.546	0.506	0.546	0.549
2001	0.544	0.512	0.544	0.547
2011	0.549	0.522	0.549	0.553

Note: This table shows the Gini coefficient for PUA according to each cutoff

Table A.8: Spline regression for Primary Urban Area and Conservative cutoff - 1

	Annual growth rates					
	1801-1811	1811-1821	1821-1831	1831-1841	1841-1851	1851-1861
00to10	0.0004 (0.0005)	0.00001 (0.001)	-0.00002 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.002* (0.001)
10to11	-0.00002 (0.0004)	-0.0004 (0.0004)	-0.0001 (0.0005)	-0.001 (0.0005)	-0.00001 (0.0005)	0.001 (0.001)
11to12	-0.0003 (0.001)	0.00002 (0.001)	0.0004 (0.001)	0.001 (0.001)	-0.0001 (0.001)	-0.0004 (0.001)
12to14	0.001 (0.003)	0.0005 (0.002)	0.0003 (0.001)	0.0004 (0.001)	0.0003 (0.001)	0.0001 (0.001)
14to17	-0.297 (0.591)	-0.008 (0.018)	-0.004 (0.006)	-0.003 (0.003)	-0.001 (0.002)	-0.0004 (0.002)
Constant	-0.002 (0.004)	0.002 (0.005)	0.002 (0.005)	0.007 (0.005)	0.010* (0.005)	0.016** (0.008)
Observations	47	49	51	54	55	55
R ²	0.029	0.045	0.038	0.142	0.083	0.080
Adjusted R ²	-0.089	-0.066	-0.069	0.053	-0.011	-0.013

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.9: Spline regression for Primary Urban Area and Conservative cutoff - 2

	Annual growth rates					
	1861-1881	1881-1891	1891-1911	1911-1921	1921-1931	1931-1951
00to10	-0.004*** (0.001)	0.0003 (0.004)				
10to11	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	0.002 (0.002)	0.004 (0.007)	-0.003 (0.002)
11to12	-0.001 (0.001)	-0.0002 (0.0004)	-0.0004 (0.0004)	-0.0001 (0.0003)	-0.001 (0.0003)	-0.001** (0.0002)
12to14	0.0001 (0.001)	-0.0003 (0.0003)	-0.0002 (0.0003)	-0.0003 (0.0002)	-0.0003 (0.0002)	-0.0002 (0.0001)
14to17	-0.0004 (0.002)	0.0003 (0.001)	0.0001 (0.001)	0.0001 (0.0004)	0.0003 (0.0004)	0.0001 (0.0003)
Constant	0.044*** (0.010)	-0.001 (0.036)	0.003*** (0.001)	-0.001 (0.001)	-0.003 (0.006)	0.004** (0.002)
Observations	56	56	55	56	55	56
R ²	0.508	0.139	0.188	0.077	0.157	0.321
Adjusted R ²	0.458	0.053	0.123	0.005	0.090	0.268

Note: *p<0.1; **p<0.05; ***p<0.01

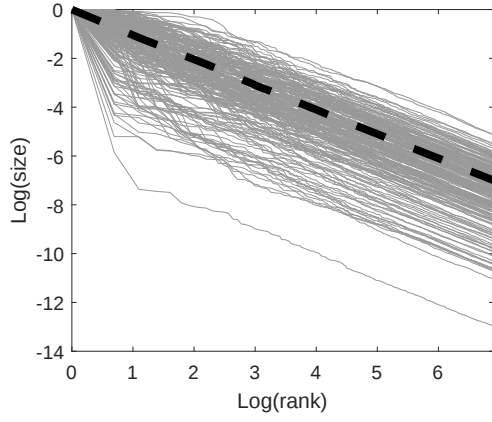
Table A.10: Spline regression for Primary Urban Area and Conservative cutoff - 3

	Annual growth rates					
	1951-1961	1961-1971	1971-1981	1981-1991	1991-2001	2001-2011
00to10						
10to11						
11to12	-0.001* (0.0004)	0.0002 (0.001)	-0.0002 (0.001)	-0.001* (0.0005)	-0.001** (0.0004)	-0.001 (0.0004)
12to14	-0.0003 (0.0002)	-0.001 (0.0004)	-0.001*** (0.0004)	-0.0005*** (0.0002)	-0.0002* (0.0001)	-0.0001 (0.0001)
14to17	-0.00001 (0.0004)	-0.0002 (0.001)	0.001 (0.001)	0.0003 (0.0003)	0.0002 (0.0002)	0.0002 (0.0002)
Constant	0.001*** (0.0003)	0.001 (0.001)	0.0003 (0.001)	0.001*** (0.0004)	0.001*** (0.0004)	0.001*** (0.0004)
Observations	56	56	56	56	56	56
R ²	0.186	0.073	0.178	0.319	0.207	0.129
Adjusted R ²	0.139	0.020	0.131	0.280	0.162	0.079

Note: *p<0.1; **p<0.05; ***p<0.01

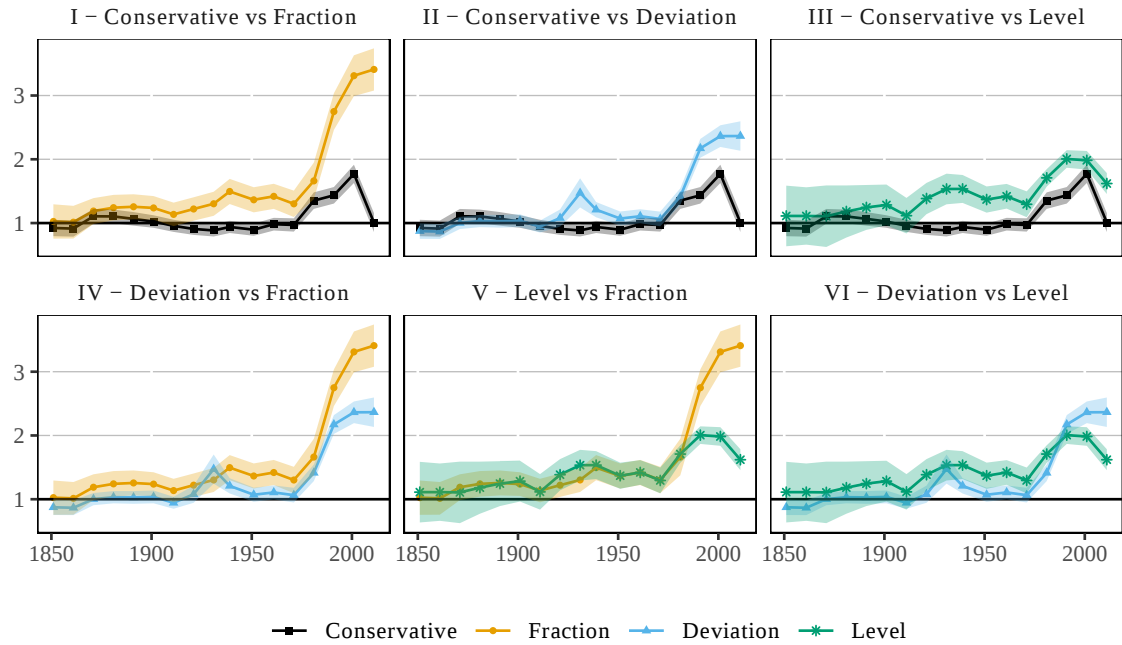
A.3 Figures

Figure A.1: Size-rank coefficient: Local Government Districts



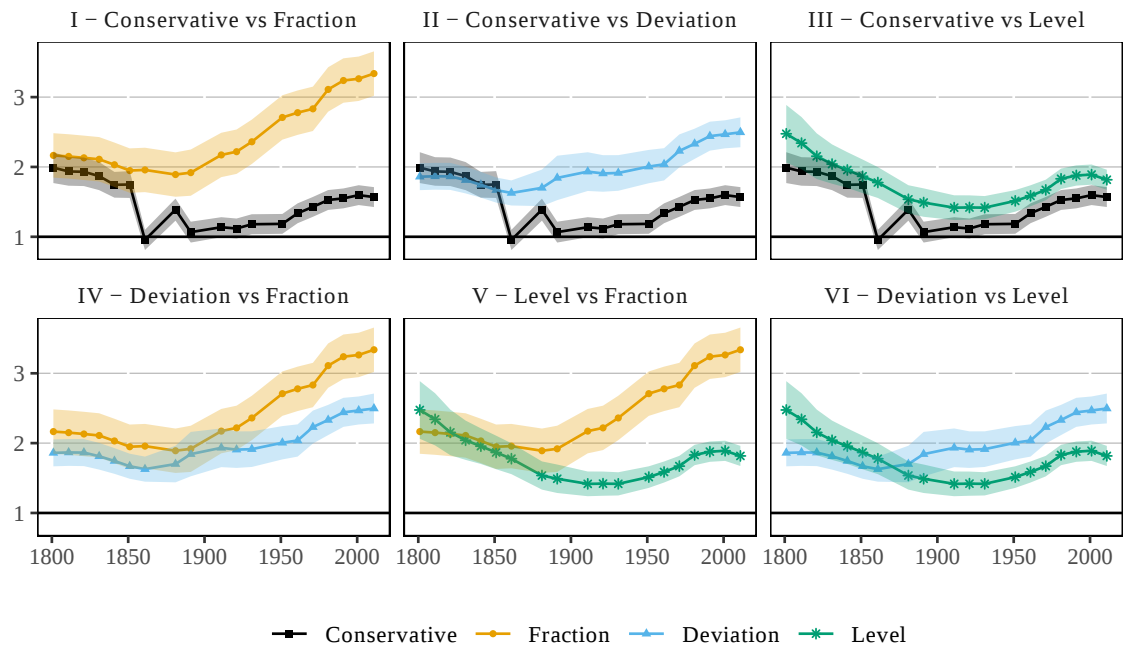
Note: This figure shows the $\log(\text{size})$ - $\log(\text{rank})$ plot of 1000 simulations of a power law distribution with true parameter equal to 1 and normalisation of the bigger city to 1. As it is visible, the error on large cities is more negative than it is on smaller cities, which induces OLS to be flatter, while minimising residuals. .

Figure A.2: Size-rank coefficient: Local Government Districts



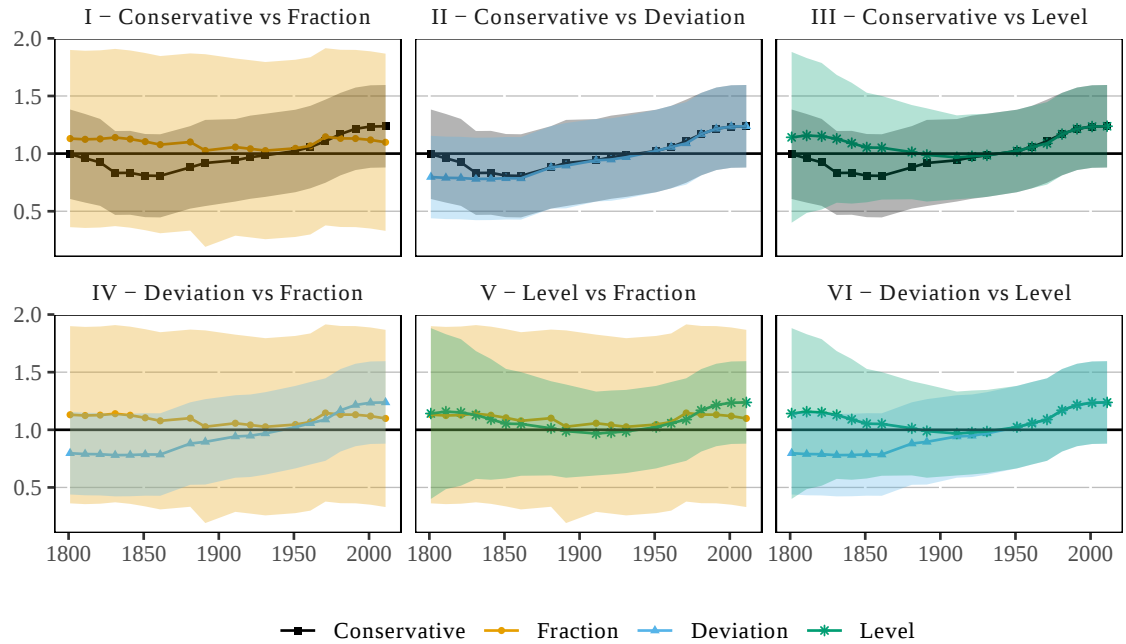
Note: The six panels in this figure show the pairwise comparison between all cutoff methods of the absolute value of the Pareto coefficient of Local Government Districts, for each Census year.

Figure A.3: Size-rank coefficient: District/Unitary Authority



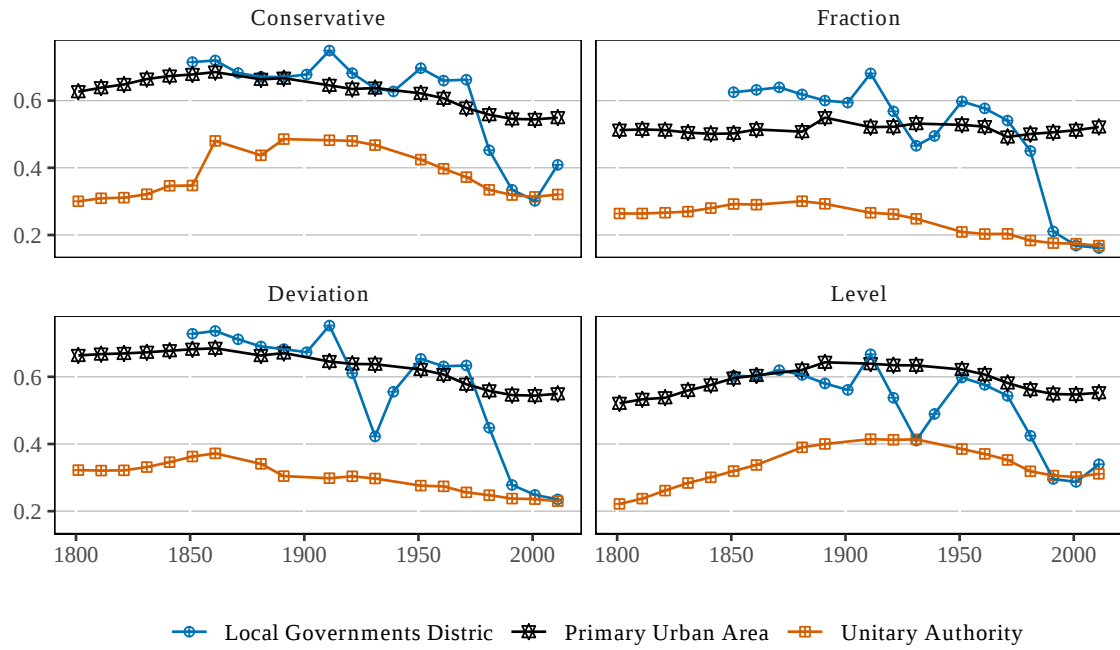
Note: The six panels in this figure show the pairwise comparison between all cutoff methods of the absolute value of the Pareto coefficient of District/Unitary Authority, for each Census year.

Figure A.4: Size-rank coefficient: Primary Urban Area



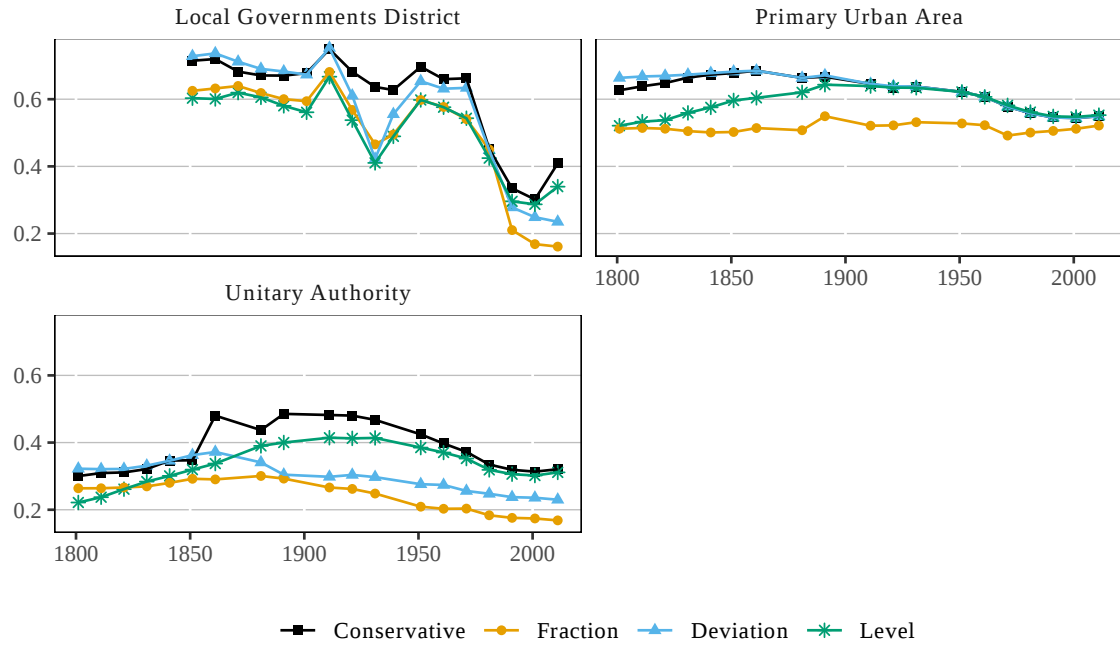
Note: The six panels in this figure show the pairwise comparison between all cutoff methods of the absolute value of the Pareto coefficient of Primary Urban Area, for each Census year.

Figure A.5: Gini coefficient for four different units



Note: This figure shows the evolution of the Gini coefficient for all different cutoff in each area.

Figure A.6: Gini coefficient for four different cutoffs



Note: This figure shows the evolution of the Gini coefficient for all different units for each different cutoff.

Figure A.7: Lorenz curve for Conservative cutoff - Local Government District

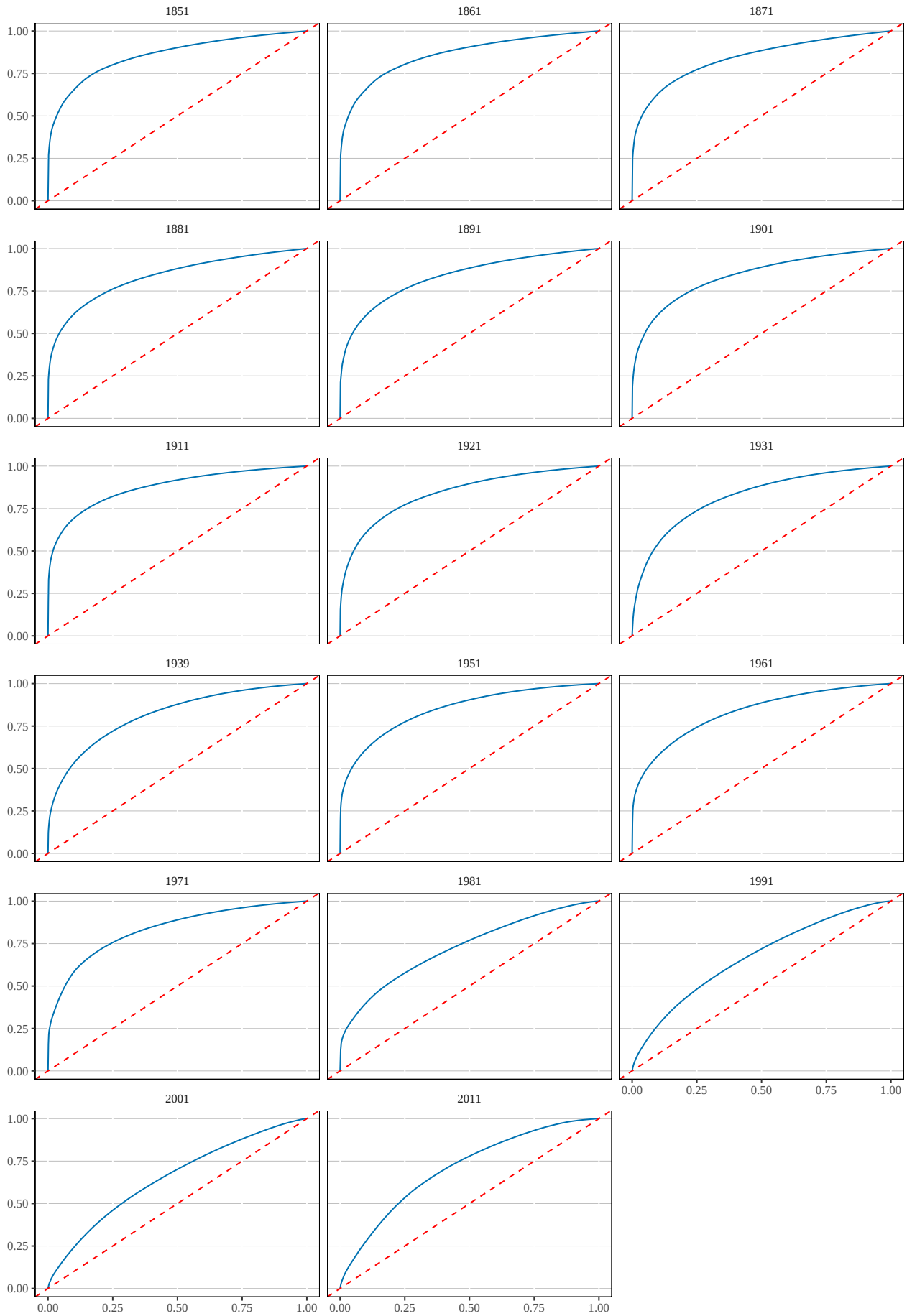


Figure A.8: Lorenz curve for Fraction cutoff - Local Government District

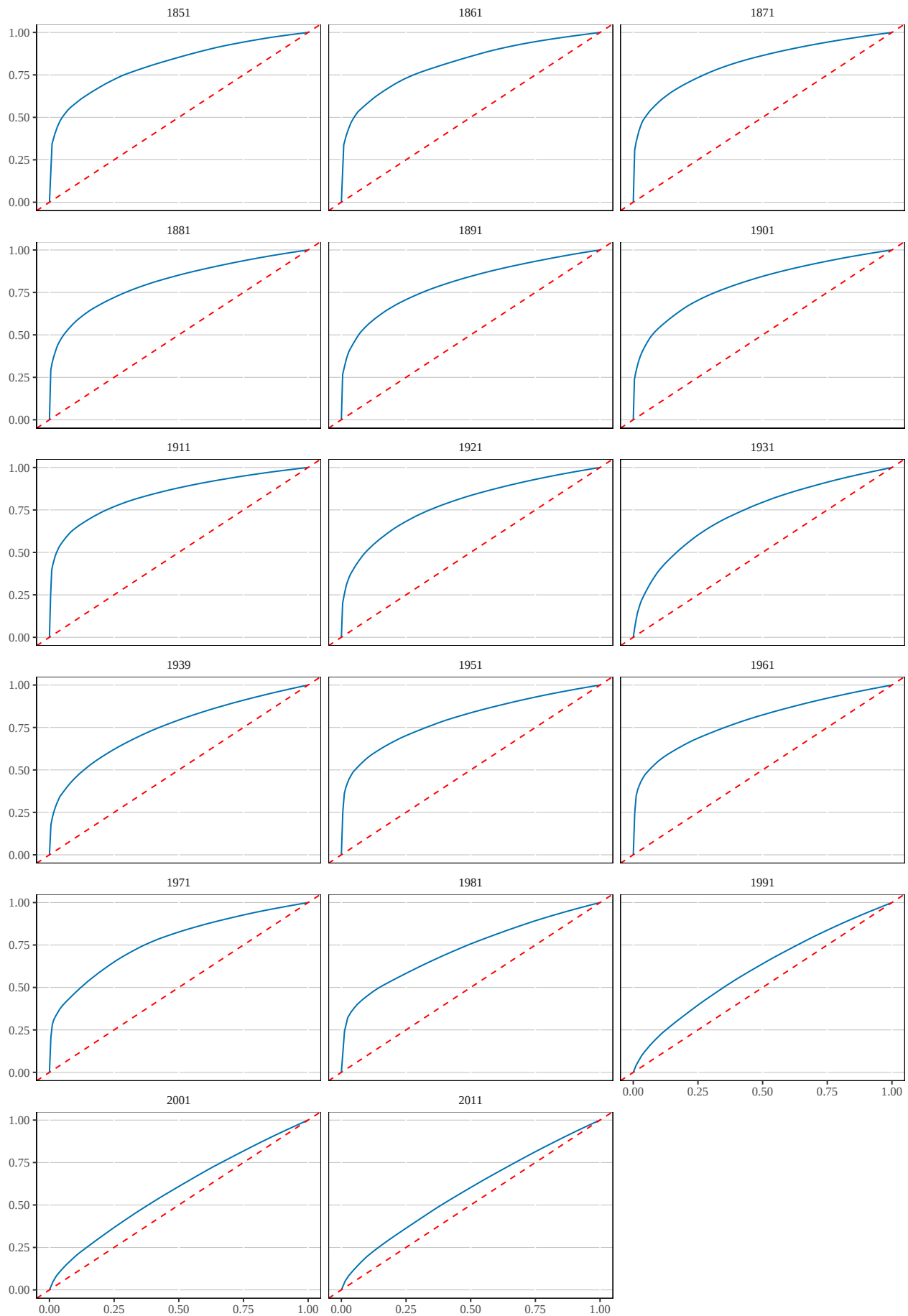


Figure A.9: Lorenz curve for Level cutoff - Local Government District

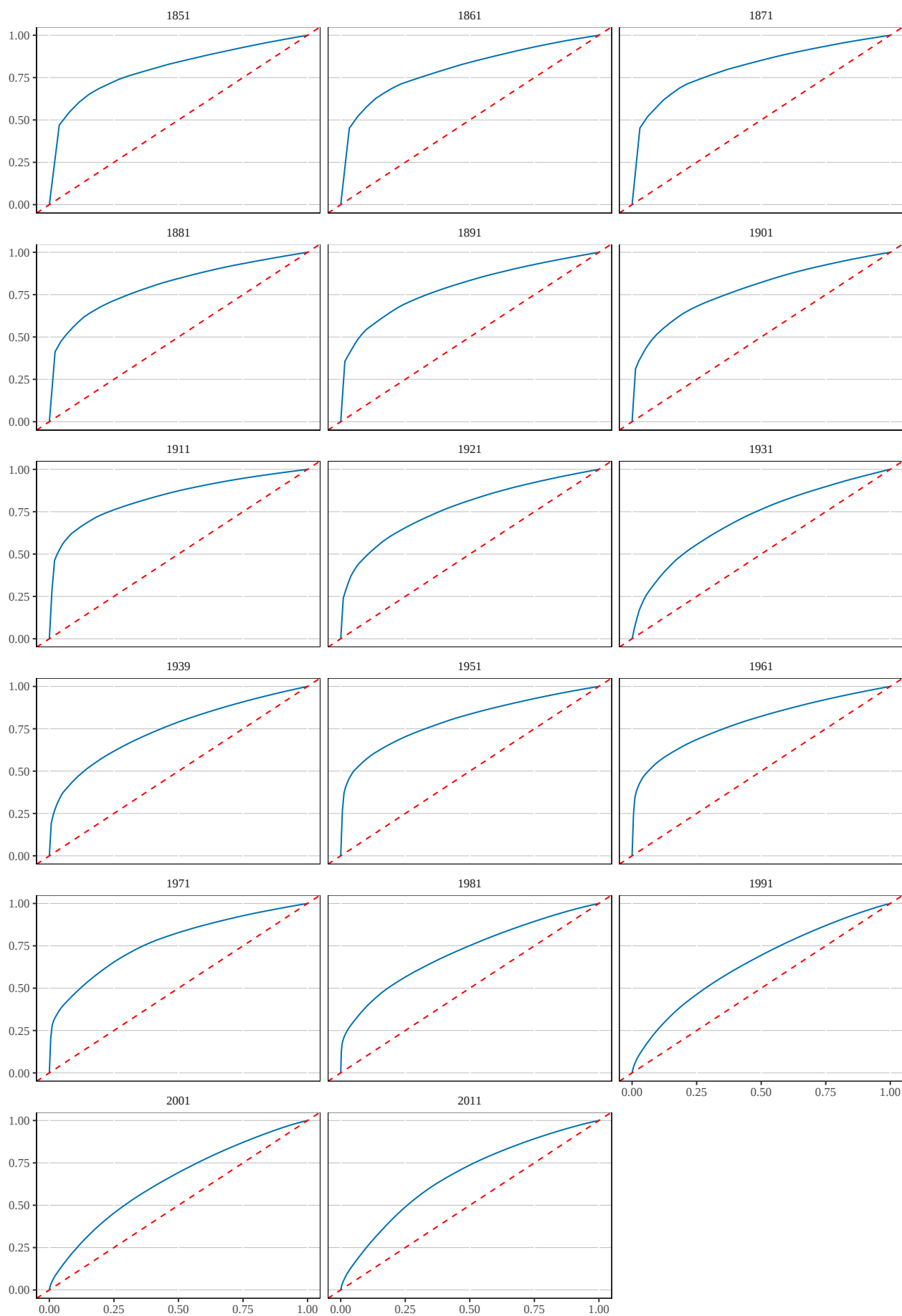


Figure A.10: Lorenz curve for Deviation cutoff - Local Government District

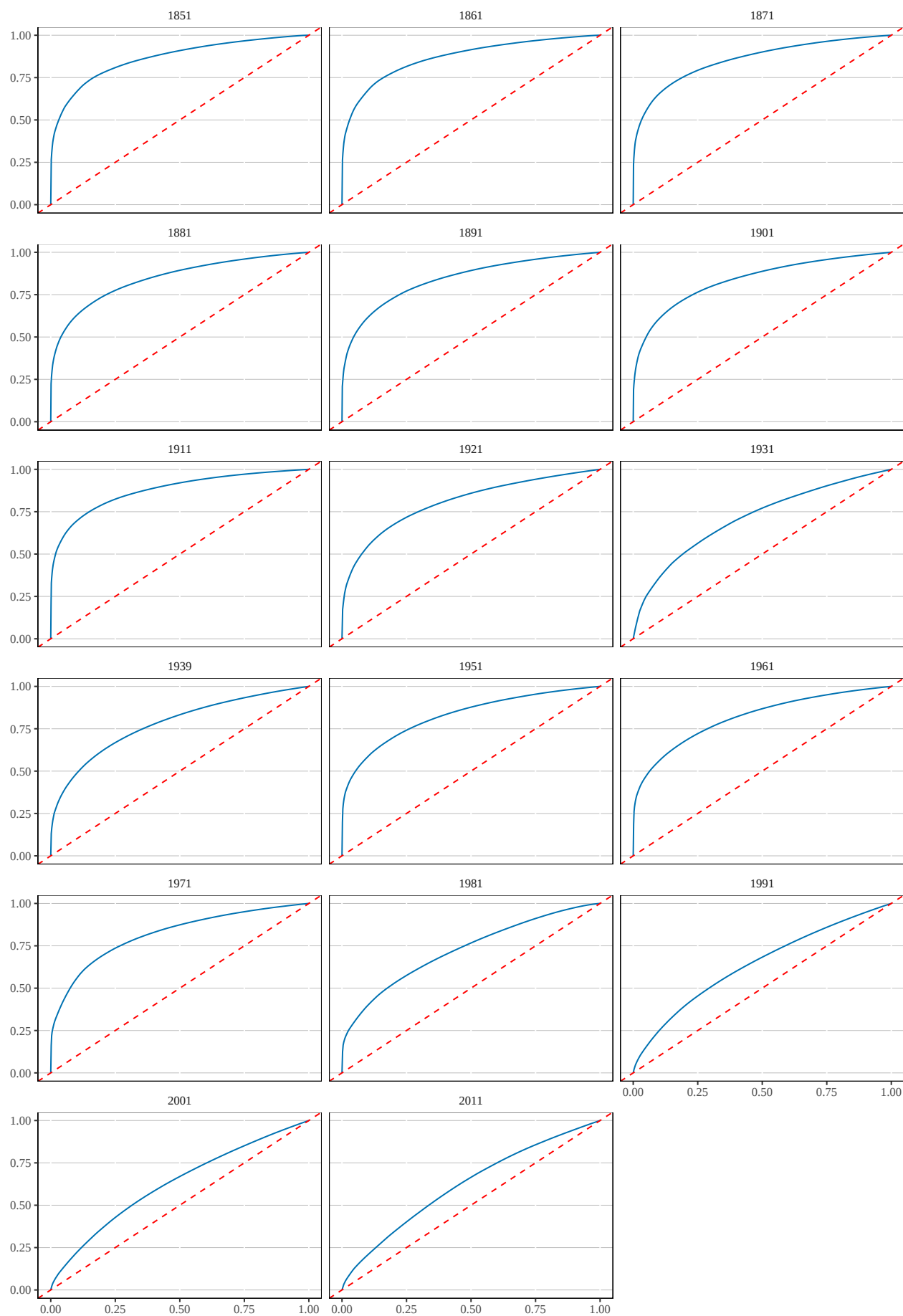


Figure A.11: Lorenz curve for Conservative cutoff - Unitary Authority

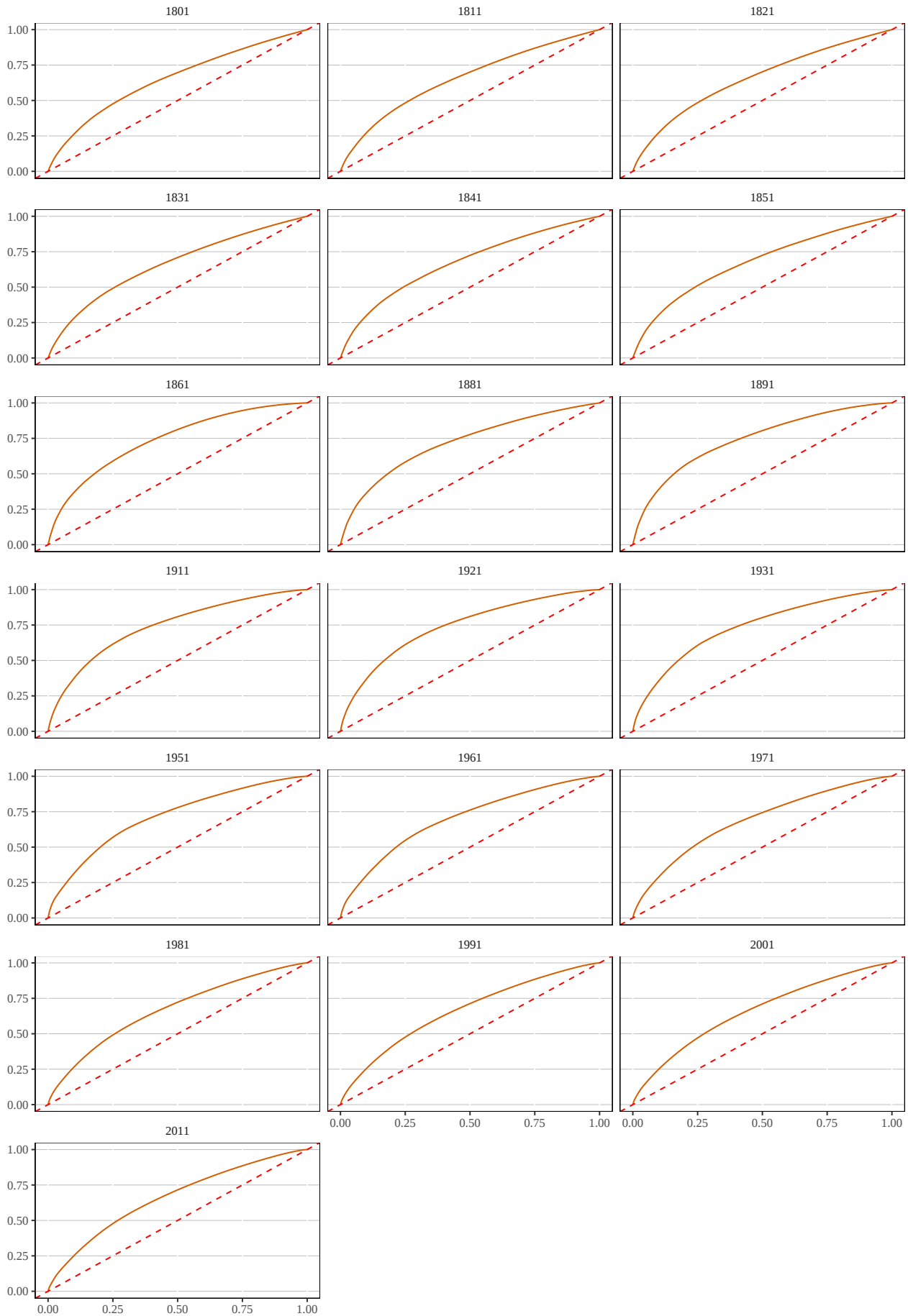


Figure A.12: Lorenz curve for Fraction cutoff - Unitary Authority

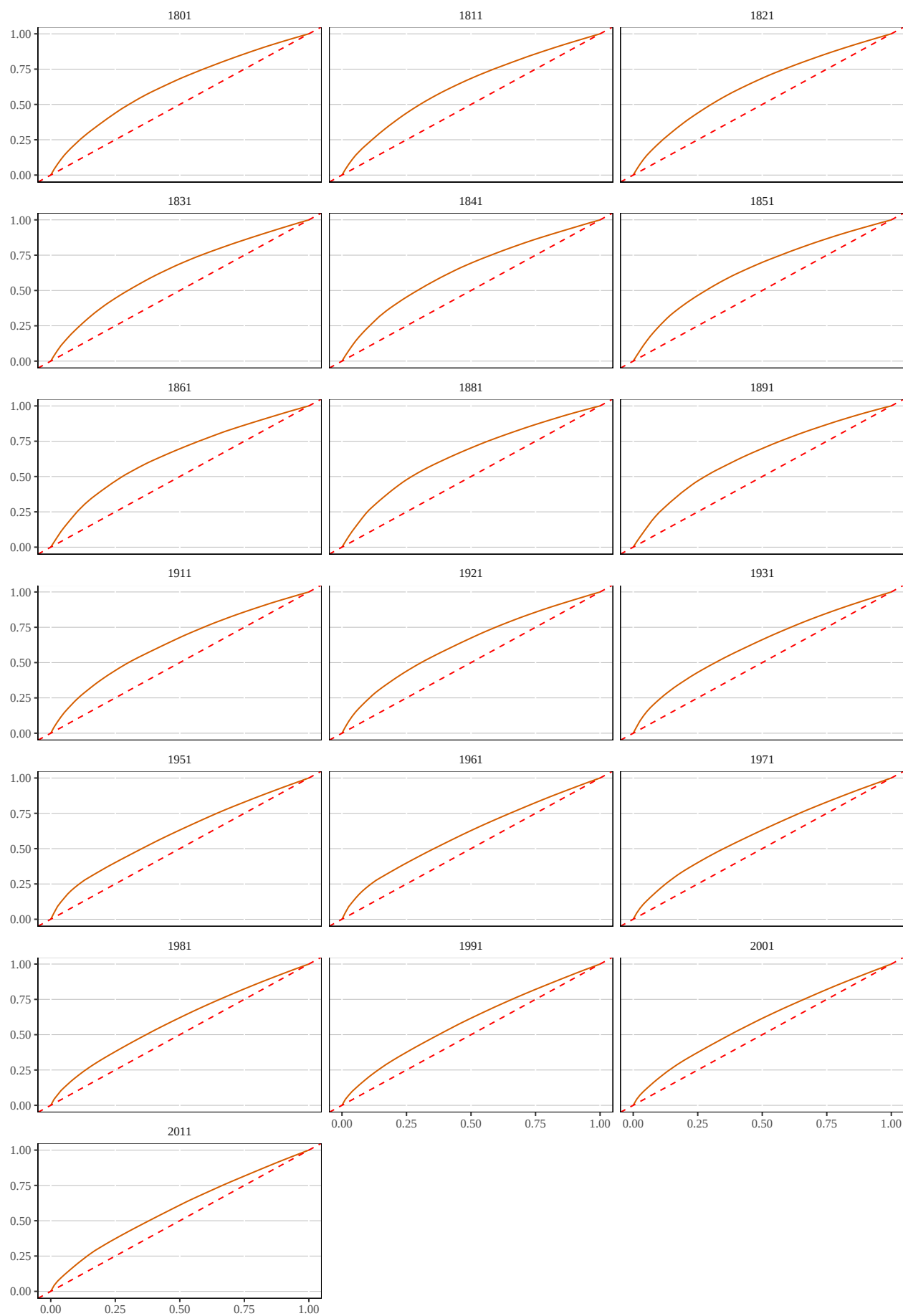


Figure A.13: Lorenz curve for Level cutoff - Unitary Authority

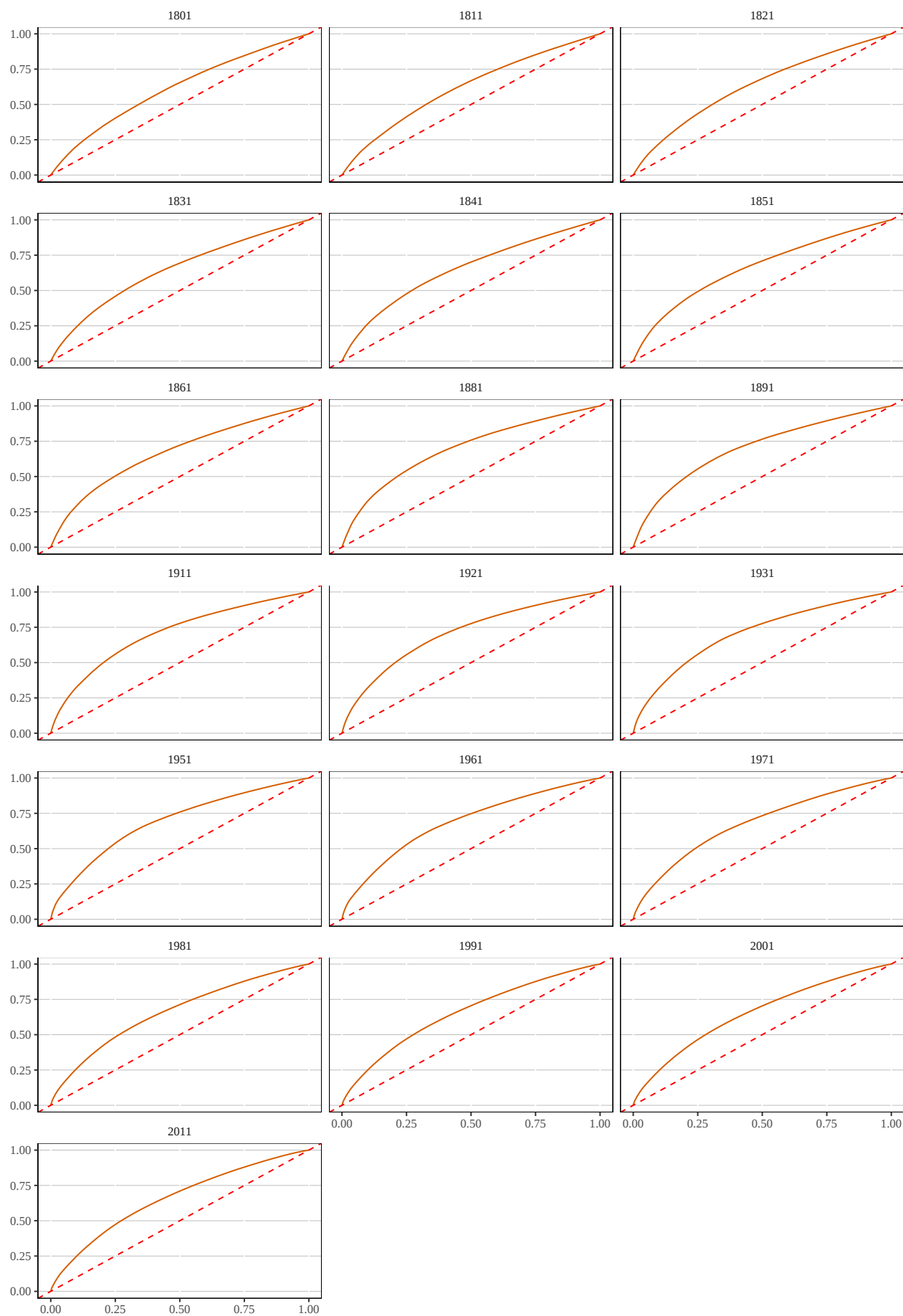
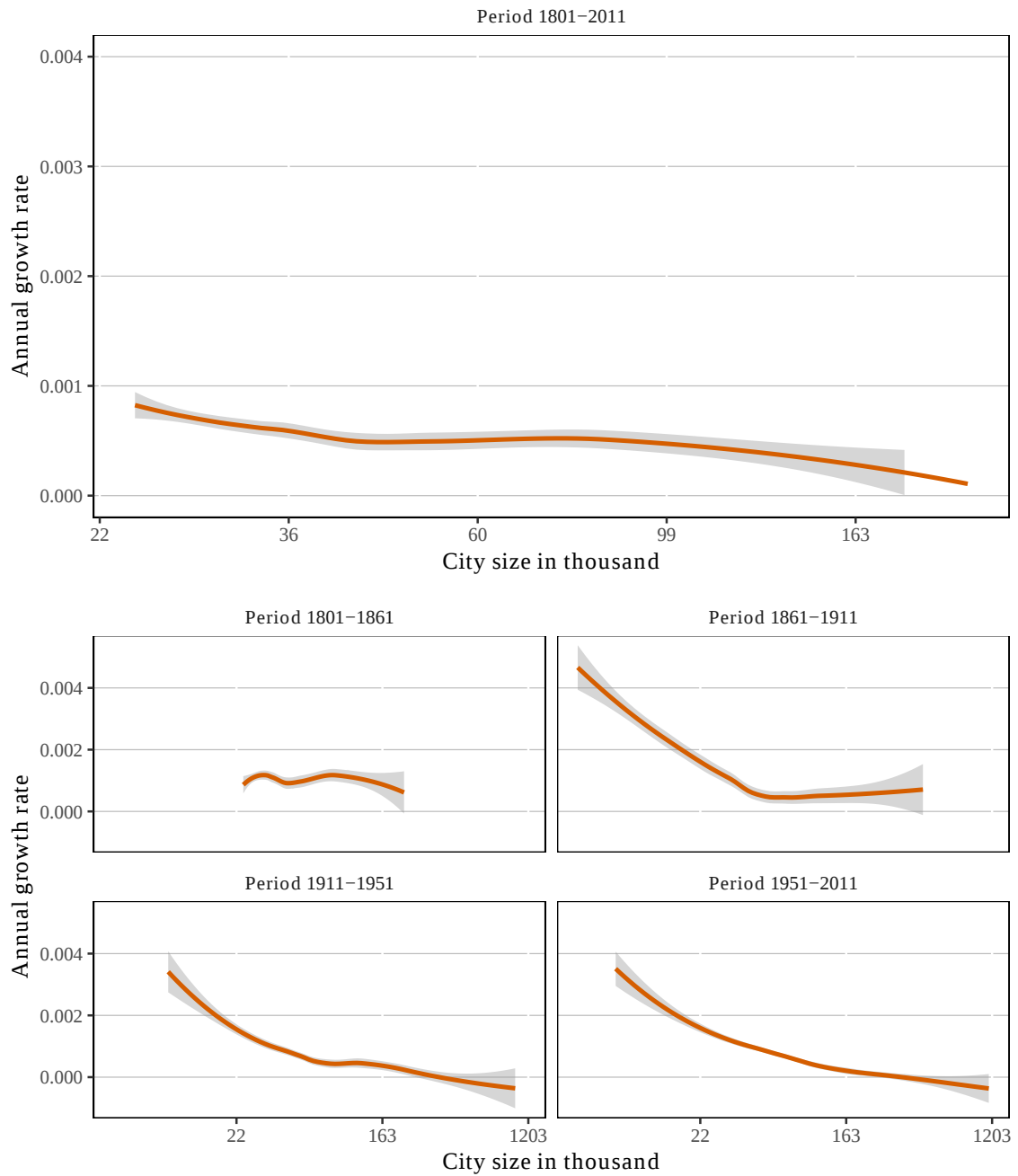


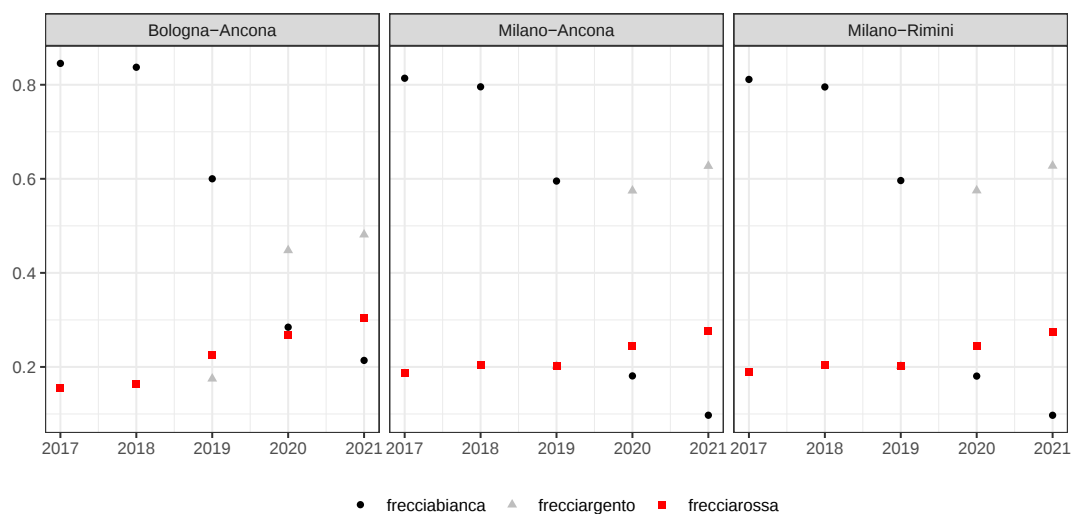
Figure A.14: Kernel regression: Unitary Authority with Conservative cutoff



Note: Kernel regression: growth rate for the whole period 1801 to 2011 plotted against initial city size in 1801, and intermediate periods 1801-1861, 1861-1911, 1911-1951 and 1951-2011 plotted against initial size.

A.4 Market segmentation and long-run competition: the case of the Italian rail market: Appendix

Figure A.15: Train type share on tourist segment



Note: This figure shows the share of train type on the tourist routes where competition starts in 2020. Increase in price can be also justified by higher share of Frecciargento and Frecciarossa trains, which are higher level trains.

Table A.11: Pooled difference in difference on both segments

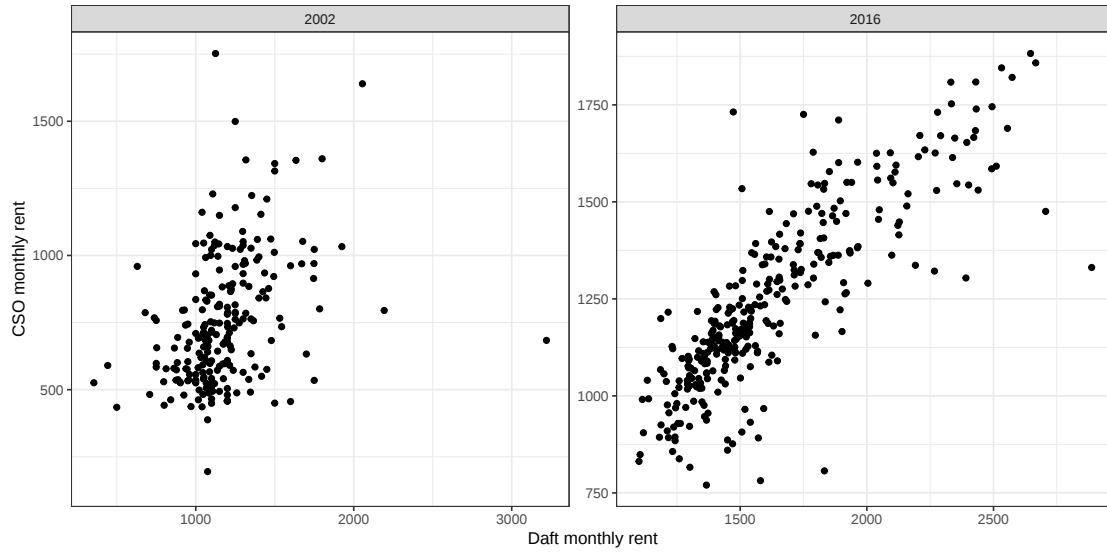
	log(price)			
	1 day before		20 days before	
	2017 to 2020	2017-2019-2020-2021	2017 to 2020	2017-2019-2020-2021
Trip length	−0.002*** (0.00000)	−0.002*** (0.000)	−0.003*** (0.000)	−0.002*** (0.000)
Treated	0.010 (0.011)	−0.052*** (0.014)	−0.075 (0.041)	−0.296*** (0.042)
Post-entry	0.072*** (0.005)	0.204*** (0.013)	−0.066* (0.027)	0.034 (0.023)
Tourist	0.106*** (0.005)	0.104*** (0.017)	0.097*** (0.025)	−0.103*** (0.019)
DID-Core	−0.022 (0.019)	−0.030 (0.022)	−0.092 (0.066)	−0.107 (0.058)
Treated-Tourist	0.008 (0.012)	−0.149*** (0.021)	0.306*** (0.065)	0.270*** (0.031)
DID-Tourist	0.063 (0.025)	0.061 (0.031)	0.130 (0.088)	0.271*** (0.063)
Constant	−1.062*** (0.003)	−0.992*** (0.009)	−1.206*** (0.017)	−1.160*** (0.017)
Route FE	Yes	Yes	Yes	Yes
Treated-Tourist	Yes	Yes	Yes	Yes
Observations	2,871	3,271	2,527	3,104
R ²	0.905	0.963	0.802	0.875
Adjusted R ²	0.904	0.963	0.800	0.873
Residual Std. Error	0.071	0.047	0.131	0.105
F Statistic	967.970***	3,037.993***	360.895***	765.670***

Note:

* p<0.05; ** p<0.01; *** p<0.001

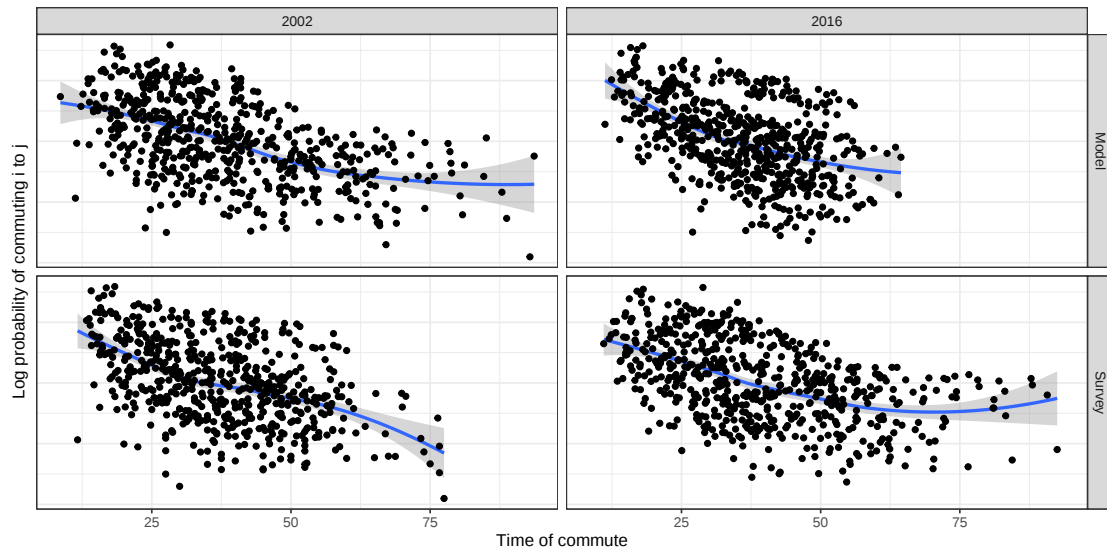
A.5 Skill-biased transportation change: Population, gentrification and Dublin's Luas light rail: Appendix

Figure A.16: CSO rents vs Daft rents



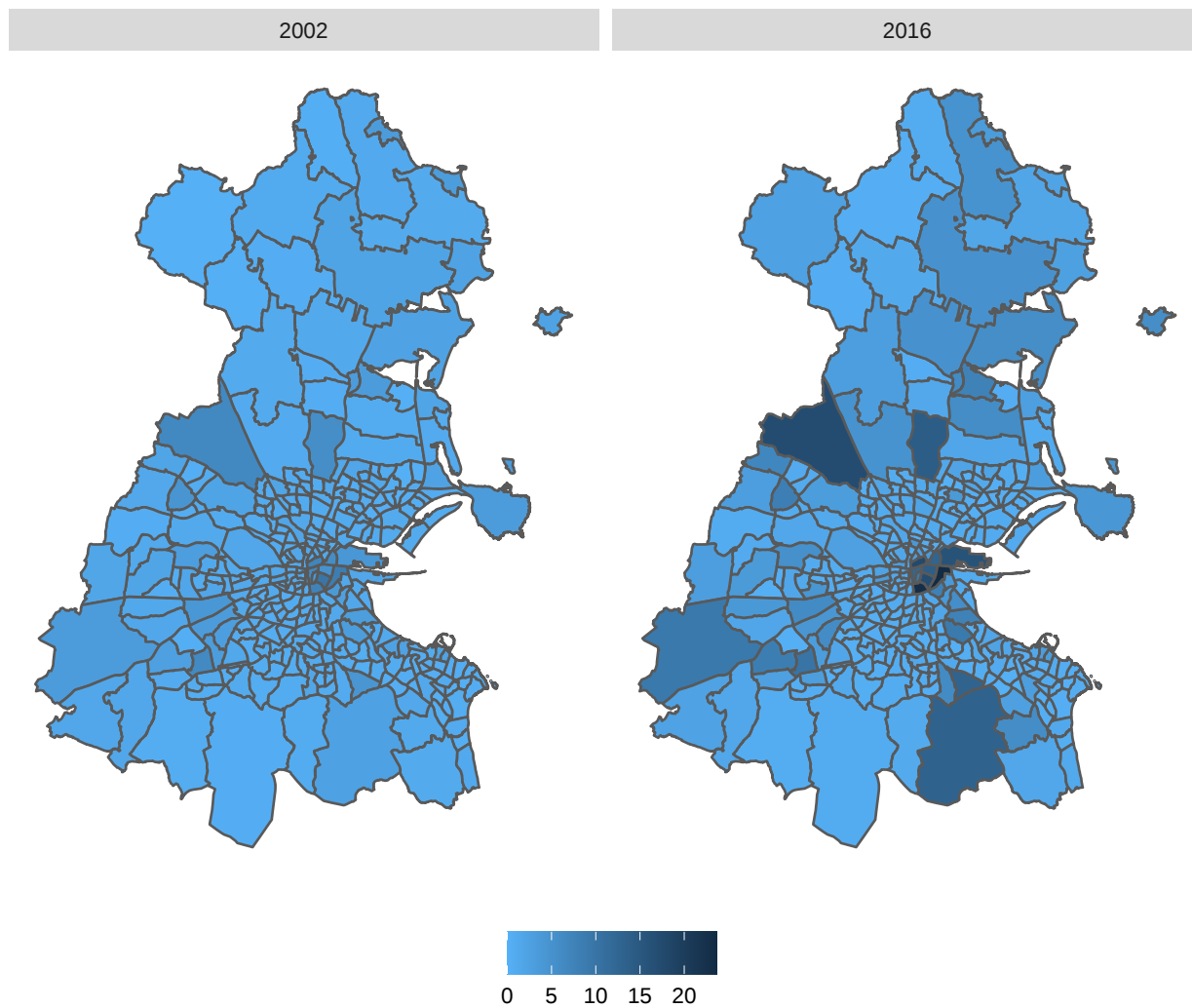
This figure is here just to show you and have an opinion on whether something can be done with daft rents.

Figure A.17: Estimation of $\kappa\epsilon$



The four panels show the estimation of $\kappa\epsilon$ with the four proposed methods; using the survey for years 2002 or 2016, and using my network again for 2002 and 2016. The estimate of $\kappa\epsilon$ we keep in the rest of the paper is the one from year 2016 and the survey.

Figure A.18: Wage calibration



The left panel shows the level of calibrated wage per ED in 2002, while the right panel shows the level of calibrated wages in 2016. The darker the shade of blue the higher the wage level in a given ED

Table A.12: Effect of introduction of LUAS on ED rents

	log(residential rent)			
	GREEN line		RED line	
	OLS	IV	OLS	IV
Δ RCMA	−0.028 (0.045)	−0.007 (0.045)	−0.030 (0.045)	−0.008 (0.045)
No heating	0.088 (0.060)	0.090 (0.060)	0.088 (0.060)	0.090 (0.060)
Old house	0.119* (0.052)	0.118* (0.052)	0.118* (0.052)	0.117* (0.051)
No public sewerage	0.182 (0.106)	0.129 (0.067)	0.183 (0.106)	0.129 (0.067)
PC FE	Yes	Yes	Yes	Yes
ED area	Yes	No	Yes	No
Intercept	No	No	No	No
F 1 st stage	6958.451***		7706.782***	
Observations	322	322	322	322
R ²	0.180	0.179	0.180	0.179
Adjusted R ²	0.098	0.100	0.099	0.100
Residual Std. Error	0.273	0.273	0.273	0.273
F Statistic	2.208***		2.210***	

This table shows the estimation of equation 3.14, where residential rent is the outcome and the regressor is Residential Commuting Market Access (RCMA), in the counterfactual scenario where only the green line was introduced (columns 1 and 2) and in the scenario where only the red line was introduced.

Note:

*p<0.05; **p<0.01; ***p<0.001

Table A.13: Placebo Effect of introduction of LUAS on probability of commuting

	<i>Dependent variable:</i>		
	log(share of pop)	log(share of HS)	log(share of LS)
$\Delta \log(\text{Placebo RCMA})$	0.0193 (0.47)	0.0244 (0.53)	0.0183 (0.49)
PC FE	Yes	Yes	Yes
ED area	Yes	Yes	Yes
Constant	No	No	No
Observations	322	322	322
R ²	0.300	0.249	0.320
Adjusted R ²	0.238	0.182	0.260
Residual Std. Error	0.327	0.369	0.340
F Statistic	1.775***	2.333***	3.862***

This table shows the estimation of the main specification, with the placebo placement of the Luas. Log share of population is the log of the share of resident population of each ED over the total population resident in Dublin, Log share of HS is the log of the share of higher-skilled workers resident in each ED over the total workers resident in each ED, Log share of LS is the log of the share of lower-skilled workers resident in each ED over the total worker in each ED and log(RCMA) is log of calibrated residential market access defined as per equation 3.9, with Luas placed along the Dublin & Lucan Electric Tramway.

Note:

*p<0.05; **p<0.01; ***p<0.001

Table A.14: Effect of introduction of LUAS on share on probability of commuting

	<i>Dependent variable:</i>		
	log(share of workers)	log(share of HS)	log(share of LS)
$\Delta \log(\text{FCMA})$ Placebo	0.0383 (0.43)	0.0341 (0.68)	0.0471 (0.49)
PC FE	Yes	Yes	Yes
ED area	Yes	Yes	Yes
Constant	No	No	No
Observations	319	307	307
R ²	0.356	0.212	0.411
Adjusted R ²	0.299	0.139	0.357
Residual Std. Error	0.435	0.593	0.426
F Statistic	3.205***	2.898***	3.3***

This table shows the estimation of the main specification. Log share of workers of is the log of the share of workers of each ED over the total workers in Dublin, Log share of HS is the log of the share of higher-skilled workers working in each ED over the total workers working in each ED, Log share of LS is the log of the share of lower-skilled workers working in each ED over the total workers working in each ED and log(FCMA) is log of calibrated residential market access defined as per equation 3.10, with Luas placed along the Dublin & Lucan Electric Tramway.

Note:

*p<0.05; **p<0.01; ***p<0.001

