Year : 2018

Three Essays in Régional and Urban Economies

Himbert Alexander

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT D'ÉCONOMIE

Three Essays in Regional and Urban Economics

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences Économiques, mention « Économie politique »

par

Alexander HIMBERT

Co-directeurs de thèse
Prof. Olivier Cadot
Prof. Marius Brülhart

Jury

Prof. Christian Zehnder, Président
Prof. Rafael Lalive, expert interne
Prof. Céline Carrère, experte externe

LAUSANNE
2018

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT D'ÉCONOMIE

Three Essays in Regional and Urban Economics

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences Économiques, mention « Économie politique »

par

Alexander HIMBERT

Co-directeurs de thèse
Prof. Olivier Cadot
Prof. Marius Brülhart

Jury

Prof. Christian Zehnder, Président
Prof. Rafael Lalive, expert interne
Prof. Céline Carrère, experte externe

LAUSANNE
2018

UNIL | Université de Lausanne
HEC Lausanne
Le Décanat
Bâtiment Internef
CH-1015 Lausanne

IMPRIMATUR

Sans se prononcer sur les opinions de l'auteur, la Faculté des Hautes Études Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Monsieur Alexander HIMBERT, titulaire d'un bachelor en Sciences Économiques de l'Université de Göttingen, et d'un master en Sciences Économiques de l'Université de Maastricht et de l'Université catholique de Louvain, en vue de l'obtention du grade de docteur ès Sciences Économiques, mention « Économie Politique ».

La thèse est intitulée :

THREE ESSAYS IN REGIONAL AND URBAN ECONOMICS

Lausanne, le 27 août 2018

Le doyen

Jean-Philippe Bonardi

UNIL | Université de Lausanne

Members of the thesis committee

Prof. Olivier Cadot
Full Professor of Economics, University of Lausanne
Thesis co-supervisor

Prof. Marius Brülhart
Full Professor of Economics, University of Lausanne
Thesis co-supervisor

Prof. Rafael Lalive
Full Professor of Economics, University of Lausanne
Internal member of the thesis committee

Prof. Céline Carrère
Full Professor of Economics, University of Geneva
External member of the thesis committee

Université de Lausanne
Faculté des Hautes Études Commerciales

Doctorat ès sciences économiques,
mention Economie Politique

Par la présente, je certifie avoir examiné la thèse de doctorat de

Alexander HIMBERT

Sa thèse remplit les exigences liées à un travail de doctorat.
Toutes les révisions que les membres du jury et le soussigné ont demandées durant le colloque de thèse ont été prises en considération et reçoivent ici mon approbation.

Signature :  Date : 28.08.2018

Prof. Olivier CADOT
Co-directeur de thèse

Université de Lausanne
Faculté des Hautes Études Commerciales

Doctorat ès sciences économiques,
mention Economie Politique

Par la présente, je certifie avoir examiné la thèse de doctorat de

Alexander HIMBERT

Sa thèse remplit les exigences liées à un travail de doctorat.
Toutes les révisions que les membres du jury et le soussigné ont demandées
durant le colloque de thèse ont été prises en considération
et reçoivent ici mon approbation.

Signature :   Date : 28.08.2018

Prof. Marius BRÜLHART
Co-directeur de thèse

Université de Lausanne
Faculté des Hautes Études Commerciales

Doctorat ès sciences économiques,
mention Economie Politique

Par la présente, je certifie avoir examiné la thèse de doctorat de

Alexander HIMBERT

Sa thèse remplit les exigences liées à un travail de doctorat.
Toutes les révisions que les membres du jury et le soussigné ont demandées durant le colloque de thèse ont été prises en considération et reçoivent ici mon approbation.

Signature : R. Lalive   Date : 27.8.18

Prof. Rafael Lalive
Membre interne du jury

Université de Lausanne
Faculté des Hautes Études Commerciales

Doctorat ès sciences économiques,
mention Economie Politique

Par la présente, je certifie avoir examiné la thèse de doctorat de

Alexander HIMBERT

Sa thèse remplit les exigences liées à un travail de doctorat.
Toutes les révisions que les membres du jury et le soussigné ont demandées durant le colloque de thèse ont été prises en considération et reçoivent ici mon approbation.

Signature : Date : 27/08/2018

Prof. Céline CARRÈRE
Membre externe du jury

Three Essays in Regional and Urban Economics

Alexander Himbert
University of Lausanne

September, 2018

Contents

Acknowledgments

I would like to thank my thesis advisor and Doktorvater Olivier Cadot for his very generous support throughout the last five years and his impact both on my doctoral studies as well as on my career and personal development. I could not have imagined anyone better equipped to guide me through this important period of my life.

I am grateful to the thesis committee members, Céline Carrère and Rafael Lalive, for their very constructive and valuable feedback which helped me improve the quality of my thesis.

Furthermore, I would like to thank Marius Brülhart for countless hours of interesting discussions that helped me become a better economist and the patience he has shown in working with me.

I also want to thank my friend and co-author Jayson Danton for proposing a collaboration on a paper that turned into the first academic publication of my life at a point in time where I was more than unsure about continuing my PhD.

I want to thank Anne-Sophie Farines for helping me get through the inevitable crises of self-doubt most PhD students and all men of the Himbert family are familiar with.

Lastly, I wish to thank the two most important people in getting me to this point of my life: I thank my mother for always listening to me when I wanted to talk, for making me talk when I did not want to and for ensuring that I finish high-school. And I thank my father for all the things that I do exactly like he did them, for being the standard by which I measure myself and for teaching me the value of having a plan in life.

Preface

Since the 1960s, the fields of regional and urban economics have analysed economic questions that have a spatial dimension. On one hand, space and location affect the behaviour of both individuals and households as well as of enterprises and governments. On the other hand, economic decisions have an impact on the distribution of activity and people in space. In recent years, the growing availability of georeferenced data that makes it possible to analyze the interdependence of the decisions of economic agents at different locations in space has opened up new pathways for empirical research. This thesis is composed of three chapters focusing on different empirical questions of regional and urban economics. The first chapter studies the role verticality has in explaining the valuation of different locations for residential use and the possible implications of different bidding behavior at different vertical locations on questions like residential sorting and gentrification. The second and third chapter analyse the impact of trade liberalization on the distribution of economic activity in border regions. Chapter 2 measures economic activity through nighttime lights captured by satellites, allowing for an analysis on a global scale. This permits in particular the comparison of the impact of trade on border regions in developing and in industrialized countries. Chapter 3 complements this study by using georeferenced data on firm level activity in 24 countries. The use of firm level data allows the identification of cross-sectoral differences in the impact of trade on the spatial distribution of economic activity and by this sheds light on underlying causal channels.

The first chapter is written jointly with my fellow PhD student Jayson Danton. We study how the bidding behavior for residential apartments is influenced by the floor level that apartments are located at. While location has long been analyzed as one of the key factors to explain the value of residential space, locations have essentially been defined purely along the horizontal dimension. This means that the value of residential space has been analyzed in light of the region, city or neighborhood a building is located in, but within a given building, all residential units were implicitly assumed to have the same value of location. This neglects the possibility of vertical housing driving differences in rental prices within a given building. As the analysis of the value of different locations within a city has long been used as a tool to identify policy relevant patterns of residential sorting and gentrification, incorporating verticality is an important step to augment our understanding of urban dynamics.

While the literature has long been silent on the implications of vertical housing on patterns of residential prices and sorting, recent empirical work by Ahlfeldt and McMillen (2015) and Liu et al. (2015) address the relationship between verticality and locational value. Most closely to our study, Liu et al. (2015) take the first step to enrich the standard urban model by accounting for rental price patterns at a given location *within* commercial buildings, finding that pricing and spatial structure vary vertically in ways that standard urban models fail to capture. Our study complements their findings by analyzing vertical rent curves in a residential context.

In our study, we exploit variation of rental prices per square meter *within* buildings in order to identify the vertical rent curve. In order to do so, we make use of two precisely geocoded datasets recording location and characteristics of all residential buildings in Switzerland as well as rental price postings for the period of 2004-2014. This allows us to control for any time-invariant building-specific characteristics as well as for very local neighborhood characteristics that we identify through precise data on the location of each analysed building. In an extension of the vertical rent curve analysis, we explore the potentially heterogeneous equilibrium relationship between rental prices at different vertical locations in buildings and the proximity to initially high-priced areas.

We find that residential rental prices at a given horizontal location vary along a non-monotonic vertical rent curve. While households value the ease of access to the street provided exclusively by ground floor apartments by paying a premium of 1.5% to 3%, which is equivalent to the value of moving 250 to 450 m closer to the Central Business District of a city, apartments at higher vertical locations command higher prices due to amenities like views that compensate for the lack of easy access. Finally, we find an equilibrium relationship between rental prices and proximity to initially high-priced areas that is consistent with spillover effects. More precisely, the closer one is located to an initially high-priced area, the stronger the equilibrium relationship. Furthermore, when splitting the average effect across different vertical locations, we find that the equilibrium relationship is strongest for dwellings located at higher vertical locations in buildings. This finding could provide a starting point for the analysis of vertical residential sorting in future research.

Chapters 2 and 3 broaden the scope of the analysis from the urban to a regional perspective by analysing economic activity in border regions defined as locations of less than 200km distance to the nearest land border. In most countries, locations close to land borders are less economically developed than interior or coastal locations. While this phenomenon can partly be explained by external geographic or historical factors, borders themselves, by segmenting markets, act as an impediment to regional economic development. Chapters 2 and 3 aim to provide an economic mechanism to overcome to relatively low development of border regions by quantifying the effect of opening up international land borders on the economic development of border regions.

The second chapter is co-authored with Marius Brülhart and Olivier Cadot. In our study, we analyse the impact of liberalizations of international trade on the spatial distribution of economic activity in order to contribute to the debate of whether international trade helps or hinders the economic development of border regions.

In the literature, international trade has been seen both as a factor contributing to a dispersion of economic activity towards remote border areas as well as a potential force driving further concentration of activity in urban centers. From a theoretical point of view, the answer depends on whether the lower transportation costs to the foreign market that border regions benefit from are seen as stronger or weaker than the advantages urban agglomerations in the center of countries reap from potential economies of scale. In the empirical literature, a majority of papers have found international trade to contribute to a catching up of border regions, even if the results of within country studies remain more mixed. Our study offers three main extensions to the existing body of empirical research. First, we seek to quantify effects that were mostly captured only in qualitative terms in the existing quasi-experimental work. By taking measured changes in trade intensities as our explanatory variables instead of the binary before-after analyses of the Iron Curtain studies, we can compute magnitudes of border-region responses with respect to measurable magnitudes of changes in trade openness. Second, we seek to quantify effects at the border as well as gradients as one moves away from border crossing points. Third, we extend the analysis to essentially the entire world economy, allowing us in particular to explore border-region trade effects in developing countries.

In order to analyse spatial patterns of economic activity on a very fine spatial resolution while at the same time keeping the ability to draw conclusions on both developing as well as industrialized countries, we construct a dataset drawing on satellite night lights data. This provides us with a proxy of economic activity available on a resolution of about $1km^2$.

Measuring light intensity along all major cross-border road corridors over the 1995-2010 period, we detect a distinct "border shadow", whereby average light intensity progressively decreases as one gets closer to the border. The effect is robust to the inclusion of geographical controls (altitude, presence of ports and airports) and to the inclusion of region-year fixed effects to control for confounding political-economy influences. However, we also show that trade liberalization, measured by the volume of exports between the two countries straddling a border, reduces the intensity of the border shadow. Again, the effect is robust to the inclusion of an array of controls.

The third chapter is a single-authored research that expands on the analysis in chapter 2 by analysing potential mechanisms behind the redistributive effect of trade liberalization on border effect. To this end, I construct a geocoded firm level dataset based on ORBIS data. Through their headquarter address, I locate over 1.3 million firms in 24 countries, which enables me to estimate gradients of firm level employment and turnover along distance to the nearest land border similar to the "border shadows" measured in chapter 2. Again, liberalizations in international trade diminish the initial impact of remoteness induced by land borders.

Furthermore, the use of firm level data allows me to distinguish between the effects on different sectors of the economy as well as to split the effects of exports, imports excluding inputs and imports of inputs. The results draw a more nuanced picture of the effects of trade on border regions. While it can generally be said that trade facilitation attenuates the initial gradient of firm activity along distance within the border region, with firms being located more closely to the border reaping the biggest part of the gains from trade, the exact effects depend both on the sectoral composition of the border region as well as on the relative trade power compared to the neighbouring country. Border regions with a high share of manufacturing firms tend to gain least from increasing exports to the neighbour country and even risk losing employment and turnover in cases where the increase in imports due to trade facilitation outweigh the increase in exports. Furthermore, trade facilitation is unsurprisingly related to improvements in the infrastructure of border regions as reflected by increased activity of construction firms close to the border. These cross-sectoral similarities and differences could not be explored using the night-light data of Brülhart et al. (2018) and contribute to our understanding of the underlying causal channels.

The three chapters follow.

Chapter 1

Residential Vertical Rent Curves

JAYSON DANTON AND ALEXANDER HIMBERT

We estimate the equilibrium bid-rent function for vertical locations within residential buildings using geocoded rental price data. We find that the vertical rent curve is non-monotonic with respect to floor level. Specifically, residential bidders pay a 1.5 to 3.5% ground floor premium. For the remaining vertical locations, the slope of the vertical rent curve is positive with respect to floor level. Our analysis shows that the ground floor premium is indeed a valuation of access similar to the valuation of access to the Central Business District (CBD) in a monocentric setting. Based on the cities in our sample, the ground floor premium is roughly equivalent to moving between 250-450 meters closer to the CBD. The positive slope of the vertical rent curve, with respect to floor level, is predominately explained by amenities (e.g. views) that compensate residents for the lack of easy access to the street. Furthermore, our analysis provides evidence of an equilibrium relationship between rental prices and proximity to initially high-priced areas that is consistent rental price spillovers. More precisely, dwellings at higher vertical locations exhibit stronger rental price growth if they are located closer to an initially expensive neighborhood compared to dwellings on lower floor levels.

1.1 Introduction

Since its origin, a key concern of the standard model of urban economics as developed by Alonso (1964), Mills (1967) and Muth (1969) has been the analysis of rental price patterns at different locations within a metropolitan area. However, the model only allows for a very limited treatment of verticality and in essence implicitly assumes the existence of only one rental price at a given location within a city.[1] In practice, this is not the case as the possibility of vertical housing implies a distribution of rental prices at any horizontal location in the city. This paper addresses this gap by analyzing the bid-rent function for vertical locations within residential buildings. We find a convex and non-monotonic vertical rent curve. The non-monotonicity is mainly explained by the ease of access to the road that the ground floor exclusively provides.

In this paper, identification of the vertical rent curve comes from *within* building variation of rental prices per square meter. In particular, the analysis controls for any time-invariant building-specific characteristics, such as construction year, that could influence the variation in rental prices. To complement the within building identification, we include both grid-year and municipality-year fixed effects. On the one hand, the 200x200m grid-year cell fixed effects control for very local neighborhood characteristics that are time varying, e.g. improvements to the local park that makes the environment more attractive, putting upward pressure on rental prices. On the other hand, municipality-year fixed effects account for local policies at the municipal level that have an impact on rental prices, i.e. Swiss municipalities enjoy fiscal autonomy on both local public expenditures and tax revenues. These factors are likely to be capitalized into local rental prices.[2]

In an extension of the vertical rent curve analysis, we explore the potentially heterogeneous equilibrium relationship between rental prices at different vertical locations in buildings and the proximity to initially high-priced areas. To accommodate the analysis, we aggregate up to grid-cell and municipality levels in order to measure average rental price growth rates at various locations within buildings. We identify the equilibrium relationship by controlling for initial municipality-level characteristics as well as changes in the average building and dwelling quality. In other words, the controls help partial out factors that may conflate local common shocks with proximity to initially high-priced units.

Our analyses yield four main results. First, the ground floor premium is driven predominantly by upper floor access costs (lack of a lift) rather than ground floor amenities (e.g. a garden). More precisely, a building that hosts a lift service will display a greatly reduced ground floor premium compared to a building that does not host a lift service. Second, the ground floor premium is statistically significant and ranges between 1.5% and 3.5%. Our results are smaller in magnitude than the commercial building ground floor premia found in Liu, Rosenthal and Strange(2015), which range between 30-50%. This seems reasonable given that we analyze a residential setting in which access to the road is not as important as customers' access to stores in a commercial setting. In connection with the standard urban model, the ground floor premium corresponds to a consumer locating approximately 250 to 450 meters closer to the CBD. Third, the vertical rent curve is non-monotonic with respect to floor level. Intuitively, this result can be explained by amenities (e.g. views) that compensate consumers for the lack of easy access to the street, which is similar to the result in Liu et al. (2015) that produces sorting within a commercial building where access-oriented retailers bid for lower floors and amenity-oriented firms bid for upper floors. Finally, we find an equilibrium relationship between rental prices and proximity to initially high-priced areas (grid-cell or municipality) that is consistent with spillover effects. More precisely,

[1] Residential development can employ varying levels of capital-to-land at different locations within the city, which functions as a proxy for building height.

[2] See for example, Oates (1969), Black (1999), Bayer, Ferreira and McMillan (2007), Fack and Grenet (2010) or Gibbons, Machin and Silva (2013).

the closer one is located to an initially high-priced area, the stronger the equilibrium relationship. Furthermore, when splitting the average effect across different vertical locations, we find that the equilibrium relationship is strongest for dwellings located at higher vertical locations in buildings.

The remainder of the paper is structured as follows. Section 1.2 briefly covers relevant literature. Section 1.3 describes the data we use. Sections 1.4 and 1.5 explain and discuss our empirical framework and ensuing results. Finally, section 1.6 concludes.

1.2 Literature Review

In a unified analysis of spatial economic regularities Brueckner (1987) discusses patterns of urban land-use throughout real-world cities, developed in earlier work by Alonso (1964), Mills (1967) and Muth (1969). Notably, *within* cities, buildings tend to be taller towards the center compared to those found in more distant suburban areas. Rental prices also decrease as one moves further away from the city center in order to compensate commuting costs, which is also known as the horizontal rent curve. However, the standard monocentric model neglects the verticality of rental prices within cities at a given location. For a more recent survey of the literature, see Duranton and Puga (2015).

In an extension of the monocentric model, Wright (1971) shows that by introducing horizontal and vertical location directly into the consumer's utility function, one can derive both a horizontal and a vertical rent curve. The well-known horizontal rent curve is a downward sloping function of the distance to the CBD (see, among others, McMillen (1996) for empirical evidence). In other words, in order to keep utility constant throughout the monocentric city, consumers are compensated by lower rental prices at further locations from the CBD in lieu of their increased commuting costs. Gibbons and Machin (2005) document that urban residents do indeed value the ease of access to CBD, using a quasi-experimental approach to study the effect on local house prices after the construction of new railway stations in London in the 1990's.[3] Furthermore, Ahlfeldt and Wendland (2013) show that innovations in transport technology in early 20th century Berlin led to a flatter, yet still downward sloping, house price curve with respect to distance from the CBD.

The vertical rent curve is at the center stage of this paper. Without positive vertical location preferences, the vertical rent curve would be downward sloping since time spent in vertical transportation generates disutility for urban residents. However, it is reasonable to assume the existence of positive vertical location preferences, for example, through the valuation of views on higher vertical locations. Liu et al. (2015) show that tenants in commercial buildings are likely to have preferences for different vertical locations depending on the access value that lower vertical locations offer versus the amenity value generated at higher vertical locations.[4]

Recent empirical work by Ahlfeldt and McMillen (2015) and Liu et al. (2015) address two notable elements of the standard monocentric model. On the one hand, Ahlfeldt and McMillen (2015) document a robust relationship between building height and land rent. They find that developers indeed respond to higher land prices by producing taller buildings. More precisely, in 2000, they estimate the elasticity of building height with respect to land price to be approximately 45% for tall commercial structures and around 30% for residential structures. On the other hand, Liu et al. (2015) take the first step to enrich the standard urban model by accounting for rental price patterns at a given location *within* commercial buildings.[5] First, they conclude that

[3] NG (2008) shows that, in theory, differences in household preferences over amenities generate various types of horizontal location patterns in housing prices. In other words, households may be willing to accept longer commutes in order to locate closer to amenities.

[4] In related work on building height, Helsley and Strange(2008) provide a game-theoretical analysis of why there is an inherent value in a building being classified as the tallest.

[5] As noted by Wong, Chau, Yau and Cheung (2011), the real estate literature has rarely studied the variation in rental price per square meter *within* buildings.

pricing and spatial structure vary vertically in ways that standard urban models fail to capture. Second, vertical rent curves are non-monotonic due to the presence of forces driven by accessibility and amenities. These forces produce sorting within a commercial building where access-oriented retailers bid for lower floor levels and amenity-oriented firms bid for upper floor levels.

In a similar fashion, Brueckner and Rosenthal (2009) address a third important aspect of the standard monocentric model. They augment the static model to accommodate dynamic neighborhood evolution over time. Their analysis studies how the age of the build stock affects the process of gentrification. They find that due to the fact that high-income households have high demand for housing services, they tend to be drawn to areas where the housing stock is relatively young. Furthermore, they find that if spatial variation in dwelling ages were eliminated, disparities in neighborhood economic status would be reduced but up to 50 percent.

This paper contributes to the literature in various ways. First, we are among the first to study the vertical rent curve within residential buildings using detailed geocoded micro data on observed rental prices. Second, we show that the vertical rent curve exists at varying horizontal locations within the city, but becomes less pronounced at further horizontal locations from the CBD. That said, the ground floor premium may not necessarily appear at all horizontal locations within the city. Finally, we provide empirical evidence of heterogeneous equilibrium relationships between rental prices across vertical locations in buildings and proximity to initially high-priced areas. For example, we find that the average rental prices of top floor dwellings, in a given grid-cell, experience higher growth on average the closer the grid-cell is to an initially relatively more expensive grid-cell.[6] In a similar fashion, lower floor dwelling rental prices tend to experience higher growth in close proximity to the CBD. The observed equilibrium relationships would be consistent rental price spillovers driven by a gentrification mechanism.

1.3 Data

1.3.1 Residential Buildings

The first panel data set covers an exhaustive record of geocoded residential building data.[7] In particular, we observe the number of floors within a building, the building period as well as characteristics such as the number of individual housing units within the building. We restrict our sample to the five biggest cities in Switzerland, namely, Zurich, Geneva, Basel, Lausanne and Bern.[8] These city agglomerations account for over 50% of Switzerland's population.

The Swiss building stock exhibits interesting characteristics that will help put the analysis into perspective. First, by the end of 2014, the stock of residential buildings amounted to around 1,7 million buildings or around 4,3 million residential dwellings. Second, around 38% of the residential buildings are apartment blocks, hosting around 68% of Switzerland's residential population. In addition, around 56% of these apartment blocks are located in cities. Third, around 94% of all residential buildings have less than 5 floors (around 87% if one removes the buildings with only one story). Fourth, around 42% of the building stock was constructed after

[6] Guerrieri, Hartley and Hurst (2013) also use proximity to initially high-priced neighborhoods to study house price dynamics in the context of endogenous gentrification.

[7] The confidential data were kindly provided by Swiss Federal Registry of Buildings and Housing.

[8] Cities in our context are monocentric circles of 10km radius around the main train station, which acts as the CBD.

the 1980's. Finally, Switzerland is a renter-occupier nation, with around 37,4% owner-occupancy.[9] As a result, analysing rental price variation within residential buildings is particularly relevant in this setting.[10]

The analysis in this paper will be carried out on the stock of apartment block buildings that are either entirely destined for residential use or may have a mixture of residential and commercial activities within the building. First, this is important given that we want to study the shape of the residential vertical rent curve, therefore we exclude single story buildings, i.e. individual homes. Second, we include buildings that exhibit a mixture of residential and commercial activity. These types of buildings are typically found closer to the center of the city. It is worth noting that the commercial activities generally include bars, clothing stores, pharmacies, small offices and shops. Throughout our analysis we will account for the difference between purely residential building blocks and building blocks with mixed residential and commercial activity.

Panel A of Table 1.1 summarizes building level characteristics as well as additional characteristics calculated using ArcGIS. For example, the former include information on variables within the building such as building height measured in total number of floors, construction period, internal building density and building type, i.e. whether it is a building that hosts a mixture of residential and commercial activity or not. The ArcGIS information describes how a given building compares to others within a 100m radius, such as how tall the building is with respect the average building height in its vicinity, the neighborhood density (measured in number of buildings or number of dwellings). Furthermore, we include topographic information on road distance to CBD, the altitude and the slope of the terrain the building is situated on.[11] Note that the road distance to the CBD can exceed the 10km radius around the main train station.

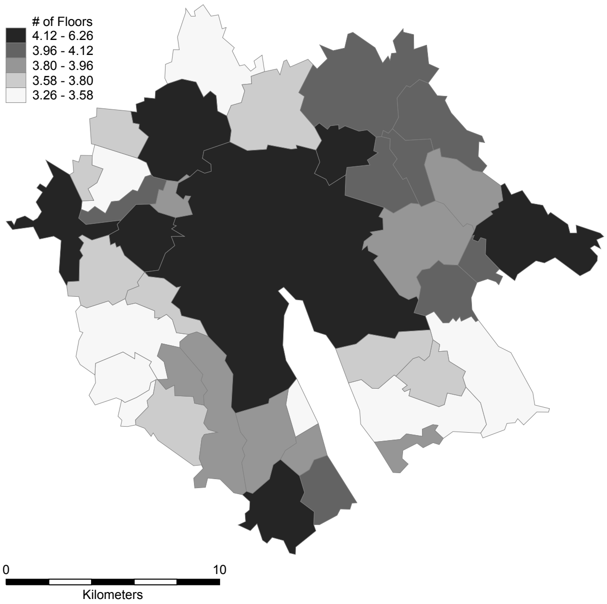
Figure 1.1 displays the average building height per municipality throughout the city of Zurich for illustrative purposes. The CBD is located in the central municipality, where we find the tallest buildings. As one moves away from the central municipality, average building height decreases, which is a well-known result of the standard urban monocentric model. This feature is driven by the decrease in land prices at further locations from the CBD. In turn, the capital-to-land ratio (building height) decreases.

[9] See the Swiss Federal Statistics Office publication *Construction and Housing* (2014).

[10] As of 2017, around 50% of rental dwellings were owned by private individuals, around 5% were owned by the state, and construction and real estate companies accounted for around 15% of rental dwelling ownership. The remainder was owned by legal entities. See the Swiss Federal Statistics Office for more information.

[11] The Digital Height Model (DHM25) data, a product of the Swiss Federal Statistics Office, are confidential and provided by Dr. Alexandre Hirzel from the Graphic Information System (GIS) department at the University of Lausanne. The raster layers contained contours of altitude and grid-cells measuring 100x100m with the average slope of the terrain.

Figure 1.1: Average Building Height (Floors) Variation across Locations in Zurich*



Note: The figure is constructed by calculating the average building height (in number of floor levels) per municipality. Dark regions represent municipalities with tall buildings, on average, and lighter regions represent municipalities with smaller buildings, on average. Gray lines represent municipal administrative borders. Produced using the command *spmap* in Stata.

1.3.2 Rental Prices

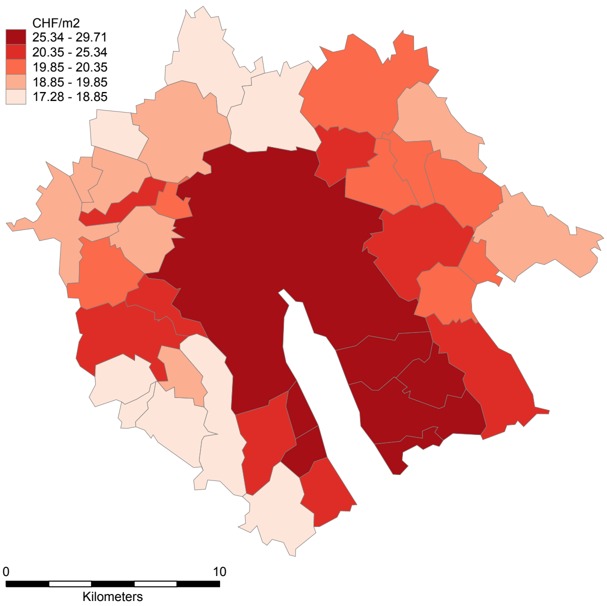
The second panel data set contains annual advertised residential rental prices spanning 2004-2014.[12] It contains the structural attributes, namely the rental price per square meter, of approximately 1,200,000 rental objects as well as their corresponding geographic coordinates. For example, the dwelling characteristics include: floor level, the number of rooms and the existence of a balcony, view, parking spot or garage. Roughly one quarter of these observations take place in the top 5 cities. The high quality geocoding ensures that the localization of the rental objects within the city provides reliable matching of rental price postings to buildings. Panel B of Table 1.1 summarizes the dwelling level characteristics.

Figure 1.2 displays the average rental price per square meter per municipality throughout the city of Zurich. The central municipality hosts the main train station, which is considered to be the CBD in our setting. Dark red zones indicate high rental prices and the lighter municipalities depict lower rental prices. Moving away from the central municipality, we notice a decreasing rental price, which is a typical feature of the standard urban monocentric model.

Panel C in Table 1.1 describes additional covariates used in the spillover analyses in section 1.5.4. The main variables measure the distances to initially high-priced units of observations, i.e. municipalities or grid-cells, depending on the level of aggregation. In addition, we include cross-sectional municipality level variables such as an index of accessibility, presence of a lake in the municipality, median annual taxable income in 2004, the 90th income percentile personal income tax rate in 2004 and the resident population in 2004.[13]

[12] The rental price data were provided by Meta-Sys, who collected postings from online internet platforms.
[13] See Parchet (2014) for more detail regarding local income taxation in Swiss municipalities.

Figure 1.2: Average Rental Price (CHF/m^2) Variation across Locations in Zurich*



Note: The figure is constructed by calculating the average (over the whole sample period 2004-2014) rental price per square meter (CHF) per municipality. Dark regions represent expensive municipalities, on average, and lighter regions represent cheaper municipalities, on average. Gray lines represent municipal administrative borders. Produced using the command *spmap* in Stata.

Table 1.1: Summary Statistics

Variable	Unit	Mean	Std. Dev.	Min.	Max.	N
***Panel A* : Building level characteristics**						
Building height	Floors	4.98	1.84	2.00	20.00	38,198
Building construction period	–	1,962.80	25.88	1,919.00	2,015.00	38,198
Internal building density	# of Dwellings	10.05	8.08	1.00	137.00	38,198
Commercial building	Binary	0.27	0.45	0	1	38,198
Building has a lift	Binary	0.43	0.50	0	1	38,198
Building has garden amenities	Binary	0.46	0.50	0	1	38,198
Building has a view at any floor level	Binary	0.53	0.50	0	1	38,198
Building higher than average in vicinity	Binary	0.53	0.50	0	1	38,198
Neighborhood building density	# of Buildings	24.88	15.01	1.00	107.00	38,198
Neighborhood dwelling density	# of Dwellings	209.00	146.62	1.00	1,162.00	38,198
Commute distance to CBD	km	4.65	2.95	0.04	15.32	38,198
Altitude	Meters	420.71	103.85	247.60	862.90	38,198
Slope of terrain	Slope Percentage	3.15	3.23	0.00	31.40	38,198
***Panel B* : Dwelling level characteristics**						
Price per square meter	CHF	22.33	7.15	10.34	57.39	270,516
Ground floor	Binary	0.18	0.38	0	1	270,516
Top floor	Binary	0.02	0.13	0	1	270,516
Floor level	–	2.05	1.76	0.00	17.00	270,516
Number of rooms	–	3.04	1.22	1.00	6.50	270,516
Surface area	Meters squared	73.69	31.96	11.00	660.00	270,516
View reported in posting	Binary	0.22	0.41	0	1	270,516
Balcony reported in posting	Binary	0.10	0.30	0	1	270,516
Access to outside parking	Binary	0.18	0.39	0	1	270,516
Access to parking in garage	Binary	0.34	0.47	0	1	270,516
***Panel C* : Additional Covariates**						
Median taxable income (2004)	CHF	52,580.42	7,373.61	38,200.00	73,800.00	143
Personal income tax rate (2004)	%	12.87	2.01	8.47	16.51	143
Resident population (2004)	–	13,971.85	38,419.30	455.00	364203.00	143
Presence of a lake in municipality	Binary	0.15	0.36	0	1	143
Index of accessibility (2004)	–	3.27	1.41	1.00	6.00	143
Distance to nearest high priced municipality	km	5.79	3.59	1.22	17.57	143
Distance to nearest high priced grid-cell	km	0.45	0.52	0.20	5.89	4,256

Panel A: The underlying variable of *Building Construction Period* is coded as follows: 8011 = ¡ 1919, 8012 = [1919 , 1945], 8013 =[1946 , 1960], 8014 = [1961 , 1970], 8015 = [1971 , 1980], 8016 = [1981 , 1985], 8017 = [1986 , 1990], 8018 = [1991 , 1995], 8019 = [1996 , 2000], 8020 = [2001 , 2005], 8021 = [2006 , 2010] and 8022 = [2011 , 2015]. The presentation in the table is for legibility purposes. The variable, *Building Age Period*, used in the regressions is created by 8022 minus the building's construction period. The variables *Building has a lift*, *Building has garden amenities* and *Building has a view at any floor level* are defined using the rental price posting data. *Panel B*: *Number of rooms* is top coded where 6.5 means 6.5 rooms or more. *Panel C*: describes additional variables that are used in the spillover exercises in section 1.5.4. Grid cells measure 200x200m in size.

1.4 Empirical Framework

Our baseline estimation strategy entails regressing the rental prices per square meter on the floor level of a dwelling. In order to allow for the potential non-monotonicity in the vertical rent curve, we include a dummy variable that takes the value one if the dwelling is on the ground floor and zero otherwise, as well as an equivalent dummy variable for the top floor.

Identification of the ground floor premium necessitates controls on three different dimensions. First, time-invariant building-specific characteristics that could influence rental price patterns over different floor levels in a building such as building age, are captured by building fixed effects. As a result, identification comes from *within* building variation of the rental price per square meter over time. Second, local rental prices are likely to be driven by neighborhood characteristics that vary over time (e.g. the neighborhood becoming more or less attractive over time). We control for very local time varying neighbourhood characteristics by using grid-cells measuring 200x200 meters as grid-year fixed effects. Third, local policy at the municipality level will affect rental prices in various ways. For example, among others, Swiss municipalities enjoy fiscal autonomy in levying and deploying tax revenues, which will capitalize into rental prices.[14] We include municipality-year fixed effects to account for these features in grid-cells that straddle municipal borders.

Inspired by Liu et al. (2015), let $\ln P_{ibgmt}$ represent the log per square meter rental price of dwelling i in building b, grid-cell g, municipality m and year t. Our baseline specification then takes on the following form:

$$\begin{aligned}\ln P_{ibgmt} = {} & \beta_1 GF_i + \beta_2 GF_i \times BH_b + \gamma_1 FL_i + \gamma_2 FL_i \times BH_b \\ & + \theta_1 TF_i + \theta_2 TF_i \times BH_b + \boldsymbol{\mu}\mathbf{Z}_i + \delta_b + \delta_{gt} + \delta_{mt} + \epsilon_{ibgmt}\,, \end{aligned} \tag{1.1}$$

where GF is a dummy variable for the dwelling being on the ground floor or not, FL is the floor level, TF is a dummy for the top floor and BH_b is the building height of building b.[15] The vector **Z** contains a variety of dwelling controls such as dwelling size in number of rooms, the presence of a balcony and/or a view and access to parking facilities. The full battery of building, grid-year and municipality-year fixed effects are δ_b, δ_{gt} and δ_{mt}, respectively. A positive β_1 coefficient should be interpreted as a premium paid for ease of access that ground floor dwellings enjoy *compared to the rental price of dwellings on all other floors in the building*. We allow for the ground floor premium to vary with building height, measured by β_2. In turn, γ_1 measures the per square meter rental price semi-elasticity with respect to floor level, which may also vary with building height (γ_2). Finally, for completeness, θ_1 accounts for a top floor premium that may be associated with dwellings located at the top of a building, which is also allowed to vary with building height (θ_2).

It is worth noting that while equation (1.1) provides clean identification of the ground floor premium, there may still be characteristics at the building-year level that affect rental prices, such as changes in ownership in parts of the building. To address this concern, we implement a within building-year (δ_{bt}) estimation given by:

$$\begin{aligned}\ln P_{ibt} = {} & \beta_1 GF_i + \beta_2 GF_i \times BH_b + \gamma_1 FL_i + \gamma_2 FL_i \times BH_b \\ & + \theta_1 TF_i + \theta_2 TF_i \times BH_b + \mu\mathbf{Z}_i + \delta_{bt} + \epsilon_{ibt}\,. \end{aligned} \tag{1.2}$$

[14] Schmidheiny(2006) shows how differences in municipal income taxation in the city of Basel influence household location choices and provide a channel for income segregation within a city. Brülhart, Bucovetsky and Schmidheiny (2015) discuss how most cities are divided into 74 municipalities, on average, which enjoy varying degrees of fiscal autonomy.

[15] Following Liu et al. (2015), we use the floor level as a continuous variable.

This particular identification strategy has a significant drawback in that approximately 35% of our estimation sample come from building-years with only one observation and therefore, would not contribute to the identification of the ground floor premium. In addition, around 40% of the remaining observations come from building-years of two or three rental postings. The likelihood of having multiple rental postings per building-year increases with building height (higher buildings contain more dwellings), which means that the within building-year specification systematically changes the share between tall and small buildings in the estimation sample.

1.5 Results

1.5.1 Baseline Results

Table 1.2 displays the results of our baseline estimation of equations (1.1) and (1.2). Estimating equation (1.1) without dwelling controls in column (1), we find a premium for ground floor dwellings of 3.3% compared to the average rental price per square meter in the rest of the building. Once dwelling controls are included in column (2), this premium drops slightly to 3.0%. This premium decreases with building height, which might reflect the disutility that ground floor dwellings receive being located near to the building's thoroughfare to dwellings on upper floors in taller buildings. However, the ground floor only starts to be cheaper than the average dwelling in buildings of five floors or more. Concerning the vertical rent curve, we find a semi-elasticity of the per square meter rental price with respect to the floor level of 1.4% without controlling for dwelling characteristics and 1.1% once dwelling characteristics are included. The coefficient on the interaction is negative but a factor of 50 times smaller than the main effect. We explore the determinants of the vertical rent curve in section 1.5.3, where the interaction loses significance. Finally, there appears to be a top floor discount in smaller buildings, which is most likely associated with increased running costs, such as heating due to the dwelling's location in the building. Note that throughout our results, the top floor interaction with building height tends to negate the underlying discount for buildings that have three or more stories.

Columns (3) and (4), report the building-year fixed effects results from the specification given by equation (1.2). The point estimates remain stable across the two columns in comparison to columns (1) and (2). The assertion that changes in ownership of dwellings or other factors that vary over time within a building-year appear to have no systematic effect on the pattern of rental price across floor levels. For reasons discussed in section 1.4 the remainder of the paper presents results using the specification as described by equation (1.1), i.e. with building fixed effects instead of building-year fixed effects.

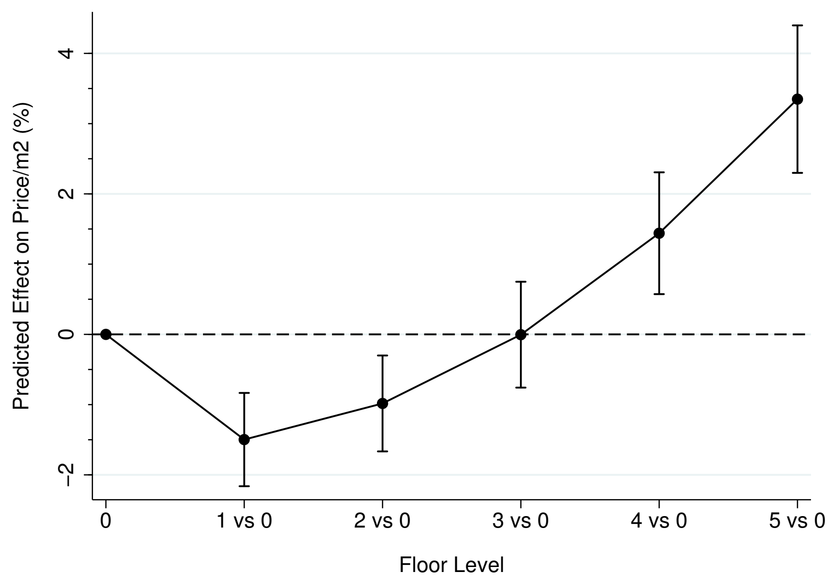
Figure 1.3 illustrates the findings regarding the non-monotonicity of the convex vertical rent curve. The figure shows that first and second floor dwellings are, on average, between 1% and 1.5% cheaper per square meter than ground floor dwellings. In addition, the amenity value of higher floors only negates the ground floor access premium from the third floor onwards. The next section investigates the driving forces behind this non-monotonic vertical rent curve.

Table 1.2: Baseline Estimations

Dependent variable:	Log Rental Price Per Square Meter			
	(1)	(2)	(3)	(4)
Ground Floor	0.033***	0.030***	0.032***	0.028***
	(0.004)	(0.003)	(0.004)	(0.004)
Ground Floor × Building Height	-0.003***	-0.006***	-0.003***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Floor Level	0.014***	0.011***	0.015***	0.013***
	(0.001)	(0.001)	(0.001)	(0.001)
Floor Level × Building Height	-0.000***	-0.000***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Top Floor	-0.010	-0.028***	-0.012	-0.026**
	(0.010)	(0.010)	(0.012)	(0.012)
Top Floor × Building Height	0.005**	0.008***	0.005*	0.007***
	(0.002)	(0.002)	(0.003)	(0.003)
Buildings	38,198	38,198	27,559	27,559
Observations	270,516	270,516	179,412	179,412
Building FE	YES	YES	NO	NO
Building-Year FE	NO	NO	YES	YES
Grid-Year FE	YES	YES	NO	NO
Municipal-Year FE	YES	YES	NO	NO
Dwelling controls	NO	YES	NO	YES

Standard errors clustered at building level in parentheses. Ground floor is a binary variable that equals one when the floor level is the ground floor, and zero otherwise. Top floor is a binary variable that equals one when the floor level is the top floor of the building and zero otherwise. Dwelling characteristics include: the number of rooms and the existence of a balcony, view, parking spot or garage (see panel B of Table 1.1 for more details). Grids measure 200x200m in size.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure 1.3: The Vertical Rent Curve



Note: Figure constructed as follows. We run a log-linear polynomial regression of order two to capture the convex shape of the vertical rent curve. We use the estimates reported in column (2) of Table 1.7.6. More precisely, we use the coefficients on *Ground Floor*, *Floor Level* and *Floor Level Squared* (and their corresponding standard errors) to produce the stylized curve relative to the ground floor, with 95% confidence bands. The graph shows that first and second floors are, on average, cheaper per square meter than the ground floor.

1.5.2 Ground Floor Premium

The existence of a ground floor premium can be broadly motivated by two factors, access value versus amenity value. We extend our analysis in order to provide empirical evidence supporting the existence of these driving factors. On the one hand, the ground floor premium could reflect the value ground floor dwellings generate from having direct access to the road, the same way retail-oriented firms in Liu et al. (2015) pay a premium for the direct access. On the other hand, ground floor dwellings in residential buildings potentially host amenities such as gardens or terraces, which are typically not available to upper floor dwellings. To determine which of these two channels (access vs. amenity) drives the ground floor premium, we employ the following modified version of our baseline specification:[16]

$$\begin{aligned}\ln P_{ibgmt} = \beta_1 GF_i + \beta_2 GF_i \times BH_b + \beta_3 GF_i \times AM_b + \beta_4 GF_i \times AC_b \\ + \beta_5 GF_i \times AM_b \times AC_b + \gamma_1 FL_i + \gamma_2 FL_i \times BH_b \\ + \theta_1 TF_i + \theta_2 TF_i \times BH_b + \boldsymbol{\mu}\mathbf{Z}_i + \delta_b + \delta_{gt} + \delta_{mt} + \epsilon_{ibgmt}\,, \end{aligned} \quad (1.3)$$

where β_3 measures the effect of having a garden or a terrace and β_4 reports the effect of the availability of a lift on the ground floor premium. Furthermore, β_5 determines which of the two forces is dominant (i.e. buildings that have both a lift and a garden). Since the sign on β_3 is expected to be positive (i.e. a garden increases the ground floor premium) and the sign on β_4 is expected to be negative (as a lift decreases the premium on the ground floor), we interpret a positive sign on β_5 as evidence of a dominant amenity effect, while a negative sign on β_5 would be evidence of a dominant access effect that drives the ground floor premium.

There are a number of takeaways from the estimation results in Table 1.3. First, the ground floor premium ranges from 1.5-3.5%, which roughly corresponds to moving between 250-450 meters closer to the CBD.[17] Second, the amenity value of a garden or a terrace increases the ground floor premium by 0.6 percentage points (p.p.). Third, the value that the ground floor dwelling generates from direct access to the road decreases by around 1.4 p.p. in buildings with a lift. In other words, as access to upper floors is made easier, the premium enjoyed by ground floor dwellings (without a garden or terrace) is greatly reduced. Fourth, the triple interaction term between the ground floor, amenity and access binary variables is negative (-0.9 p.p.). The result indicates in the presence of both a garden and a lift, the negative effect of increased accessibility to higher floors on the ground floor premium outweighs the positive effect of amenity value that a garden adds to the ground floor premium. In other words, the value of the ground floor dwelling is driven to a greater extent by accessibility than by its amenity value. Fifth, the interaction with the commercial building binary variables in practically zero, suggesting that the residential dwelling ground floor premium is not affected by the mixture of activities in the building. Sixth, building age, internal density (number of dwellings in the building) as well as the external neighborhood (within 100 meters) density do not appear to affect the ground floor premium.

Finally, column (6) relates the ground floor premium to the standard monocentric model and analyzes its evolution depending on a building's location within a city. Here, we interact the ground floor premium with the commute distance from a building to the CBD. Controlling for the distance to the CBD reduces the ground floor premium to zero at the CBD center, but it gradually increases by 0.2 p.p. per kilometer as one moves away from the CBD center. We also construct two variables capturing the neighborhood density: we control for the number of buildings and the number of rental units within a 100 meter radius of a building. Results

[16] See Table 1.7.7 for point estimate comparisons between competing specifications of column (3) from Table 1.3.

[17] See Table 1.7.8 in the appendix for results on the horizontal rent curve.

Table 1.3: Determinants of the Ground Floor Premium

Dependent variable:	Log Rental Price Per Square Meter					
	(1)	(2)	(3)	(4)	(5)	(6)
Ground Floor	0.033***	0.030***	0.026***	0.018***	0.018***	-0.003
	(0.004)	(0.003)	(0.004)	(0.005)	(0.005)	(0.007)
Ground Floor × Building Height	-0.003***	-0.006***	-0.004***	-0.004***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ground Floor × Amenity			0.006**	0.006**	0.006**	0.006**
			(0.003)	(0.003)	(0.003)	(0.003)
Ground Floor × Access			-0.014***	-0.013***	-0.013***	-0.012***
			(0.004)	(0.004)	(0.004)	(0.004)
Ground Floor × Amenity × Access			-0.009**	-0.008*	-0.008*	-0.008*
			(0.005)	(0.005)	(0.005)	(0.005)
Ground Floor × Building Age Period				0.001**	0.001**	0.002***
				(0.000)	(0.000)	(0.000)
Ground Floor × Commercial Building					-0.001	0.000
					(0.003)	(0.003)
Ground Floor × Internal Building Density (# of Dwellings)						0.000
						(0.000)
Ground Floor × Neighborhood Dwelling Density (in 100 units)						0.002
						(0.002)
Ground Floor × Neighborhood Building Density (in 10 buildings)						0.000
						(0.002)
Ground Floor × Commute Distance from CBD (km)						0.002***
						(0.000)
Floor Level	0.014***	0.011***	0.011***	0.011***	0.011***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Floor Level × Building Height	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Top Floor	-0.010	-0.028***	-0.028***	-0.028***	-0.028***	-0.028***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Top Floor × Building Height	0.005**	0.008***	0.008***	0.008***	0.008***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Buildings	38,198	38,198	38,198	38,198	38,198	38,198
Observations	270,516	270,516	270,516	270,516	270,516	270,516
Building FE	YES	YES	YES	YES	YES	YES
Grid-Year FE	YES	YES	YES	YES	YES	YES
Municipal-Year FE	YES	YES	YES	YES	YES	YES
Dwelling controls	NO	YES	YES	YES	YES	YES

Standard errors clustered at building level in parentheses. Ground floor is a binary variable that equals one when the floor level is the ground floor, and zero otherwise. Top floor is a binary variable that equals one when the floor level is the top floor of the building and zero otherwise. *Building Age Period* corresponds to how old the building is (see panel A of Table 1.1 for more details.). *Internal Building density* is measured by the number of individual dwellings in the building. *Neighborhood Density* refers the number of buildings or individual dwellings in a 100 meter radius around the given building. Dwelling characteristics include: the number of rooms and the existence of a balcony, view, parking spot or garage (see panel B of Table 1.1 for more details). Grids measure 200x200m in size.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

indicate that neither play a significant role. In summary, the effect of distance to the CBD is most likely due to the fact that ground floor dwellings in the CBD are subject to more than residential density, but also general commercial and leisure activity. Intuitively, ground floor locations may become less desirable as the number of surrounding commercial and leisure activities increase. In addition, ground floor dwellings in central locations are likely to be less desirable due to the potentially higher risk of burglary.[18]

[18] In a similar spirit, Table 1.7.9 provides results for estimations including dummies for the different cities. While the cities are heterogeneous in terms of population, crime or pollution, we find no significant differences on the ground floor premium coefficient.

1.5.3 Vertical Rent Curve

In this section, we focus on the determinants of the slope of the vertical rent curve. To do so, we allow for the slope of the curve to vary with different control variables. For simplicity, we estimate β_1 and β_2 without allowing for any other interactions. More precisely, we modify our baseline specification as follows:

$$\begin{aligned} \ln P_{ibgmt} = \beta_1 GF_i + \beta_2 GF_i \times BH_b + \gamma_1 FL_i + \gamma FL_i \times \mathbf{X}_b \\ + \theta_1 TF_i + \theta_2 TF_i \times BH_b + \boldsymbol{\mu} \mathbf{Z}_i + \delta_b + \delta_{gt} + \delta_{mt} + \epsilon_{ibgmt}\,, \end{aligned} \tag{1.4}$$

where $\mathbf{X}$ is a vector of variables expected to affect the slope of the vertical rent curve, such as building age, density, height, neighbourhood density, and topographical features.

Five broad results appear in Table 1.4. First, the ground floor premium is approximately 2.5%, which is in line with what we previously estimated. In addition, the top floor discount remains roughly around 2.5%, but disappears for buildings with three or more stories. Second, the rental price per square meter semi-elasticity with respect to the floor level stabilizes around 2% once controlling for different factors affecting the steepness of the slope of the vertical rent curve. The commercial nature of the building, availability of a view and neighborhood density (in number of dwelling units) appear to steepen the slope of the vertical rent curve, a feature that is consistent with an amenity value accompanying verticality. Point estimates indicate that a view or the fact that the building is higher than the average building within a 100m radius increases the rental price per square meter semi-elasticity with respect to floor level by 0.4, or 0.6 percentage points if both characteristics are present.

Third, the slope of the vertical rent curve gets flatter the older the building as well as when the building is densely constructed, housing many individual dwellings. The flatter curve due to building age reflects the fact that older buildings, while potentially desirable, may exhibit out-dated structural features. The urban resident in our setting will bid for higher vertical locations as long as those locations provide compensating amenity value. Densely constructed buildings are likely to be perceived as low cost, high density living quarters and no longer provide the required amenity value to compensate the lack of street access.

Fourth, the vertical rent curve flattens with every kilometer from the CBD center, i.e. steep rent curves are predicted to be in the city center. This is explained by the fact that the relative distance to the CBD captures density features of commercial and leisure activities such as noise, pollution and crime. Figure 1.4 illustrates the relative steepness of vertical rent curves depending on the distance to the CBD. The black curve represents the vertical rent curve at the core of the CBD and the light gray dotted curve represents the vertical rent curve at the CBD perimeter. As depicted in the figure, ground floor dwellings do not enjoy a premium with respect to upper floors when located at the core of the CBD. In contrast, ground floor dwellings appear to enjoy a significant premium when located toward the CBD's perimeter. Note that the upward sloping vertical rent curve still persists in each of the curves, i.e. rental prices per square meter increase as one moves up the vertical locations within a building.

Finally, column (6) in Table 1.4 incorporates an additional interaction with floor level that prepares the basis for the spillover analysis in the following section. More precisely, we interact floor level with the distance to the nearest initially high-priced grid-cell neighbor. An initially high-priced grid-cell is determined by the city-specific rental price distribution of rental prices per square meter in 2004. If the average rental price in 2004 in a given grid-cell was in the top quartile of the city-specific rental price distribution, it is attributed the status of ”initially high-priced.” With these initially high-priced grid-cells at hand, we determine, for each grid-cell,

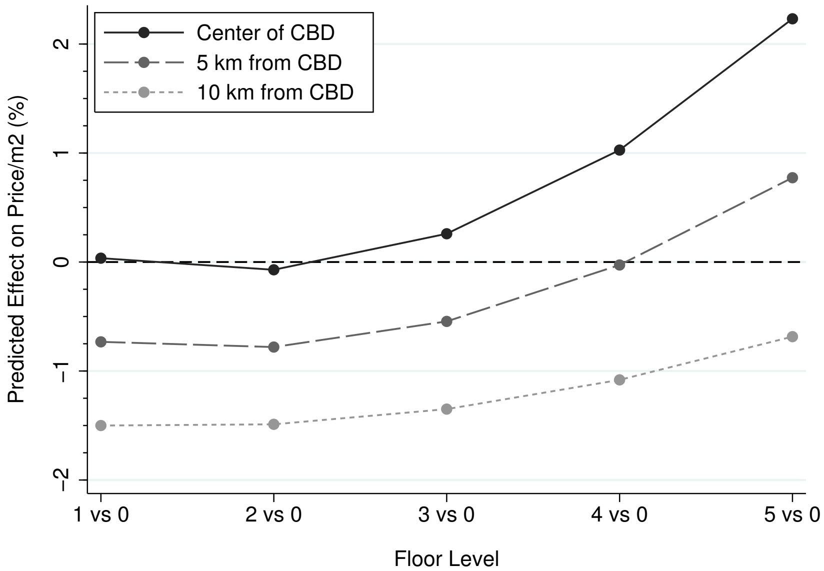
Table 1.4: Determinants of the Vertical Rent Curve

Dependent variable:	Log Rental Price Per Square Meter					
	(1)	(2)	(3)	(4)	(5)	(6)
Ground Floor	0.030***	0.034***	0.030***	0.026***	0.021***	0.021***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Ground Floor × Building Height	-0.006***	-0.006***	-0.006***	-0.005***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Top Floor	-0.028***	-0.031***	-0.029***	-0.024**	-0.023**	-0.023**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Top Floor × Building Height	0.008***	0.009***	0.009***	0.008***	0.008***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Floor Level	0.011***	0.022***	0.020***	0.018***	0.024***	0.025***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Floor Level × Building Height	-0.000***	-0.001***	-0.001***	-0.000**	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Floor Level × Building with View		0.003***	0.003***	0.004***	0.004***	0.004***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Floor Level × Building Age Period		-0.001***	-0.001***	-0.001***	-0.002***	-0.002***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Floor Level × Building Higher Than Average in Vicinity			0.002***	0.003***	0.003***	0.002***
			(0.001)	(0.001)	(0.001)	(0.001)
Floor Level × Commercial Building			0.003***	0.003***	0.001*	0.001**
			(0.001)	(0.001)	(0.001)	(0.001)
Floor Level × Internal Building Density (# of Dwellings)				-0.000***	-0.000***	-0.000***
				(0.000)	(0.000)	(0.000)
Floor Level × Neighborhood Building Density (in 10 buildings)				-0.001**	-0.001**	-0.001**
				(0.000)	(0.000)	(0.000)
Floor Level × Neighborhood Dwelling Density (in 100 units)				0.001***	0.001***	0.001***
				(0.000)	(0.000)	(0.000)
Floor Level × Altitude (in 100m)					-0.000	0.000
					(0.000)	(0.000)
Floor Level × Slope of Terrain					0.001***	0.001***
					(0.000)	(0.000)
Floor Level × Commute Distance from CBD (km)					-0.001***	-0.001***
					(0.000)	(0.000)
Floor Level × Distance to Closest Initially High Price Cell (km)						-0.002**
						(0.001)
Buildings	38,198	38,198	38,198	38,198	38,198	38,198
Observations	270,516	270,516	270,516	270,516	270,516	270,516
Building FE	YES	YES	YES	YES	YES	YES
Grid-Year FE	YES	YES	YES	YES	YES	YES
Municipal-Year FE	YES	YES	YES	YES	YES	YES
Dwelling controls	YES	YES	YES	YES	YES	YES

Standard errors clustered at building level in parentheses. The variable Log Floor is the log of floor level + 1, which removes the missing values created by log of zero. Ground floor is a binary variable that equals one when the floor level is the ground floor, and zero otherwise. Top floor is a binary variable that equals one when the floor level is the top floor of the building and zero otherwise. *Building Age Period* corresponds to how old the building is (see panel A of Table 1.1 for more details.). *Internal Building density* is measured by the number of individual dwellings in the building. *Neighborhood Density* refers the number of buildings or individual dwellings in a 100 meter radius around the given building. Dwelling characteristics include: the number of rooms and the existence of a balcony, view, parking spot or garage (see panel B of Table 1.1 for more details). Grids measure 200x200m in size. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

the nearest initially high-priced neighbor based on the euclidean distance between two grid-cells. The results indicate that being further away from an initially high-priced grid-cell flattens the slope vertical rent curve. In the next section, we explore the effect of proximity to initially high-priced grid-cells in greater detail.

Figure 1.4: The Vertical Rent Curve In The Monocentric Model*



Note: Figure constructed as follows. Combining elements from the ground floor premium and vertical curve analyzes, we run a log-linear polynomial regression of order two with all the control variables at our disposal. We recover the coefficients on *Ground Floor*, *Ground Floor* × *Commute Distance from CBD*, *Floor Level* and *Floor Level Squared* (as well as their interactions with commute distance from CBD). We then selected three commute distances from the CBD: 0km, 5km and 10km to create the vertical rent curve relative to the ground floor in each scenario. The graph shows that vertical rent curve exists at various locations within a city, but may be less pronounced at further locations from the CBD.

1.5.4 Vertical Rent Curve Spillovers

The previous sections established the non-monotonic nature of the vertical rent curve in residential housing. In this section, we aim to use our framework of vertical bidding to provide a starting point for the analysis of a potential vertical dimension in sorting and gentrification. To do so, we analyze the equilibrium relationship between rental prices across different vertical locations in buildings and proximity to initially high-priced areas. The equilibrium relationship that we identify would be consistent with rental price spillovers that are driven by neighborhood gentrification. As pointed out in Brueckner and Rosenthal (2009), understanding whether or not gentrification is a neighborhood-specific curiosity or a more systemic phenomenon that affects location decisions, is important for local policy makers and real estate developers.

The potential gentrification mechanism can be described as follows. Assuming full mobility of agents, rental prices in relatively cheaper areas are bid up by an inflow of high income households which ultimately forces the existing lower income households out. In other words, rental prices in relatively cheaper neighborhoods experience higher rental price growth rates the closer they are to initially relatively more expensive neighborhoods.

To the best of our knowledge, the empirical literature has not yet fully analyzed the potentially heterogeneous effects of gentrification at different vertical locations (see Zuk et al. (2015) for an overview). Maloutas and Spyrellis (2016) provide evidence of vertical segregation, using 2011 census data for the city of Athens by mapping the stratification of households with higher income and higher social status into higher floor levels rather than lower floors. However, their analysis abstracts from the temporal dimension of gentrification.

In a similar vein, Guerrieri et al. (2013) use a Bartik-style instrument in order to explain spillovers in housing prices between different neighborhoods. The spillovers are a result of exogenous city-wide demand shocks that cause higher income households to migrate into previously low priced neigborhoods. Our approach is closely related to Guerrieri et al. (2013), but differs on two important dimensions. First, we account for verticality in the equilibrium relationship between rental prices and proximity to initially high-priced areas. Second, we claim to identify an equilibrium relationship that would be consistent with rental price spillovers, but not the underlying gentrification mechanism that is driving the potential spillovers. Given that we do not observe household characteristics (e.g. income or education) across different vertical locations, we are unable to associate the rental price spillovers with the endogenous behavior of individual households in reaction to changes in the average behavior of their social or income group (see Manski (1993) for a discussion of the reflection problem). As a result, this section documents the existence of an equilibrium relationship between rental prices across different vertical locations and proximity to initially high-priced areas.

The context of gentrification necessitates the adaptation of our initial empirical strategy in two ways. First, gentrification is a dynamic process that describes changes in neighborhood characteristics over time. Therefore, the outcome of interest is now rental price growth rates over time at a given vertical location, instead of measuring rental prices in levels, as was done in previous sections. Secondly, gentrification is a process that takes place at the neighborhood level. In order to accommodate this important conceptual basis, we aggregate up from building level, obtaining average rental price growth rates at the grid-cell and municipality levels respectively.

Our analysis uses the following specification, which we estimate separately for different vertical locations within buildings (i.e. ground floor, lower and upper halves of buildings and top floor dwellings):

$$\Delta \ln P_{ijk} = \beta_1 dist_{ijt}^{high} + \beta_2 dist_{ijt}^{CBD} + \theta \mathbf{X}_{mjt} + \gamma \Delta \mathbf{Z}_{ijk} + \eta \ln P_{ijt} + \delta_{jk} + \epsilon_{ijk} \tag{1.5}$$

where $\Delta \ln P_{ijk} = \ln P_{ijk} - \ln P_{ijt}$ measures the percentage change in average rental prices per square meter in unit of observation $i \in \{\text{grid-cell, municipality}\}$ within city j between time $t = 2004$ and $t + k$, where $k \in \{2005, 2006, ..., 2014\}$. The variables of interest, $dist_{ijt}^{high}$ and $dist_{ijt}^{CBD}$, measure the distance (in kilometers) between the unit of observation i and the closest initially high-priced reference unit of observation (grid-cell or municipality) within city j at time t and the distance to the CBD, respectively.[19] The vector $\mathbf{X}_{mjt}$ is a set of municipality-specific controls that account for initial cross-sectional differences at time t. These are the log of median taxable income, log of local personal income tax rate, log of resident population and an index of accessibility. We further include an indicator of the presence of a lake. The vector $\Delta \mathbf{Z}_{ijk} = \mathbf{Z}_{ijk} - \mathbf{Z}_{ijt}$ contains the changes in average building and dwelling quality for each unit of observation. The changes in average building quality are proxied for by the changes in the average building height (measured in floor levels), changes in the average building density (measured by number of dwellings in the building) and changes in the average building age. The changes in dwelling quality are proxied for by the changes in the average number of rooms, changes in the average share of dwellings with balconies, changes in the average share of dwellings with views, changes in the average share of dwellings with outside parking and changes in the average share of dwellings with garage parking. In addition, we also include the initial rental price level to account for the base of the rental price growth rates. Finally, the city-year fixed effects, δ_{jk}, control for unobservables that change at the city-year level.

[19] An initially high-priced reference observation unit $i \in \{\text{municipality, grid-cell}\}$ is defined as follows. If the average rental price in 2004 in a given unit of observation i is in the top quartile of the city-specific rental price distribution, it is attributed the status of ”initially high-priced.”

Table 1.5 reports the results of the so-called spillover exercise. The results are presented in two panels: Panel A aggregates rental prices at the municipal level, while panel B uses grid-cells as unit of observation. In both panels, we present regression results with and without building and dwelling controls, which control for average quality in the unit of observation. Column (1) reports the results that pool the average rental price growth rates at any vertical location within buildings, while columns (2) to (5) present results for sub-sample analyzes when limiting our sample to ground floor dwellings, dwellings located in the lower half or upper half of a building, or top floor dwellings. We also report tests of coefficient equality between columns (2) and (5) for both $\hat{\beta}_1$ and $\hat{\beta}_2$, where the superscripts indicate the column of reference. In both panels, the results are robust to the inclusion of control variables for the changes in average rental unit quality in the municipality or grid-cell over time.[20]

Four main results arise. First, we find an equilibrium relationship between rental prices and proximity to an initially high-priced reference unit of observation. More precisely, the closer a unit of observation is located to an initially high-priced reference unit of observation, the higher the average rental price in the unit of observation. Second, when differentiating the equilibrium relationship across different vertical locations in buildings, we find that the equilibrium relationship is stronger for dwellings located at higher vertical locations. Furthermore, tests of coefficient equality indicate that the equilibrium relationship for top floor dwellings is statistically different from the equilibrium relationship for ground floor dwellings. In terms of the vertical rent curve, conditional on being a given distance from the CBD, this means that the vertical rent curve becomes steeper when in proximity of an initially high-priced reference unit of observation. Third, we also find evidence of an equilibrium relation between rental prices and proximity to the CBD. Our results indicate that the equilibrium relationship is stronger for dwellings at lower vertical locations within buildings. Finally, the results are qualitatively comparable across levels of aggregation, however, the coefficients are several times larger in magnitude when using grid-level aggregation. Therefore, respecting the appropriate level of geographic disaggregation (census tract, neighborhood, grid-cell) appears to play a key role in identifying these effects. The stark contrast in the magnitude of the estimated coefficients between grid-cell level and municipality level aggregation is likely due to attenuation bias. More precisely, municipalities tend to be larger geographic units and, as a result, potentially contain several heterogeneous neighborhoods. Therefore, increasing the size of the observation units from grid-cells to municipalities yields less representative distances between the units of observation and the reference unit of observation. This measurement error in the distance to the nearest initially high-priced reference unit will bias the estimated coefficients towards zero.

In summary, the empirical model given by equation (1.5), provides empirical evidence of an equilibrium relationship between rental prices across vertical locations within buildings and proximity to initially high-priced areas that would be consistent with residential sorting via gentrification.

[20] We also provide robustness checks for Table 1.5 in the appendix. In a robustness check, we remove the reference units of observation that determine the initially high-priced areas. See Table 1.7.10. Overall, the results are qualitatively stable.

Table 1.5: Vertical Rent Curve Spillovers

Dependent variable:	Growth in Average Log Rental Price				
Vertical Location in a building:	All	Ground Floor	Lower Half	Upper Half	Top Floor
	(1)	(2)	(3)	(4)	(5)
***Panel A*: Municipality-level aggregation**					
Without building and dwelling controls					
Distance to closest high-price municipality (km)	-0.003***	-0.001	-0.003***	-0.004***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Distance to CBD (km)	-0.005***	-0.008***	-0.007***	-0.003**	-0.001
	(0.001)	(0.002)	(0.001)	(0.001)	(0.003)
Test of equality: $\beta_1^{(2)} = \beta_1^{(5)}$ [p-value]			[0.047]		
Test of equality: $\beta_2^{(2)} = \beta_2^{(5)}$ [p-value]			[0.045]		
With building and dwelling controls					
Distance to closest high-price municipality (km)	-0.004***	-0.002**	-0.003***	-0.005***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Distance to CBD (km)	-0.005***	-0.008***	-0.006***	-0.004***	-0.002
	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)
Test of equality: $\beta_1^{(2)} = \beta_1^{(5)}$ [p-value]			[0.082]		
Test of equality: $\beta_2^{(2)} = \beta_2^{(5)}$ [p-value]			[0.091]		
Clusters	50	50	50	50	50
Observations	4,734	1,261	2,569	2,165	924
City-Year FE	YES	YES	YES	YES	YES
***Panel B*: Grid-level aggregation**					
Without building and dwelling controls					
Distance to closest high-price grid-cell (km)	-0.012***	-0.009***	-0.011***	-0.015***	-0.031***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.007)
Distance to CBD (km)	-0.018***	-0.017***	-0.018***	-0.017***	-0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Test of equality: $\beta_1^{(2)} = \beta_1^{(5)}$ [p-value]			[0.003]		
Test of equality: $\beta_2^{(2)} = \beta_2^{(5)}$ [p-value]			[0.037]		
With building and dwelling controls					
Distance to closest high-price grid-cell (km)	-0.013***	-0.010***	-0.012***	-0.014***	-0.031***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.007)
Distance to CBD (km)	-0.017***	-0.016***	-0.017***	-0.016***	-0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Test of equality: $\beta_1^{(2)} = \beta_1^{(5)}$ [p-value]			[0.003]		
Test of equality: $\beta_2^{(2)} = \beta_2^{(5)}$ [p-value]			[0.073]		
Clusters	50	50	50	50	50
Observations	83,599	22,410	55,320	28,279	3,780
City-Year FE	YES	YES	YES	YES	YES

Standard errors clustered at city-year level in parentheses. Municipal controls include: log of median taxable income, log of local personal income tax rate and log of local population. In addition, we include an index of accessibility and an indicator of the presence of a lake. All regressions include initial rental price levels as a regressor. Building controls include the changes in the average building height (measured in floor levels), the average building density (measured by number of dwellings in the building) and the average building age. Dwelling controls include the changes in the average number of rooms, the average share with balconies, the average share with views, the average share with outside parking and the average share with garage parking. (see Table 1.1 for more details on the underlying variables). Building and dwelling controls are averaged at the corresponding level of aggregation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

1.6 Conclusion

Our within building analysis documents the existence of a non-monotonic and convex vertical rent curve in the residential context, adding to Liu et al. (2015). The ground floor premium is estimated to lie between 1.5-3.5% compared to the rental prices of all other dwellings in a building, which is equivalent to moving between 250-450 meters closer to the CBD in the cities in our sample. Furthermore, we are able to disentangle different determinants of the ground floor premium and the vertical rent curve. We show that the ground floor premium can be explained predominantly as a valuation of ease of access rather than a valuation of ground floor specific amenities. In terms of the slope of the vertical rent curve, it tends to be steeper when the building is taller, has a view or is located closer to the CBD. Neighborhood density (number of rental units) around the building also tends to result in steeper rent curves.

In an extension to the vertical rent curve empirical framework, we provide empirical evidence of an equilibrium relationship between rental prices and proximity to initially high-priced areas that would be consistent with residential sorting manifesting through gentrification. The results indicate that the closer one is to initially high-priced areas, the higher the growth in average rental prices. In addition, we find that the equilibrium relationship is heterogeneous across floor levels. On the one hand, top floor dwellings tend to experience greater rental price growth, on average, than ground floor dwellings. On the other hand, lower floor dwellings are more likely to be affected by the distance to the CBD. In summary, our results indicate that residential sorting may have a vertical dimension.

This paper provides clean identification of the vertical rent curve and its determinants as well as an extension aimed at providing evidence of an equilibrium relationship between rental prices and proximity to initially high-priced area that would be consistent with rental price spillovers driven by gentrification. In order to provide an explicit description of the residential sorting mechanism behind the rental price spillovers, data at the household floor level would be necessary. Future research would attempt to trace households at this spatially disaggregated level in order to provide more explicit empirical evidence of residential sorting along the vertical dimension. In addition, we have shown empirically that verticality is an important dimension to consider regarding residential location choice in the standard monocentric model.

1.7 Supplementary Tables

Table 1.7.6: Baseline Estimations - Polynomial Specification

Dependent variable:	Log Rental Price Per Square Meter			
	(1)	(2)	(3)	(4)
Ground Floor	0.020***	0.015***	0.025***	0.020***
	(0.004)	(0.004)	(0.005)	(0.004)
Ground Floor × Building Height	-0.003***	-0.005***	-0.003***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Floor Level	0.001	-0.002	0.008***	0.005***
	(0.002)	(0.002)	(0.002)	(0.002)
Floor Level × Building Height	0.000	0.001***	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Floor Level Squared	0.002***	0.002***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Floor Level Squared × Building Height	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Top Floor	0.005	-0.013	-0.002	-0.017
	(0.010)	(0.010)	(0.012)	(0.012)
Top Floor × Building Height	0.001	0.004*	0.002	0.005*
	(0.002)	(0.002)	(0.003)	(0.003)
Buildings	38,198	38,198	27,559	27,559
Observations	270,516	270,516	179,412	179,412
Building FE	YES	YES	NO	NO
Building-Year FE	NO	NO	YES	YES
Grid-Year FE	YES	YES	NO	NO
Municipal-Year FE	YES	YES	NO	NO
Dwelling controls	NO	YES	NO	YES

Standard errors clustered at building level in parentheses. Ground floor is a binary variable that equals one when the floor level is the ground floor, and zero otherwise. Top floor is a binary variable that equals one when the floor level is the top floor of the building and zero otherwise. Dwelling characteristics include: the number of rooms and the existence of a balcony, view, parking spot or garage (see panel B of Table 1.1 for more details). Grids measure 200x200m in size. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1.7.7: Determinants of the Ground Floor Premium - Alternative Specifications

Dependent variable:	Log Rental Price Per Square Meter		
	(1)	(2)	(3)
Ground Floor	0.026***	0.011***	0.022***
	(0.004)	(0.004)	(0.005)
Ground Floor × Building Height	-0.004***	-0.002***	-0.001
	(0.001)	(0.001)	(0.001)
Ground Floor × Amenity	0.006**	0.006**	0.006**
	(0.003)	(0.003)	(0.003)
Ground Floor × Access	-0.014***	-0.014***	-0.015***
	(0.004)	(0.004)	(0.004)
Ground Floor × Amenity × Access	-0.009**	-0.010**	-0.009**
	(0.005)	(0.005)	(0.005)
Floor Level	0.011***	-0.002	
	(0.001)	(0.002)	
Floor Level × Building Height	-0.000***	0.001***	
	(0.000)	(0.000)	
Floor Level Squared		0.002***	
		(0.000)	
Floor Level Squared × Building Height		-0.000***	
		(0.000)	
Top Floor	-0.028***	-0.013	-0.028***
	(0.010)	(0.010)	(0.010)
Top Floor × Building Height	0.008***	0.004*	0.010***
	(0.002)	(0.002)	(0.002)
Log Floor			0.015***
			(0.003)
Log Floor × Building Height			0.002***
			(0.000)
Buildings	38,198	38,198	38,198
Observations	270,516	270,516	270,516
Building FE	YES	YES	YES
Grid-Year FE	YES	YES	YES
Municipal-Year FE	YES	YES	YES
Dwelling controls	YES	YES	YES

Standard errors clustered at building level in parentheses. The variable Log Floor is the log of floor level + 1, which alleviates the issue of missing values created by log of zero. Ground floor is a binary variable that equals one when the floor level is the ground floor, and zero otherwise. Top floor is a binary variable that equals one when the floor level is the top floor of the building and zero otherwise. Dwelling characteristics include: the number of rooms and the existence of a balcony, view, parking spot or garage (see panel B of Table 1.1 for more details). Grids measure 200x200m in size. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1.7.8: Horizontal Rent Curve

Dependent variable:	Log Rental Price Per Square Meter	
	(1)	(2)
Road Distance to Center	-0.055***	-0.073***
	(0.011)	(0.014)
Road Distance to Center Squared		0.002**
		(0.001)
Municipalities	161	161
Observations	270,516	270,516
Municipality FE	YES	YES
City-year FE	YES	YES
Dwelling controls	YES	YES

Standard errors clustered at municipality level in parentheses. Dwelling characteristics include: the number of rooms and the existence of a balcony, view, parking spot or garage (see panel B of Table 1.1 for more details). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1.7.9: Baseline Estimations - Ground floor premium across cities

Dependent variable:	Log Rental Price Per Square Meter		
	(1)	(2)	(3)
Ground Floor	0.033***	0.030***	0.039***
	(0.004)	(0.003)	(0.008)
Ground Floor × Building Height	-0.003***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.001)
Ground Floor × Lausanne (dummy)			-0.009
			(0.008)
Ground Floor × Bern (dummy)			-0.002
			(0.008)
Ground Floor × Basel (dummy)			-0.010
			(0.007)
Ground Floor × Zurich (dummy)			-0.011
			(0.007)
Floor Level	0.014***	0.011***	0.011***
	(0.001)	(0.001)	(0.001)
Floor Level × Building Height	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Top Floor	-0.010	-0.027***	-0.027***
	(0.010)	(0.010)	(0.010)
Top Floor × Building Height	0.005**	0.008***	0.008***
	(0.002)	(0.002)	(0.002)
Buildings	38,198	38,198	38,198
Observations	270,516	270,516	270,516
Building FE	YES	YES	YES
Grid-Year FE	YES	YES	YES
Municipal-Year FE	YES	YES	YES
Dwelling controls	NO	YES	YES

Standard errors clustered at building level in parentheses. Ground floor is a binary variable that equals one when the floor level is the ground floor, and zero otherwise. Top floor is a binary variable that equals one when the floor level is the top floor of the building and zero otherwise. Dwelling characteristics include: the number of rooms and the existence of a balcony, view, parking spot or garage. Grids measure 200x200m in size. For cities, Geneva is chosen as reference city since it has the least rental price observations of all considered cities. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1.7.10: Vertical Rent Curve Spillovers - Excluding Reference Units

Dependent variable:	Growth in Average Log Rental Price				
Vertical Location in a building:	All	Ground Floor	Lower Half	Upper Half	Top Floor
	(1)	(2)	(3)	(4)	(5)
***Panel A*: Municipality-level aggregation**					
*Without **dwelling and building controls***					
Distance to closest high-price municipality (km)	-0.003***	-0.002	-0.002***	-0.004***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
Distance to CBD (km)	-0.004***	-0.007***	-0.005***	-0.002	-0.002
	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)
Test of equality: $\beta_1^{(2)} = \beta_1^{(5)}$ [p-value]			[0.039]		
Test of equality: $\beta_2^{(2)} = \beta_2^{(5)}$ [p-value]			[0.201]		
With building and dwelling controls					
Distance to closest high-price municipality (km)	-0.003***	-0.002	-0.002***	-0.004***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
Distance to CBD (km)	-0.004***	-0.007***	-0.004***	-0.002	-0.002
	(0.001)	(0.002)	(0.001)	(0.002)	(0.004)
Test of equality: $\beta_1^{(2)} = \beta_1^{(5)}$ [p-value]			[0.065]		
Test of equality: $\beta_2^{(2)} = \beta_2^{(5)}$ [p-value]			[0.313]		
Clusters	50	50	50	50	50
Observations	3,778	1,004	2,040	1,738	751
City-Year FE	YES	YES	YES	YES	YES
***Panel B*: Grid-level aggregation**					
Without building and dwelling controls					
Distance to closest high-price grid-cell (km)	-0.013***	-0.011***	-0.012***	-0.013***	-0.025***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.008)
Distance to CBD (km)	-0.017***	-0.016***	-0.017***	-0.016***	-0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Test of equality: $\beta_1^{(2)} = \beta_1^{(5)}$ [p-value]			[0.106]		
Test of equality: $\beta_2^{(2)} = \beta_2^{(5)}$ [p-value]			[0.129]		
With building and dwelling controls					
Distance to closest high-price grid-cell (km)	-0.012***	-0.011***	-0.012***	-0.012***	-0.026***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.008)
Distance to CBD (km)	-0.017***	-0.016***	-0.017***	-0.017***	-0.012***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Test of equality: $\beta_1^{(2)} = \beta_1^{(5)}$ [p-value]			[0.085]		
Test of equality: $\beta_2^{(2)} = \beta_2^{(5)}$ [p-value]			[0.226]		
Clusters	50	50	50	50	50
Observations	66,919	17,852	44,137	22,782	2,996
City-Year FE	YES	YES	YES	YES	YES

Standard errors clustered at city-year level in parentheses. Municipal controls include: log of median taxable income, log of local personal income tax rate and log of local population. In addition, we include an index of accessibility and an indicator of the presence of a lake. All regressions include initial rental price levels as a regressor. Building controls include the changes in the average building height (measured in floor levels), the average building density (measured by number of dwellings in the building) and the average building age. Dwelling controls include the changes in the average number of rooms, the average share with balconies, the average share with views, the average share with outside parking and the average share with garage parking. (see Table 1.1 for more details on the underlying variables.). Building and dwelling controls are averaged at the corresponding level of aggregation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Not-for-publication appendices

1.8 Replication Tables - 100x100m Grid-cells

Table 1.8.11: Baseline Estimations

Dependent variable:	Log Rental Price Per Square Meter			
	(1)	(2)	(3)	(4)
Ground Floor	0.030***	0.027***	0.032***	0.028***
	(0.004)	(0.004)	(0.004)	(0.004)
Ground Floor × Building Height	-0.003***	-0.005***	-0.003***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Floor Level	0.013***	0.012***	0.015***	0.013***
	(0.001)	(0.001)	(0.001)	(0.001)
Floor Level × Building Height	-0.000***	-0.000***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Top Floor	-0.006	-0.023**	-0.012	-0.026**
	(0.012)	(0.011)	(0.012)	(0.012)
Top Floor × Building Height	0.004*	0.007***	0.005*	0.007***
	(0.003)	(0.003)	(0.003)	(0.003)
Buildings	34,907	34,907	27,559	27,559
Observations	246,838	246,838	179,412	179,412
Building FE	YES	YES	NO	NO
Building-Year FE	NO	NO	YES	YES
Grid-Year FE	YES	YES	NO	NO
Municipal-Year FE	YES	YES	NO	NO
Dwelling controls	NO	YES	NO	YES

Standard errors clustered at building level in parentheses. Ground floor is a binary variable that equals one when the floor level is the ground floor, and zero otherwise. Top floor is a binary variable that equals one when the floor level is the top floor of the building and zero otherwise. Dwelling characteristics include: the number of rooms and the existence of a balcony, view, parking spot or garage (see panel B of Table 1.1 for more details). Grids measure 100x100m in size. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1.8.12: Determinants of the Ground Floor Premium

Dependent variable:	Log Rental Price Per Square Meter					
	(1)	(2)	(3)	(4)	(5)	(6)
Ground Floor	0.030***	0.027***	0.023***	0.018***	0.017***	-0.005
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.007)
Ground Floor × Building Height	-0.003***	-0.005***	-0.003***	-0.003***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ground Floor × Amenity			0.004	0.005	0.005	0.005
			(0.003)	(0.003)	(0.003)	(0.003)
Ground Floor × Access			-0.015***	-0.014***	-0.014***	-0.013***
			(0.004)	(0.004)	(0.004)	(0.004)
Ground Floor × Amenity × Access			-0.009*	-0.008	-0.008	-0.008
			(0.005)	(0.005)	(0.005)	(0.005)
Ground Floor × Building Age Period				0.001*	0.001*	0.001***
				(0.000)	(0.001)	(0.001)
Ground Floor × Commercial Building					-0.003	-0.002
					(0.004)	(0.004)
Ground Floor × Internal Building Density (# of Dwellings)						0.000
						(0.000)
Ground Floor × Neighborhood Dwelling Density (in 100 units)						0.001
						(0.002)
Ground Floor × Neighborhood Building Density (in 10 buildings)						0.001
						(0.002)
Ground Floor × Commute Distance from CBD (km)						0.002***
						(0.000)
Floor Level	0.013***	0.012***	0.011***	0.011***	0.011***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Floor Level × Building Height	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Top Floor	-0.006	-0.023**	-0.024**	-0.024**	-0.024**	-0.024**
	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Top Floor × Building Height	0.004*	0.007***	0.007***	0.007***	0.007***	0.007***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Buildings	34,907	34,907	34,907	34,907	34,907	34,907
Observations	246,838	246,838	246,838	246,838	246,838	246,838
Building FE	YES	YES	YES	YES	YES	YES
Grid-Year FE	YES	YES	YES	YES	YES	YES
Municipal-Year FE	YES	YES	YES	YES	YES	YES
Dwelling controls	NO	YES	YES	YES	YES	YES

Standard errors clustered at building level in parentheses. Ground floor is a binary variable that equals one when the floor level is the ground floor, and zero otherwise. Top floor is a binary variable that equals one when the floor level is the top floor of the building and zero otherwise. *Building Age Period* corresponds to how old the building is (see panel A of Table 1.1 for more details.). *Internal Building density* is measured by the number of individual dwellings in the building. *Neighborhood Density* refers the number of buildings or individual dwellings in a 100 meter radius around the given building. Dwelling characteristics include: the number of rooms and the existence of a balcony, view, parking spot or garage (see panel B of Table 1.1 for more details). Grids measure 100x100m in size.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1.8.13: Determinants of the Vertical Rent Curve

Dependent variable:	Log Rental Price Per Square Meter					
	(1)	(2)	(3)	(4)	(5)	(6)
Ground Floor	0.027***	0.031***	0.027***	0.023***	0.018***	0.018***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Ground Floor × Building Height	-0.005***	-0.006***	-0.005***	-0.005***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Top Floor	-0.023**	-0.027**	-0.025**	-0.021*	-0.020*	-0.020*
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Top Floor × Building Height	0.007***	0.008***	0.008***	0.007***	0.007***	0.007***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Floor Level	0.012***	0.022***	0.020***	0.018***	0.024***	0.024***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Floor Level × Building Height	-0.000***	-0.001***	-0.001***	-0.000*	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Floor Level × Building with View		0.003***	0.003***	0.003***	0.004***	0.004***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Floor Level × Building Age Period		-0.001***	-0.001***	-0.001***	-0.002***	-0.002***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Floor Level × Building Higher Than Average in Vicinity			0.002*	0.002**	0.002**	0.002**
			(0.001)	(0.001)	(0.001)	(0.001)
Floor Level × Commercial Building			0.003***	0.002***	0.001	0.001*
			(0.001)	(0.001)	(0.001)	(0.001)
Floor Level × Internal Building Density (# of Dwellings)				-0.000***	-0.000***	-0.000***
				(0.000)	(0.000)	(0.000)
Floor Level × Neighborhood Building Density (in 10 buildings)				-0.001**	-0.001**	-0.001**
				(0.000)	(0.000)	(0.000)
Floor Level × Neighborhood Dwelling Density (in 100 units)				0.001***	0.001***	0.001***
				(0.000)	(0.000)	(0.000)
Floor Level × Altitude (in 100m)					0.000	0.000
					(0.000)	(0.000)
Floor Level × Slope of Terrain					0.001***	0.001***
					(0.000)	(0.000)
Floor Level × Commute Distance from CBD (km)					-0.001***	-0.001***
					(0.000)	(0.000)
Floor Level × Distance to Closest Initially High Price Cell (km)						-0.001
						(0.001)
Buildings	34,907	34,907	34,907	34,907	34,907	34,907
Observations	246,838	246,838	246,838	246,838	246,838	246,838
Building FE	YES	YES	YES	YES	YES	YES
Grid-Year FE	YES	YES	YES	YES	YES	YES
Municipal-Year FE	YES	YES	YES	YES	YES	YES
Dwelling controls	YES	YES	YES	YES	YES	YES

Standard errors clustered at building level in parentheses. The variable Log Floor is the log of floor level + 1, which removes the missing values created by log of zero. Ground floor is a binary variable that equals one when the floor level is the ground floor, and zero otherwise. Top floor is a binary variable that equals one when the floor level is the top floor of the building and zero otherwise. *Building Age Period* corresponds to how old the building is (see panel A of Table 1.1 for more details.). *Internal Building density* is measured by the number of individual dwellings in the building. *Neighborhood Density* refers the number of buildings or individual dwellings in a 100 meter radius around the given building. Dwelling characteristics include: the number of rooms and the existence of a balcony, view, parking spot or garage (see panel B of Table 1.1 for more details). Grids measure 100x100m in size. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

1.9 Spillover exercise with control coefficients

Table 1.9.14: Vertical Rent Curve Spillovers - Panel A with controls

Dependent variable:	Growth in Average Log Rental Price				
Vertical Location in a building:	All	Ground Floor	Lower Half	Upper Half	Top Floor
	(1)	(2)	(3)	(4)	(5)
***Panel A*: Municipality-level aggregation**					
Distance to closest high-price municipality (km)	-0.004***	-0.002**	-0.003***	-0.005***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Distance to CBD (km)	-0.005***	-0.008***	-0.006***	-0.004***	-0.002
	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)
Log of initial rental price level	-0.613***	-0.627***	-0.637***	-0.585***	-0.497***
	(0.032)	(0.048)	(0.037)	(0.041)	(0.060)
log of median taxable income	0.175***	0.248***	0.208***	0.121***	0.068
	(0.031)	(0.043)	(0.042)	(0.035)	(0.070)
log of local personal income tax rate	-0.143***	-0.059	-0.067	-0.263***	-0.263**
	(0.044)	(0.070)	(0.054)	(0.067)	(0.118)
log of resident population	0.022***	0.013***	0.016***	0.031***	0.029***
	(0.003)	(0.005)	(0.004)	(0.005)	(0.010)
Index of accessibility	-0.010***	-0.003	-0.006***	-0.014***	-0.014**
	(0.002)	(0.003)	(0.002)	(0.003)	(0.007)
Presence of a lake	0.086***	0.075***	0.080***	0.090***	0.094***
	(0.011)	(0.014)	(0.011)	(0.013)	(0.023)
Changes in the average number of rooms	-0.008***	-0.004	-0.005	-0.011***	-0.013***
	(0.003)	(0.005)	(0.004)	(0.003)	(0.003)
Changes in the average share of dwellings with balconies	-0.061***	-0.040*	-0.062***	-0.052***	-0.068***
	(0.011)	(0.022)	(0.011)	(0.016)	(0.019)
Changes in the average share of dwellings with views	0.053***	0.056***	0.045***	0.062***	0.076***
	(0.010)	(0.016)	(0.015)	(0.014)	(0.020)
Changes in the average share of dwellings with outside parking	-0.003	0.021	0.010	-0.018	-0.015
	(0.010)	(0.019)	(0.015)	(0.013)	(0.019)
Changes in the average share of dwellings with garage parking	-0.013	-0.029	-0.019	-0.011	-0.007
	(0.010)	(0.020)	(0.014)	(0.012)	(0.017)
Changes in the average building height	0.183**	0.019	0.144*	0.222**	0.001
	(0.083)	(0.083)	(0.078)	(0.106)	(0.186)
Changes in the average building density	-0.028*	-0.014	-0.036*	-0.020	0.014
	(0.015)	(0.025)	(0.019)	(0.022)	(0.056)
Changes in the average building age	0.001	0.001	-0.002	0.012	0.028
	(0.012)	(0.016)	(0.011)	(0.016)	(0.031)
Clusters	50	50	50	50	50
Observations	4,734	1,261	2,569	2,165	924
City-Year FE	YES	YES	YES	YES	YES

Standard errors clustered at city-year level in parentheses. Municipal controls include: log of median taxable income, log of local personal income tax rate and log of local population. In addition, we include an index of accessibility and an indicator of the presence of a lake. All regressions include initial rental price levels as a regressor. Building controls include the changes in the average building height (measured in floor levels), the average building density (measured by number of dwellings in the building) and the average building age. Dwelling controls include the changes in the average number of rooms, the average share with balconies, the average share with views, the average share with outside parking and the average share with garage parking. (see Table 1.1 for more details on the underlying variables). Building and dwelling controls are averaged at the corresponding level of aggregation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1.9.15: Vertical Rent Curve Spillovers - Panel B with controls

Dependent variable:	Growth in Average Log Rental Price				
Vertical Location in a building:	All	Ground Floor	Lower Half	Upper Half	Top Floor
	(1)	(2)	(3)	(4)	(5)
***Panel B*: Grid-level aggregation**					
Distance to closest high-price grid-cell (km)	-0.013***	-0.010***	-0.012***	-0.014***	-0.031***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.007)
Distance to CBD (km)	-0.017***	-0.016***	-0.017***	-0.016***	-0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Log of initial rental price level	-0.519***	-0.512***	-0.514***	-0.531***	-0.615***
	(0.012)	(0.014)	(0.011)	(0.015)	(0.026)
log of median taxable income	0.115***	0.171***	0.130***	0.112***	0.087
	(0.020)	(0.032)	(0.022)	(0.023)	(0.060)
log of local personal income tax rate	-0.073**	-0.064	-0.054	-0.077*	-0.192*
	(0.036)	(0.047)	(0.037)	(0.044)	(0.101)
log of resident population	-0.003	-0.006	-0.005	0.003	0.010
	(0.003)	(0.004)	(0.003)	(0.004)	(0.007)
Index of accessibility	-0.001	-0.006**	-0.001	-0.002	-0.009**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Presence of a lake	0.082***	0.081***	0.082***	0.087***	0.085***
	(0.010)	(0.012)	(0.010)	(0.011)	(0.021)
Changes in the average number of rooms	-0.016***	-0.016***	-0.016***	-0.015***	-0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Changes in the average share of dwellings with balconies	-0.018***	-0.008***	-0.013***	-0.023***	-0.011
	(0.002)	(0.002)	(0.002)	(0.003)	(0.008)
Changes in the average share of dwellings with views	0.036***	0.020***	0.027***	0.040***	0.040***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.007)
Changes in the average share of dwellings with outside parking	0.014***	0.021***	0.018***	0.006**	0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.007)
Changes in the average share of dwellings with garage parking	0.011***	0.008***	0.008***	0.018***	0.010
	(0.002)	(0.003)	(0.002)	(0.002)	(0.007)
Changes in the average building height	-0.028***	-0.027*	-0.023***	-0.040***	-0.020
	(0.009)	(0.014)	(0.009)	(0.011)	(0.028)
Changes in the average building density	0.003***	-0.004	0.001	0.005***	0.014
	(0.001)	(0.003)	(0.001)	(0.001)	(0.009)
Changes in the average building age	0.010***	0.009***	0.009***	0.011***	0.011*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)
Clusters	50	50	50	50	50
Observations	83,599	22,410	55,320	28,279	3,780
City-Year FE	YES	YES	YES	YES	YES

Standard errors clustered at city-year level in parentheses. Municipal controls include: log of median taxable income, log of local personal income tax rate and log of local population. In addition, we include an index of accessibility and an indicator of the presence of a lake. All regressions include initial rental price levels as a regressor. Building controls include the changes in the average building height (measured in floor levels), the average building density (measured by number of dwellings in the building) and the average building age. Dwelling controls include the changes in the average number of rooms, the average share with balconies, the average share with views, the average share with outside parking and the average share with garage parking. (see Table 1.1 for more details on the underlying variables). Building and dwelling controls are averaged at the corresponding level of aggregation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Chapter 2

Let There Be Light: Trade and the Development of Border Regions

MARIUS BRÜLHART, OLIVIER CADOT AND ALEXANDER HIMBERT

Does international trade help or hinder the economic development of border regions? We investigate this question by estimating how changes in bilateral trade volumes affect economic activity along roads running inland from international land borders. Our measure of economic activity is satellite night-light intensity on and around 2,186 border-crossing roads in 138 countries. We observe a significant border shadow: on average, lights are 37 percent dimmer at the border than 200 kilometers inland. This difference is reduced by trade expansion as measured by exports and instrumented with tariffs on the opposite side of the border. We find that, at the mean, a doubling of exports to a particular neighbor country reduces the gradient of light from the border by some 23 percent. This qualitative finding applies to developed and developing countries, and to rural and urban border regions. Proximity to cities on either side of the border amplifies the effects of trade. We provide evidence that local export-oriented production is a significant mechanism behind the observed effects.

2.1 Introduction

In most countries, locations close to land borders are less economically developed than interior or coastal locations. Border regions literally are darker: night lights captured by satellites are on average 37 percent less intense at land borders than 200 road kilometers inland. Such “border shadows” are both a cause and a consequence of national boundaries. On the one hand, country borders typically run through naturally inhospitable regions such as mountain ranges or deserts, and on the other hand borders themselves, by segmenting markets, act as an impediment to regional economic development. In this paper, we aim to explore the latter phenomenon by quantifying the effect of opening up international land borders on the economic development of border regions.

The relative underdevelopment of border regions is a regularity observed in countries across all levels of income. The stakes are likely to be highest, however, in developing countries, where unequal spatial development can generate resentment and predation among local populations. Lack of development is then not just an economic problem but a political one as well: developing-country border areas are particularly prone armed conflict (e.g. Myanmar, Uganda, DR Congo, Nigeria, Colombia and Paraguay). In the most nefarious configuration, ‘artificial’ colonial-era borders divide ethnic homelands in low-income countries. Michalopoulos and Papioannou (2016) find that African ethnicities partitioned by a border experience an 57 percent higher incidence of violence than non-partitioned ethnicities.

We show that the opening of cross-border trade can contribute to the economic development of border areas. In light of the recent literature on civil conflict, this hitherto unexplored non-traditional gain from trade liberalization may be of importance not only economically but also in broader political and societal terms.

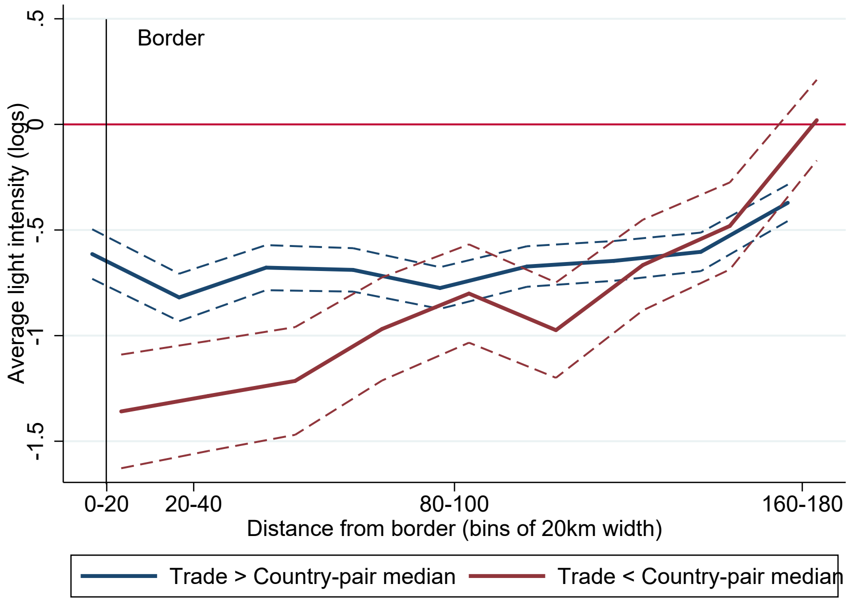
Economic activity is generally less precisely recorded at the sub-national than at the national level, especially in developing countries. As initially demonstrated by Henderson, Storeygard and Weil (2012), this potentially severe measurement problem can be overcome by drawing on satellite night lights data. We follow this approach and test how cross-border trade affects light gradients with respect to distance from the border.

Measuring light intensity along all major cross-border road corridors over the 1995-2010 period, we detect a distinct border shadow, whereby average light intensity progressively decreases as one gets closer to the border. The effect is robust to the inclusion of geographical controls (altitude, presence of ports and airports) and to the inclusion of region-year fixed effects to control for confounding political-economy influences. However, we also show that trade liberalization, measured by the volume of exports between the two countries straddling a border, reduces the intensity of the border shadow. Again, the effect is robust to the inclusion of an array of controls.

Figure 2.1 illustrates our main result. To construct this graph, we pool our observations into 20-kilometer bins in terms of distance from the border. We then split the sample into “high trade” (blue line) and “low trade” (red line) years. For each group we estimate our baseline empirical model with these bins as categorical explanatory variables. The graph plots point estimates and confidence intervals for all distance bins up to 200 kilometers from the border, with the most distant bin taken as the reference group.

Two significantly different gradients emerge: the border shadow is visibly stronger in low-trade years than in high-trade years, and this up to a distance of some 150 kilometers. Given the construction of this graph, the observed difference in border shadows is identified through time variation alone and therefore cannot be attributed to natural features that facilitate both cross-border trade and border-region development. In what

Figure 2.1: Within-road light gradients as a function of trade intensity



Note: The graph represents the point estimates for a within-road regression according to equation (3.1), where distance is expressed as a set of dummy variables for bins of 20 km width. The sample is split into "high trade" (blue line) and "low trade" (red line) years. A road-year observation is counted as high-trade (low-trade) if overall exports from country i to country j in this year were above (below) the sample average of exports between the two countries. Estimations based on road-cells only. Dashed lines are 95% confidence intervals.

follows, we describe in detail how this finding is obtained, how robust it is, and what mechanisms likely underpin it.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature, Section 3 describes the data, Section 4 discusses estimation issues, Section 5 presents baseline results, Section 6 presents extensions, and Section 7 concludes.

2.2 Literature background

2.2.1 Theory

Within-country spatial effects of external trade liberalization have been modeled in a number of recent theoretical contributions, which have yielded tractable quantitative spatial models with rich underlying geographies (Allen and Arkolakis, 2014; Atkin and Donaldson, 2015; Cosar and Fajgelbaum, 2016; Fajgelbaum and Redding, 2014; Redding, 2016; Rossi-Hansberg, 2005).[1] In these models, market access typically is only one of several determinants of regional economic activity, combining with exogenously given features such as immo-

[1] For a survey of this literature, see Redding and Rossi-Hansberg (2017). Earlier theoretical approaches included 'urban systems' models, featuring unique equilibria in perfectly competitive settings (e.g. Henderson, 1982; Rauch, 1991), and 'new economic geography' models featuring imperfectly competitive settings with multiple equilibria (e.g. Krugman and Livas Elizondo, 1996; Monfort and Nicolini, 2000). Both of those modeling approaches are compatible with trade liberalization either increasing or decreasing economic activity in border regions. In urban-systems models, this essentially depends on whether border regions are specialized in comparative-advantage or comparative-disadvantage sectors; whereas in new economic geography settings it is assumptions on the size of regions and strength of agglomeration economies that determine whether cross-border liberalization will end up drawing activity toward the border or pushing it further inland. These theories therefore do not offer any clear predictions on the impact of trade liberalization on the economic development of border regions.

bile factor endowments, productivity levels and/or amenities. Hence, even if better market access is associated with greater economic activity *ceteris paribus*, the disadvantages of border regions in terms of overall market access could be offset by advantages in terms of other locational determinants, thus making border shadows a likely but not necessarily pervasive phenomenon.

Improved market access acts as a potential boon for regional economies in all these models. Specifically, all spatial general-equilibrium models known to us that feature heterogeneous regions and labor mobility within countries can generate disproportionate economic growth in border regions as a consequence of external trade liberalization – where we associate 'border regions' with relatively low transport costs to the international border. The nature of the response depends on the specifics of the model. In counterfactual simulations of falling crossborder trade costs, border regions experience a combination of higher employment, higher real wages, higher consumer surplus and/or higher land rents.

However, alternative configurations are also possible. In Rossi-Hansberg (2005), for example, trade liberalization can change the sectoral specialization of border regions. Depending on the relative labor intensities of sectors, this may draw labor toward or away from the border region. Moreover, effects may be heterogeneous across different border regions. Redding (2016) and Redding and Rossi-Hansberg (2017) simulate multi-location models with rich geographies in which falling trade costs generate additional activity in some border regions but not in others. Consideration of intra-national transport costs makes the effective gain in market access associated with a certain drop in the cross-border trade cost unequal along the border, depending on the size of the adjacent market in the neighboring country. In our estimations, we therefore take account of urbanization patterns on either side of the border. General-equilibrium effects furthermore imply that trade liberalization triggers a reallocation of activity among different segments of the border region, such that the gains in economic activity in some border areas might come at the expense of economic activity in some other border areas.

Redding (2016, Section 5.5) simulates a hypothetical two-country world with a road running perpendicular to the border. This is the theoretical setup that comes closest to our empirical configuration. Interestingly, he finds that the effect of trade liberalization on both population and real wages is positive at the point where the road crosses the border and then decreases monotonically along the road as one moves inland. Our aim in this paper is to quantify the average effect at the border and its decay in space using actual rather than simulated data. Redding's (2016) analysis also illustrates how in general equilibrium border regions away from the border-crossing road may well experience net losses in terms of population and/or wages, at the expense of border regions closer to the road. We shall explore the empirical relevance of this prediction as well.

2.2.2 Empirics

The dominant theoretical prediction whereby trade liberalization is favorable to the development of border regions is borne out in a majority of existing empirical analyses. Following the seminal paper by Ades and Glaeser (1995), a considerable number of studies have found trade openness to be associated with the spatial dispersion of activities within countries.[2] This is consistent with economic catch-up by border regions. Within-country studies show more mixed results, although partly because many of them focus on the case of Mexico, where maquiladora activity concentrated heavily in the northern part of the country, creating a second agglomeration pole which came to overtake the traditional one (Mexico city) in terms of manufacturing production (e.g. Hanson, 1998). A similar pattern has been observed in China, where rising trade openness has been associated with intensified concentration of industrial activity in the southeastern coastal region (Kanbur and Zhang, 2005).

[2] Ten out of eleven cross-country analyses surveyed by Brülhart (2011) diagnosed trade-related spatial dispersion.

More recent recent papers have used the closing-off of central and eastern European markets after World War II and their reopening after the fall of the Berlin Wall as a natural experiment. This allowed researchers to uncover plausibly causal evidence of the effect of cross-border market access on the economic fortunes of border regions. Cross-border liberalization is found to have had a significantly positive effect on the population growth of border regions in Germany (Redding and Sturm, 2008) and on employment and wages in Austrian border regions (Brülhart, Carrre and Trionfetti, 2012). Both papers document border shadows, whereby, in the Cold War years, population density, employment density and wages progressively fell as one got closer to the Iron Curtain, which represented an almost insuperable barrier to trade. After the demise of the communist bloc, growth was stronger in German and Austrian cities close to the old Iron Curtain, consistent with cross-border trade liberalization disproportionately favoring the economic development of border regions.

In this paper we offer three main extensions to this existing body of research. First, we seek to quantify effects that were mostly captured only in qualitative terms in the existing quasi-experimental work. By taking measured changes in trade intensities as our explanatory variables instead of the binary before-after analyses of the Iron Curtain studies, we can compute magnitudes of border-region responses with respect to measurable magnitudes of changes in trade openness. Second, we seek to quantify effects at the border as well as gradients as one moves away from border crossing points. Third, we extend the analysis to essentially the entire world economy, allowing us in particular to explore border-region trade effects in developing countries.[3]

2.3 Data

2.3.1 Construction of the dataset

The uses and limitations of *night lights* data as a proxy for economic activity have been widely discussed.[4] On the whole, night lights have been found to represent a good proxy for economic activity and human development at the local level (Bruederle and Hodler, 2017). The collection and cleaning of night lights data recorded by satellites is a five-step process that includes cloud masking, filtering out of light signals (radiance) from ambient noise, aggregation and geo-referencing, filtering in terms of persistence (to exclude e.g. flares of lightning and fires), and quantifying radiance on a bounded scale ranging from zero to 63.[5] While this scale represents the luminosity of light proportionally, pixels with the value of 63 may be top censored. This on average concerns some 0.1 percent of pixels in our sample, mostly in advanced-economy cities. By contrast, the proportion of zero-light pixels is high in developing countries, ranging from an average of 45 percent in South Asia to 92 percent in Sub-Saharan Africa.

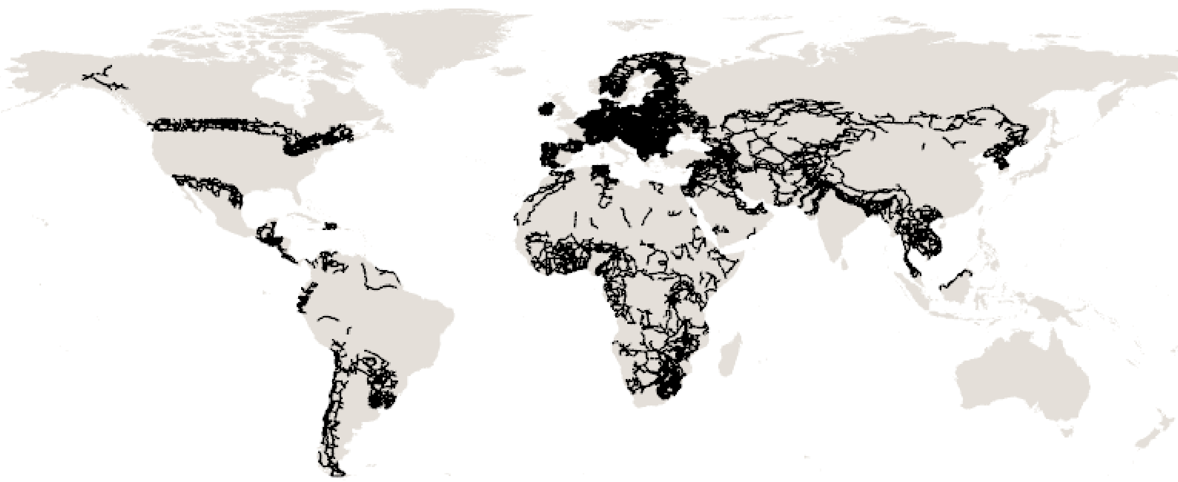
Our analysis focuses on locations within up to 200 kilometers from international land borders. Distance to the border is measured along *road corridors*. We consider all border-crossing "major roads" according to the 1990 version of the ESRI Roads and Highways dataset (see Figure 2.2). Our analysis is thus based on 2,230 border crossings. As shown in Table 2.1, we observe 70 land border crossings on average in advanced economies but only 20 in developing countries, reflecting the lower density of the road network in the latter group. Given the larger number of developing countries, they nonetheless account for some 62% of border crossings observed in our data.

[3] Hirte, Lessmann and Seidel (2018) also use night-lights data to study the effect of international trade on within-country regional inequality with world-wide country coverage. Their analysis focuses on indices of within-country regional inequality without considering border regions specifically.

[4] See e.g. Sutton, Elvidge and Ghsosh (2007); Henderson, Storeygard and Weil (2012); Donaldson and Storeygard (2016); Pinkovskiy and Sala-i-Martin (2016).

[5] We in addition drop cells featuring lights emitted by gas flares – which do not proxy well for economic activity – using readily available information on their location (see Henderson, Storeygard and Weil, 2012).

Figure 2.2: Cross-border roads



Note: Major cross-border roads up to 200km from the border, as defined in the ESRI Roads and Highways dataset.

Table 2.1: Borders and border crossings

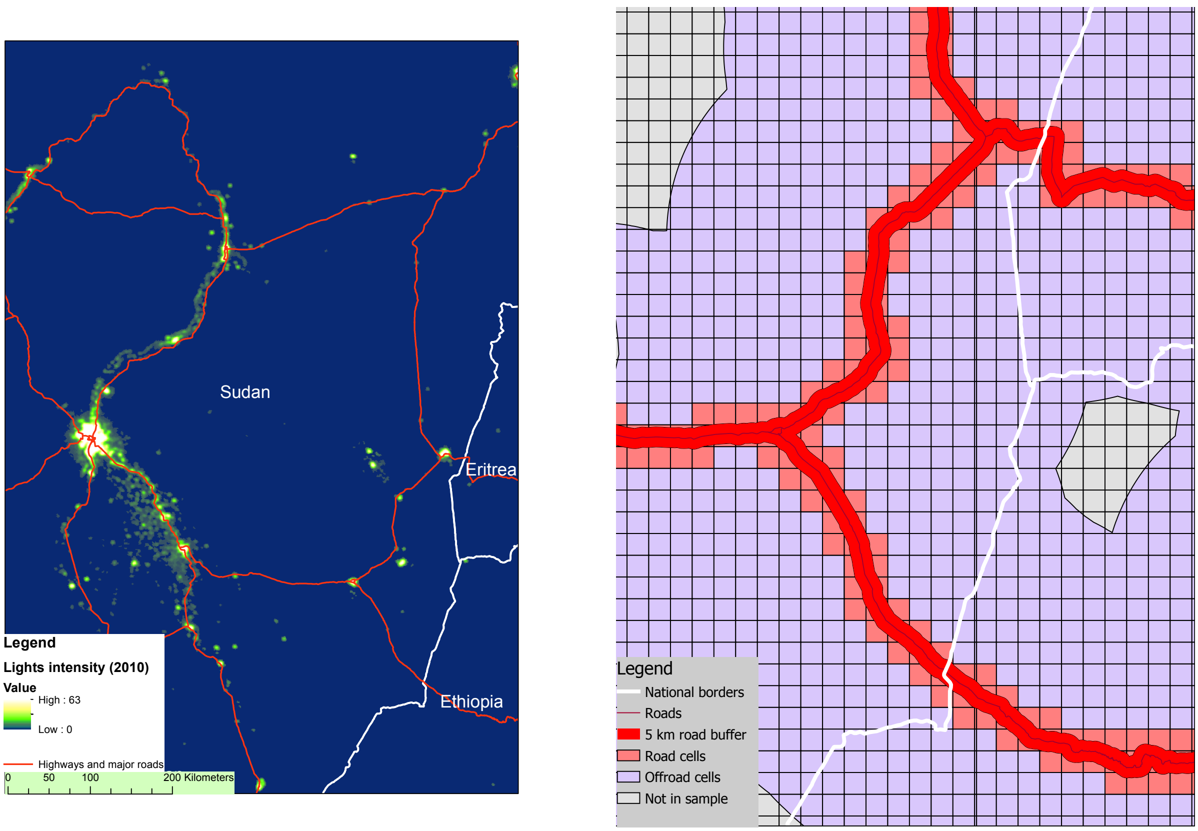
	Land borders per country	Border crossings per country	Border crossings per border	Total number of border crossings	Total number of on-road grid cells
Advanced economies	3.96	70.07	17.71	1,159	149,944
Developing economies	4.31	20.40	4.74	1,071	499,165

Note: Countries grouped according to 2015 World Bank classification.

Based on our sample of border-crossing roads, we perform a number of operations on the raw lights data using GIS software. An illustration is given in Figure 2.3. Panel (a) shows our sample roads in the case of the border region between Sudan, Eritrea and Ethiopia. To be part of our analysis, a road needs to cross a land border and be classified as either a "highway" or a "major road" in the ESRI dataset. The figure illustrates how lights cluster along major roads. Panel (a) also offers an example of the border shadow: light intensity diminishes gradually as one moves away from the Sudanese capital Khartoum toward the Ethiopian border.

Figure 2.3: Roads, lights and grid cells

(a) Roads and lights (b) Units of observation



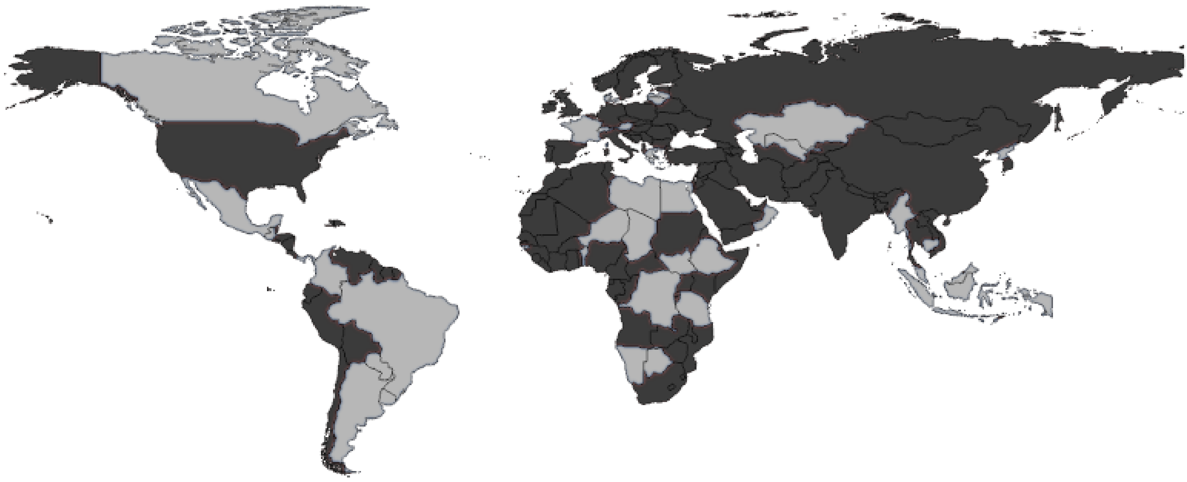
Note: National borders in white, major roads in red. Cells in panel (b) enter the baseline sample if their road distance from the closest border is $<= 200km$ and their geodesic distance from the closest road is $<= 100km$. Source: ESRI ArcGIS.

In panel (b) of Figure 2.3, we zoom in further to illustrate the construction of our units of observation. Our basic units are 10×10 kilometer grid cells. In order to be part of our sample, a cell needs to be within 200km of distance along the road from the border. Within each of these cells, we compute the average light intensity of all contained light pixels. We then construct buffers with a width of ten kilometers around the border-crossing roads. Around these road buffers, we construct additional outer buffers with a width of 200 km. This allows us to distinguish between cells that are located directly on a road (on-road cells) and cells located in border regions but away from the main roads (off-road cells). By doing this, we obtain some 162,000 grid cells for each of the years 1995, 2000, 2005 and 2010 (the years for which gridded population data are also available).[6][7] For each

[6]The DMSP satellites were discontinued in 2013 and replaced by a new system of satellites called VIIRS. As so far there is no consensus on how to convert values from different satellites into a unified scale (see e.g. Chen and Nordhaus, 2015), we limit our panel to the years 1995-2013.

[7]In the appendix we also provide results for estimations including the lights data of 2013, showing that results are not sensitive to the newer data.

Figure 2.4: Border region development worldwide



Note: Sample countries are displayed according to the average light intensities in border regions in relation to the respective country average before conditioning on any covariates. In dark gray countries, border regions, defined as within up to 200 kilometers, are on average darker than interior and coastal regions, and vice-versa for light gray countries.

on-road cell, we compute the distance of its center to the closest border along the border-crossing road, as well as the geodesic distance to the next sea port and airport. For each off-road cell, we compute the distance to the closest on-road cell as well as the distance to the border along the road from that on-road cell.[8]

Detailed information on all our data sources and definitions is provided in Appendix 2.8.

2.3.2 The border shadow

Importantly for the purpose of this paper, border shadows can easily be documented in the raw data.

Before analysing border shadows within border regions, we provide some context on the development of border regions when measured in terms of light intensity. To do so, we compute average light intensities within countries separately for grid cells located within 200 kilometers of land borders (the "border region") and for grid cells located beyond 200 kilometers of the nearest land border (the "interior region").[9] The results are shown in Figure 2.4, where all countries featuring border regions that are relatively less illuminated than the respective interior regions are colored red. The map shows that in the raw data border regions display lower light intensities than interior regions in most but not all countries: 76% of mapped countries host relatively "dark" border regions. That is 105 of the 138 countries shown in Figure 2.4. Weighted by GDP, these account for 83% of the sample. Weighted by population, they account for 80% of the sample.

Our main approach in this paper is to consider light gradients within 200 kilometers of land borders. Within this range, the raw average light intensity score in the outer distance band (100-200 kilometers from land borders) equals 4.17, but that in the inner distance band (0-100 kilometers) is only 3.12. As shown in Table 2.2, the difference is statistically significant.[10] The averages shown in Table 2.2 understate the steepness of

[8] Summary statistics for all variables are given in Appendix Tables 2.11 (all observations), 2.12 (on-road observations only) and 2.13 (off-road observations only).

[9] For countries that are too small to host an interior region according to this definition, we decrease the cutoff distance in increments of 25 kilometers until the interior region becomes non-empty.

[10] As an illustration, Appendix Figure 2.9 maps average lights within these two distance bands for the countries of Sub-Saharan Africa.

Table 2.2: Average light intensity by distance from border (scale: 0-63)

	Mean	Std. dev.	t stat.	No. obs.
$0-100$ km distance from border	3.12	8.05		345,326
$100-200$ km distance from border	4.17	12.39		303,783
Difference	-1.05		-54.32	

the gradient, because they aggregate lights by broad distance band. Our average observed light intensity for on-road cells at 100 kilometers from the border is 3.41 and at 200 kilometers it is 4.75, while the average light intensity at the border crossing is 2.97.[11] Hence, grid cells at the border are on average 13 percent darker than grid cells 100 kilometres inland and fully 37 percent darker than grid cells 200 kilometers inland.

The location of borders, of course, is not random and often coincides with inhospitable terrain. Part of the observed gradient is therefore undoubtedly explained by the endogeneity of border locations and not reflective of any man-made barriers to trade. However, as we document below, a strongly positive light gradient in distance from the border persists in the data once we control for topography. This implies that, while borders typically cross "naturally dark" regions, they cast an additional shadow over these regions.

2.4 Estimation

2.4.1 Two scenarios

Our main aim in this paper is to study the effect of trade liberalization between neighboring countries on light gradients around the border. Starting from a situation with a border shadow, theory suggests two possible scenarios, which we illustrate in Figure 2.5. If the productivity advantages of interior regions were to outweigh their disadvantage from greater distance to the border, then the interior of the country would benefit more from the liberalization than the border region, thus steepening the lights gradient (panel a). We refer to this as the "exacerbated border shadow" scenario. Conversely, in the "attenuated border shadow" scenario, trade liberalization leads to a more even distribution of economic activity in space, thus flattening the lights gradient and therefore brightening up the border shadow (panel b). As discussed in Section 2.1, theory can accommodate both configurations.

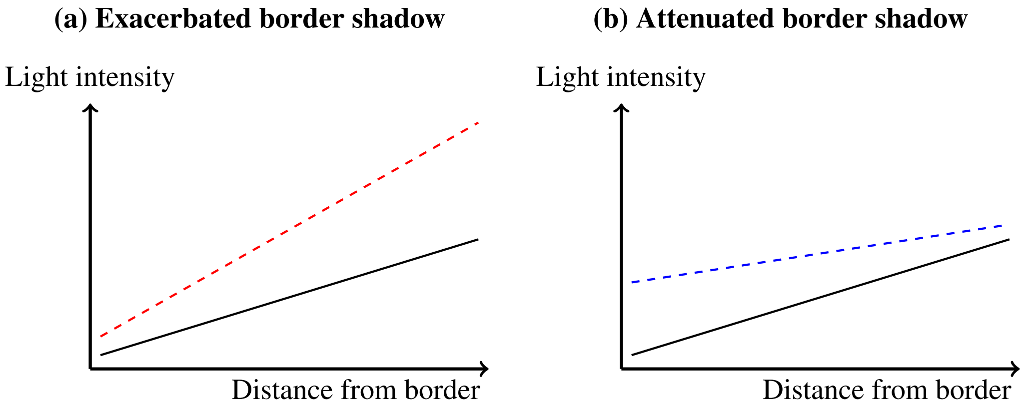
Note that our two stylized scenarios illustrated in Figure 2.5 assume positive effects of trade liberalization on local light intensity at all locations. When, as in most of our empirical specifications, 'trade' stands for exports, this assumption is consistent with all theoretical models and evidence we are aware of. However, when 'trade' is understood to mean imports, then negative regional effects could very well be possible.[12] We shall therefore explore the import channel as well, and our empirical specifications of course allow for the possibility of negative average trade effects on light intensity at any border-distance interval.

Figure 2.5 furthermore assumes a linear relationship between distance from the border and light intensity. This choice is not completely evident and one might assume for example non-monotonic pattern with a peak of light intensity directly at the border crossing followed by decreasing light intensity in the periphery of the border before light intensity then increases towards the interior region. We address this concern threefold: Table 2.14 presents results for different cut-off distances from the border and finds the gradient of light intensity along distance to be steepest on the first 50km from the border. Table 2.15 shows that our results are robust to an

[11] See Appendix Table 2.11.

[12] For evidence on potentially long-lasting negative impacts of import liberalization on particularly affected local labor markets see, e.g., Autor, Dorn and Hanson (2013), Dix-Carneiro and Kovak (2017) or Caliendo, Dvorkin and Parro (2017).

Figure 2.5: Trade and the border shadow – two scenarios



The solid lines illustrate the border shadow before trade liberalization. The dashed lines illustrate border shadows after trade liberalization.

exercise where we drop cells that are located directly at the border crossing. Finally, Table 2.16 estimates equation 2.1 as polynomial regarding distance and finds the quadratic terms to be insignificant. All these three results lend credibility to our assumption of an approximately linear relationship between distance and light intensity.

2.4.2 Baseline estimation

Our baseline empirical strategy consists of estimating night-light distance gradients along major roads across all of the world's land borders. In order to capture the causal effect of borders, we control for confounding influences arising from other exogenous sources of spatial heterogeneity such as altitude or proximity to ports and airports. Moreover, we systematically include fixed effects in order to remove country or even region-specific heterogeneity, as Pinkovskiy (2017) shows that night lights exhibit significant nation-specific variation.

Specifically, let $y_{irscc't} = y_{it}$ be the log light intensity of grid cell i located on road r in sub-national region s leading from country c to country c' in year t.[13] Roads r are defined as belonging to one country only, such that every cross-border road corridor consists of two "roads". The subscripts r, s, c and c' are implied by i, as every cell is uniquely assigned to a country, region and nearest road. We denote by d_i^{border} cell is distance from the nearest border crossing along road r. $T_{cc't}$ stands for the log value of trade of country c with neighboring country c' across that border, where trade is measured alternatively as exports from c to c' (our baseline) or as imports by c from c'.

When limited to on-road locations (red grid cells in Figure 2.3b), our empirical model can be written as follows:

$$y_{it} = \beta_0 + \beta_1 d_i^{border} + \beta_2 T_{cc't} + \beta_3 (d_i^{border} \times T_{cc't}) + \theta \mathbf{x}_i + \gamma_r + \gamma_{st} + \gamma_{c't} + u_{it} \qquad (2.1)$$

where $\mathbf{x}_i$ is a vector of grid-cell-level controls that includes average altitude, average slope, dummy variables for whether a sea port or airport respectively is closer to i than the nearest land border, interactions of those two dummies with the geodesic distance from the port or airport in question, and a dummy that takes the value of

[13] In order not to lose grid cells with zero measured lights through the log transformation, we add 0.01 to recorded lights. This approach is in line with Michalopoulos and Papaioannou (2013). Alternatively setting this value to 0.001, 0.1 or to 1 has no discernible impact on our results.

1 if at the relevant border crossing the same ethnicity dominates on both sides of the border. γ_r is a road fixed effect.

A potential threat to our identification strategy arises from unobserved political or other regional events that would change the gradient of economic activity along a certain road over time as well as being correlated with trade and trade policy. For instance, well-connected local politicians in border areas might obtain privileged access to public funding (for roads, electrification etc.) while at the same time using their influence to push for trade-facilitation reforms benefiting primarily their (border) constituencies (Hodler and Raschky, 2014). Other examples of region-time-specific confounding factors are regional outbreaks of violent conflict or the localized occurrence of extreme weather events.

We therefore estimate models that control for region-year fixed effects. Region-year fixed effects add up to year fixed effects and therefore control for common time-varying determinants of measured light intensities, notably including differences in satellite capabilities over time. Moreover, we allow for time-varying cross-border spillovers other than trade (e.g. through migration or investment surges) by controlling in addition for neighbor-country-year effects.

In order to explore the effect of trade liberalization on border-region grid cells located beyond 10 kilometers of a major border-crossing road, we expand equation (2.1) to incorporate also off-road locations (blue grid cells in Figure 2.3b):

$$\begin{aligned} y_{it} = \beta_0 + \beta_1 d_i^{border} + \beta_2 T_{cc't} + \beta_3 (d_i^{border} \times T_{cc't}) \\ + \beta_4 Off_i + Off_i \left(\beta_5 d_i^{border} + \beta_6 d_i^{road} + \beta_7 T_{cc't} + \beta_8 d_i^{border} \times T_{cc't} + \beta_9 d_i^{road} \times T_{cc't} + \gamma_r \right) \\ + \theta \mathbf{x}_i + \gamma_r + \gamma_{st} + \gamma_{c't} + v_{it}, \quad (2.2) \end{aligned}$$

where Off_i is a dummy variable that takes the value of 1 if cell i is not within 10 kilometers of a major border-crossing road; and d_i^{road} is the geodesic distance of cell i to the nearest grid cell on a border-crossing road r (hence, $d_i^{road} = 0 \Leftrightarrow Off_i = 0$). All off-road grid cells are uniquely attributed to their nearest border-crossing road r.[14]

Our empirical model implies that we identify the effects of interest at the within-road (and thus within-country) level: the gradient of night lights along a certain road is compared across years. Our coefficients of main interest are β_1, β_2 and β_3. A significantly positive estimate of β_1 is evidence for the border shadow at zero trade, as it implies that economic activity increases as one moves inland, away from the border, when $T_{cc't} = 0$. β_2 captures the effect on night lights of increased cross-border trade at the border crossing (where $d_i^{border} = 0$), and the interaction term (β_3) allows us to gauge how increased trade affects the distance gradient. When β_3 has the same sign as β_1, the data support an exacerbation of the border shadow, otherwise they support an attenuation of the border shadow. When there is a border shadow, i.e. β_1 is positive, then a negative estimate of β_3 implies an attenuation of the border shadow with a relative rise in economic activity at locations closer to the border when compared to locations towards the interior of the country, as in panel b of Figure 2.5. Finally, in specifications that include off-road grid cells, coefficients β_5 to β_9 allow for complementary evidence on the effect of trade on light gradients based on readings for those locations.

[14] We allow for different road-specific fixed effects for road and off-road cells, for the sake of comparability of coefficients across estimations with and without off-road cells. This means that the coefficient β_4 is absorbed by the fixed effects $Off_i \times \gamma_r$.

2.4.3 Identification and inference

As we seek to capture the causal effect of changed trade intensities on the geography of night lights, we need to address the potential endogeneity of trade. Not only can trade be expected to affect activity as measured through lights, but changes in border-region economic activity can in turn affect the volume of cross-border trade. We therefore estimate equations (2.1) and (2.2) by instrumenting bilateral exports $T_{cc't}$ with tariffs imposed by destination country c' on goods from origin country c. Since trade weights could be endogenous, tariffs are computed as unweighted averages across sectors. Our identifying assumption is that activity in cell i does not directly affect tariffs imposed by country c'. Given the small size of our cells, this assumption strikes us as unproblematic. The exclusion restriction we impose requires that tariffs of country c' affect economic activity in country c only through changes in the volume of exports from country c to c' – an assumption we consider similarly plausible.

Standard errors are clustered two-ways, by road r and by country-pair-year $cc't$. Roads represent the basic component of the fixed effects structure in the baseline model (2.2), and country-pair-year is the dimension of variation of our trade variable. This is a conservative approach, as can be seen in Appendix Table 2.17, where we compare standard errors. Our two-way clustered standard errors are larger than standard errors clustered one-way by road in a large majority of cases, and always with respect to our interaction coefficients of main interest, β_3 and β_9. We have also considered spatially and temporally correlated errors following Conley (1999) for OLS estimation as adapted by Colella, Lalive, Sakalli and Thoenig (2018) for panel IV regression.[15] Table 2.17 shows that this approach also yields almost uniformly smaller standard errors than our preferred (because conservative) two-way clustering.

2.5 Baseline results

2.5.1 Exports

Table 2.3 shows our baseline OLS and IV estimates, taking exports as the trade measure. In columns (1) and (2), we present estimates for on-road cells only (equation 2.1), while columns (3) and (4) show estimates for on-road and off-road cells combined (equation 2.2). For both specifications, we show regression estimates without and with instrumenting exports.

Our coefficient estimates turn out to be stable across specifications, statistically significant in most instances, and consistent with the dispersion scenario throughout. Instrumenting does not affect our qualitative findings.[16] The control variables affect light intensities in ways that correspond to expectations: high-altitude and steep locations are darker, and locations close to ports and airports are brighter.[17] Ethnic homogeneity across borders, however, is not found to affect light intensity statistically significantly.

We find strong evidence of border shadows. Estimated coefficients on the raw distance measure $\hat{\beta}_1$ are significantly positive across all specifications, ranging from 0.17 to 0.18, both on and off the main border-crossing roads. According to our preferred specification, reported in column (4) of Table 2.3, economic activity measured through night lights increases by some 18 percent with every 10 kilometers of distance from the

---

[15] For a previous application, see König, Rohner, Thoenig and Zilibotti (2017). We allow for spatial correlation up to 1,000 kilometers, and for serial correlation up to 5 years.

[16] Our instrument is strong. First-stage F-statistics, shown at the bottom of Table 2.3, are above conventional acceptance thresholds. Appendix Table 2.17 shows Table 2.3 with alternative standard error estimates. Appendix Table 2.18 shows representative first stage results for column (4).

[17] It should be noted that aerial and maritime trade potentially bias our estimated coefficients concerning $\hat{\beta}_2$ and $\hat{\beta}_3$ towards zero, since the effect of an increase in exports to the neighbor country would not be measured through changes in light intensities along border crossing roads if trade is carried out by other means of transport.

Table 2.3: Baseline estimates

Dependent variable: Average light intensity (logs)	(1) OLS	(2) IV	(3) OLS	(4) IV
	Effects on grid cells along road			
Distance from border (in 10km)	0.170***	0.183***	0.171***	0.178***
	(0.030)	(0.058)	(0.028)	(0.065)
Bilateral exports (in logs)	0.058	0.482	0.080	0.556
	(0.146)	(0.519)	(0.155)	(0.513)
Bilateral exports × Distance from border	-0.030***	-0.033**	-0.033***	-0.035**
	(0.010)	(0.016)	(0.010)	(0.016)
	Additional effects on off-road grid cells			
Off-road × Distance from border			-0.156***	-0.144*
			(0.032)	(0.076)
Off-road × Distance from road			-0.009***	-0.008***
			(0.001)	(0.002)
Off-road × Bilateral exports			0.069	0.805
			(0.187)	(1.405)
Off-road × Bilateral exports × Distance from border			0.032***	0.030*
			(0.012)	(0.018)
Off-road × Bilateral exports × Distance from road			-0.002***	-0.002**
			(0.000)	(0.001)
	Control variables			
Altitude (in 100m)	-0.115***	-0.103***	-0.062***	-0.073***
	(0.011)	(0.012)	(0.011)	(0.013)
Slope	-0.001	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.001)	(0.001)
Port closer than next land border (dummy)	1.467***	1.501***	1.452***	1.477***
	(0.353)	(0.300)	(0.271)	(0.260)
Port dummy × Distance from port	-0.144***	-0.110***	-0.136***	-0.130***
	(0.039)	(0.038)	(0.049)	(0.055)
Airport closer than next land border (dummy)	1.162***	1.006***	0.689***	0.668***
	(0.105)	(0.112)	(0.091)	(0.088)
Airport dummy × Distance from airport	-0.049***	-0.034***	-0.030***	-0.027***
	(0.020)	(0.011)	(0.010)	(0.011)
Same ethnicity on both sides of border (dummy)	-0.078	-0.070	-0.065	-0.069
	(0.101)	(0.095)	(0.088)	(0.096)
Off-road cells	NO	NO	YES	YES
Controls	ALL	ALL	ALL	ALL
Road FE	YES	YES	YES	YES
Region-Year FE	YES	YES	YES	YES
Neighbor country-Year FE	YES	YES	YES	YES
First-stage F statistic		15		13
# Clusters	812	812	1,639	1,639
# Observations	113,289	113,289	648,783	648,783

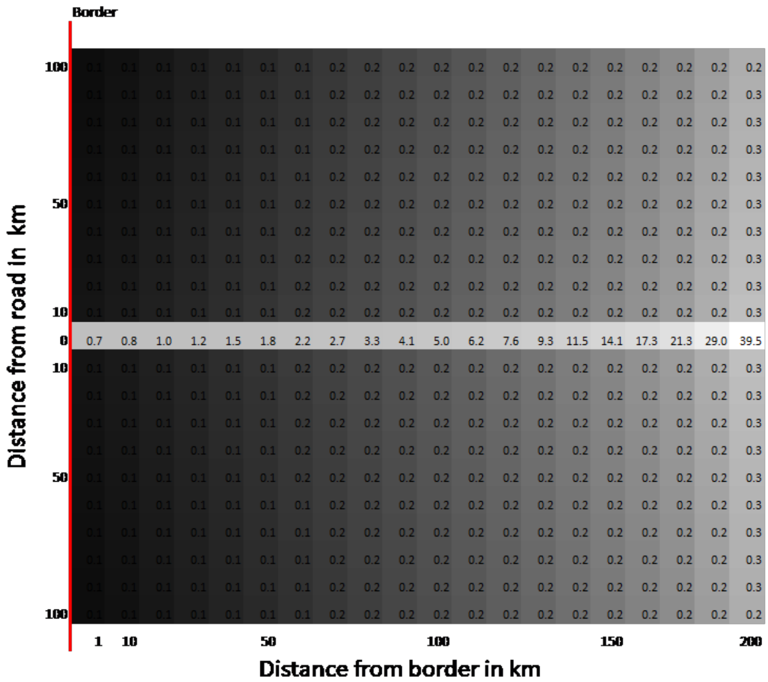
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

border in a hypothetical scenario of zero cross-border trade. For off-road cells, the distance gradient from the border is weaker – about 3 percent per 10 kilometers according to that same specification.

The multiple interaction terms of our regression model do not lend themselves individually to easy interpretation. In Figure 2.6, we therefore illustrate the border shadow implied by our preferred estimates (Table 2.3, column 4). We show a hypothetical 200 × 200 kilometer area with an international border at its western edge and a perpendicular border-crossing road running through the middle. We calibrate all variables at the 25th percentile. We show predicted grid-cell light intensities as a function of the estimated coefficients, with

Figure 2.6: The predicted border shadow



Note: The graph shows predicted light intensities based on a specification featuring road fixed effects, region-year fixed effects, neighbor-year fixed effects, all control variables and exports instrumented with tariffs (Table 2.3, column 4), with exports set to the value of the 25th percentile in our data. Darker colors symbolize lower light intensity.

variation across grid cells being determined by the spatially identified parameters $\hat{\beta}_1$, $\hat{\beta}_3$, $\hat{\beta}_4$, $\hat{\beta}_5$, $\hat{\beta}_6$, $\hat{\beta}_8$ and $\hat{\beta}_9$ (equation 2.2).[18] The shading of the grid cells is proportional to the predicted light intensities, and exact values are reported inside each cell.

It is evident from Figure 2.6 that our estimates imply pronounced border shadows also with trade intensity at the 25th percentile: predicted lights get brighter as one moves away from the border, both on and off the main road. The figure also illustrates how light intensity drops off abruptly as one moves away from the road.

Our second and main result is that the border shadows illustrated in Figure 2.6 are attenuated by cross-border exports. The interaction coefficient $\hat{\beta}_3$ is statistically significantly negative in all regression specifications. This suggests that export growth leads to stronger increases in lights along road segments close to the border than further inland. For example, the estimated coefficient in our preferred specification (Table 2.3, column 4) implies that export growth brightens up grid cells along main roads by 30 percentage points more at the point of border crossing than 200 kilometers inland.[19]

It has to be noted that our estimates for $\hat{\beta}_2$ remain positive but statistically insignificant across specifications. Table 2.19 presents results for less demanding specifications to shed light on this finding. Columns (1) to (4) show results for a within-road estimation including year fixed-effects, but not controlling for region-year and neighbor-year fixed effects. In this setting, $\hat{\beta}_2$ is consistently estimated as positive and statistically highly

[18] We retain estimated values of all these parameters, including coefficients that are not statistically significantly different from zero. The point estimates remain the values with the highest likelihood even in those instances.

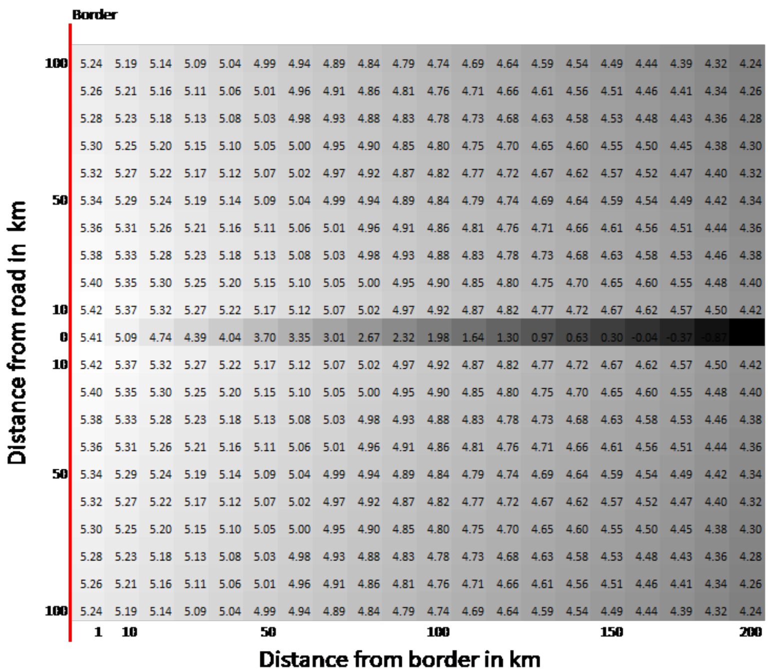
[19] As explained in Appendix 2.8, we measure light intensities only along roads that were already recorded in 1995. This potentially biases our estimated coefficients concerning $\hat{\beta}_2$ and $\hat{\beta}_3$ towards zero. This is due to the fact that in the case of new roads being constructed in the vicinity of our sample roads, trade would be likely to flow along new and more modern roads, decreasing the potential effect measured at the initial sample roads.

significant. When running the estimations within-region-year including neighbor-year fixed effects but no road fixed effects in columns(5) to (8), $\hat{\beta}_2$ is statistically insignificant as in our baseline result. This shows that the insignificance of $\hat{\beta}_2$ has to do with the region-year fixed effects, which assign each grid cell within a region to the closest neighbor country, but do not take into account the distances to other neighbor countries. This means that grid cells for which the distances from different borders are almost identical will be assigned to only one neighbor country, while in reality they may depend as much on the trade to other neighbor countries. This in turn biases our estimated effect for $\hat{\beta}_2$ towards zero.

We again provide a graphical illustration of the combined effects of our estimates. Figure 2.7 is constructed analogously to Figure 2.6 but rather than showing predicted light intensities for a given level of exports we show predicted percentage changes in light intensities for a doubling of exports, at the mean values of the remaining variables.[20] It appears clearly in Figure 2.7 that our estimates imply exports to brighten up locations close to the border more strongly than locations further inland, and that this is true both along and off the main border-crossing roads.

In a nutshell, increased trade attracts activity towards border regions, both on and off the border-crossing roads. Our estimates also imply that within our sample distance band of 200 kilometers exports are associated with increases in lights for all grid cells except for cells at the main road for which the distance from the border exceeds 150km.[21]

Figure 2.7: Predicted percentage change in light intensity associated with a 10% increase in exports



Note: The graph shows predicted percentage changes in light intensity after a 10% increase of exports starting from a scenario with trade set to the value of the 25th percentile in our data (i.e. starting from the values presented in Figure 3.6, based on a specification featuring road fixed effects, region-year fixed effects, neighbor-year fixed effects, all control variables and exports instrumented with tariffs (Table 2.3, column 4)). Darker colors symbolize lower light intensity.

[20] For our graphical illustration we set exports to the value at the 25th percentile, because when we set it at the mean value, the off-road border shadow no longer exists.

[21] Grid cells that are further than 100 km away from a major border-crossing road are found only in areas with very low population density, typically in large developing countries. As the satellites mostly do not record any measurable light emissions in these areas, it would be mechanically impossible to find a decrease in light intensity in those cells. Hence our chosen buffer width of 100 km.

2.5.2 Lights per capita

Our main dependent variable, total light emissions per grid cell and year, has the advantage of being precisely measured with constant reliability across time and space. An important limitation of this variable is that we cannot distinguish between population and income effects: do brighter lights associated with intensified trade reflect the migration of people towards border regions, do they reflect higher per-capita incomes in border regions, or do they reflect a combination of both? In order to address this question, we combine the lights data with the Gridded Population of the World dataset published by the Earth Institute of Columbia University, which are available at the same spatial resolution as the one we choose for our analyses based on lights only (see Appendix A for details).

In Table 2.4, we show estimates of our baseline models (2.2) with lights per capita and population as the dependent variable. These estimates suggest that border shadows mainly reflect lower income (proxied by lights) rather than lower population density in border regions. The estimated gradients $\hat{\beta}_1$ in per capita terms are very similar to those of the baseline regressions (Table 2.3), while the corresponding population gradients are significantly smaller.

Table 2.4: Baseline effects for light intensity per capita and population

Dependent variable:	Lights per capita (logs)		Population (logs)	
	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
Distance from border (in 10km)	0.155***	0.163***	0.070***	0.081***
	(0.024)	(0.029)	(0.019)	(0.023)
Bilateral exports (in logs)	0.095**	0.211*	0.119	0.161
	(0.041)	(0.115)	(0.093)	(0.099)
Bilateral exports × Distance from border	-0.011**	-0.016**	-0.022***	-0.024**
	(0.005)	(0.007)	(0.007)	(0.008)
Off-road cells	YES	YES	YES	YES
Controls	ALL	ALL	ALL	ALL
Road FE	YES	YES	YES	YES
Region-Year FE	YES	YES	YES	YES
Neighbor country-Year FE	YES	YES	YES	YES
Kleibergen-Paap F statistic		13		13
# Clusters	1,639	1,639	1,639	1,639
# Observations	648,787	648,787	648,787	648,787

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Regarding the effect of trade, our results of Table 2.4 imply that increased exports raise both population and incomes, in roughly equal measure.

2.5.3 Imports

Up to now, we have defined trade T_{ct} as the value of exports from country c to country c', instrumented with the tariff rate of country c' on goods from country c. We can deploy this framework to study the effect of imports, either by redefining T_{ct} as imports or by considering exports and imports simultaneously. Accordingly, T_{ct}

is instrumented with country-c unweighted tariffs on products from the neighboring country c'.[22] Results are reported in Table 2.5, with corresponding estimates for the export specification shown again for ease of comparison.

Table 2.5: Imports

Dependent variable: Average light intensity (logs)	(1) OLS	(2) IV	(3) OLS	(4) IV
Distance from border (in 10km)	0.171***	0.178***	0.161***	0.166***
	(0.028)	(0.065)	(0.029)	(0.066)
	Effects of exports			
Bilateral exports (in logs)	0.080	0.556		
	(0.155)	(0.513)		
Bilateral exports × Distance from border	-0.033***	-0.035**		
	(0.010)	(0.016)		
			Effects of imports	
Bilateral imports (in logs)			0.071	0.451
			(0.052)	(0.430)
Bilateral imports × Distance from border			-0.019*	-0.016*
			(0.009)	(0.009)
Off-road cells	YES	YES	YES	YES
Controls	ALL	ALL	ALL	ALL
Road FE	YES	YES	YES	YES
Region-Year FE	YES	YES	YES	YES
Neighbor country-Year FE	YES	YES	YES	YES
Kleibergen-Paap F statistic		13		12
# Clusters	1,639	1,671	1,429	1,429
# Observations	648,787	648,787	637,029	637,029

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Not surprisingly, we find the coefficients capturing the implied border shadow at zero trade ($\hat{\beta}_1$) to be virtually unchanged. However, both the effect of trade on economic activity at all distances from the border ($\hat{\beta}_2$) and the effect on the gradient from the border ($\hat{\beta}_3$) are mostly smaller in size and less precisely estimated for imports than for exports.

The liberalization of imports thus has qualitatively comparable effects on the economic geography of border regions, though with somewhat lesser intensity.

2.5.4 Robustness

One evidently arbitrary choice underlying our baseline specifications concerns the width of the distance band around the border. We have run extensive sensitivity tests on alternatives to our baseline 200-kilometer distance cut-off, and found the qualitative patterns to be robust: border shadows are strong when trade intensities are low, and they are significantly reduced by trade liberalization.[23] The only noticeable irregularity we observe is that the main effect on distance, $\hat{\beta}_1$, more than doubles when we reduce the distance cut-off to 50 kilometers –

[22] Own-country tariffs, even though plausibly exogenous in many cases with respect to economic conditions in individual border regions, are a less convincing instrument than neighbor-country tariffs. This is a reason for why we mostly focus on exports.

[23] See Table 2.14, where we report estimations for 150, 100 and 50 kilometer distance cutoffs.

an observation that is however entirely consistent with zero-trade average light intensity being lower the closer one gets to the border.

We also explored the sensitivity of our results to the consideration of data for 2013, the last year for which comparable night-light measurements are available. Our estimates remain virtually unchanged. [24]

We have investigated the effect on our estimates of three further considerations.[25]

First, we consider the issue of "overglow", whereby inaccuracies in the spatial precision of light measurement by satellites leads to the attribution of light emitted in one cell to neighboring cells. We follow Pinkovskiy (2017), who proposes a correction based on a measured spatial autoregression term. Applying this correction leaves our estimates virtually unaffected.[26]

Second, we exclude small countries of an area less than that of a circle of 400 km in diameter, showing that constraints on the size of our border buffers do not seem to be a major issue. This reduces the number of observations by some 10% and does not significantly alter any estimation results.

Third, we estimate our regression model on landlocked countries only, thus dropping all observations where some bilateral shipments could travel by sea and "interior" regions include ocean coasts. This shrinks the size of the sample by almost 90%, but the qualitative results are again identical to the baseline estimates of Table 2.3. The estimated main effects $\hat{\beta}_1$, capturing the implied border shadow at zero trade, however are two to three times as large as in the full sample, consistent with the view that land borders matter more in the absence of access to ocean transport.[27]

Finally, as a complementary and more demanding estimation approach again, we estimate a version of our baseline equation (2.1) in which we consider grid-cell fixed effects instead of road fixed effects.[28] In this variant of the model, identification is attained solely from intertemporal variation. Time-invariant variables drop out, which means that we cannot in this specification estimate the border shadow for a given level of trade. We can, however, assess the impact of a growth in exports. Our estimates confirm the dispersing effect of increased trade: export growth is again found to have the strongest brightening effect on grid cells located close to the border and to the main road.

2.5.5 EU enlargement as an event study

In a further attempt at ascertaining the causal nature and pervasiveness of our detected effects, we focus on the two most recent enlargements of the European Union (EU) as an event study. The timing of accession to the EU is arguably exogenous to the economic fortunes of specific border regions. And even though EU-related trade liberalization has always been a gradual process of which accession was only the culminating conclusion, there is evidence that accession provides discrete additional impetus to trade flows between incumbents and accession countries.[29]

We therefore limit the sample to countries that joined the EU either in 2004 or in 2007, considering the effect on the side of the border that was previously not in the EU.[30] In this setting, our trade variable is a

[24] See Table 2.20.

[25] See Table 2.21.

[26] It is not surprising that this issue seems of little concern in our context whereas is matters significantly for Pinkovskiy (2017), because that analysis is focused on sharp discontinuities at borders whereas ours focuses on rather wide border-region bands inside each country (see columns (3) and (4) of Table 2.21).

[27] See column (4) of Table 2.21.

[28] See Table 2.22.

[29] See e.g. Cheptea (2013).

[30] The 2004 accession countries were Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovenia and Slovakia. The 2007 accession countries were Bulgaria and Romania. Because they are islands, Cyprus and Malta do not inform our estimates.

dummy that takes the value of 1 if the border between the two countries belongs to the European Union in year t. As a complementary exercise, we run "placebo" regressions for the same sample countries but considering their borders with non-EU countries. So, for example, the dummy for Poland's border with Germany switches from 0 to 1 between 2000 and 2005 in the accession regressions; whereas in the placebo regression we switch Poland's border with Ukraine from 0 to 1 in that same time interval, even though the institutional setting for Poland-Ukraine trade remained essentially unchanged over that period.

Table 2.6: Results for countries that accessed the EU in 2004 and 2007 (borders to EU-members)

Dependent variable: Average light intensity (logs)		
	(1)	(2)
	OLS	OLS
Treatment	Accession borders	Placebo borders
Distance from border (in 10km)	0.638***	0.298***
	(0.088)	(0.107)
Border within EU (dummy)	0.591***	0.149
	(0.161)	(0.123)
Border within EU × Distance from border	-0.410***	-0.171
	(0.112)	(0.155)
Off-road cells	YES	YES
Controls	ALL	ALL
Road FE	YES	YES
Region-Year FE	YES	YES
Neighbor country-Year FE	YES	YES
# Clusters	144	240
# Observations	20,708	38,616

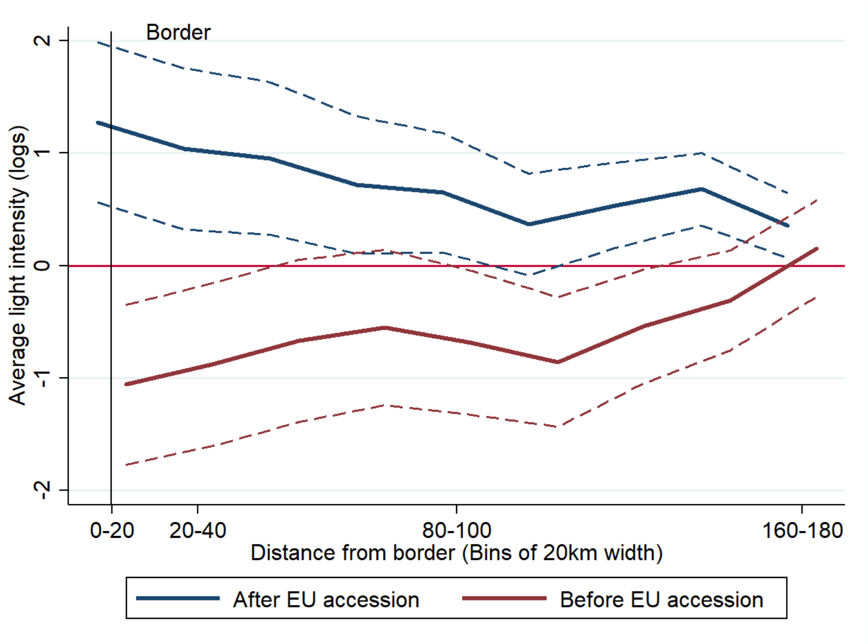
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Once again, we find strong evidence of border shadows being attenuated by trade liberalization. The coefficients in the first two columns of Table 2.6 all have the expected signs and are statistically significant. Border shadows are found to be very pronounced before accession, with night lights increasing by some 63 percent with every 10 kilometers of distance from the border – an effect that is two to three times as strong as the no-trade lights gradient implied by our world-wide baseline estimates of Table 2.3. This illustrates the activity-depressing impact of the Iron Curtain. Once a border becomes part of the EU, however, the gradient of lights within the 200 kilometer border region disappears altogether, and the border region on average becomes some 59 percent brighter. We illustrate these effects in Figure 2.8, which visualizes the stark reversal in the fortunes of border regions after accession to the EU.

The placebo regressions reported in the right half of Table 2.6 show that border shadows also exist in the east of the accession countries, but they were not discernably reduced after EU accession. This further supports our interpretation of the accession effects as being caused by improved cross-border market access made possible by EU enlargement. We do not find evidence of reduced activity in eastern border regions, which suggests that growth in regions bordering the EU did not come at the expense of growth in regions bordering non-EU countries.

Figure 2.8: The effect of accessing the European Union



Note:The graph represents the point estimates for a within-road regression according to equation (3.1), where distance is expressed as a set of dummy variables for bins of 20 km width. The sample is split into "before EU accession" (red line) "after EU accession" (blue line) years.

2.6 Extensions

2.6.1 Effects by world region

We now explore the extent to which our results estimated for the world as a whole also hold for subsets of countries. We focus on two natural sample divisions: developing versus advanced economies, and individual continents.

Table 2.7 reports estimates of our on-road baseline model (2.1) separately for developing and advanced economies, using an interaction specification with a binary variable that is set to one for advanced economies. We attribute countries to the "advanced category if they were classified as "high income" in the World Bank's 2015 country classification (GNI per capita above USD12,476). According to this definition, our sample contains 36 advanced and 102 developing economies.

Qualitatively, our main results hold for both subsets of countries: locations close to borders are darker, *ceteris paribus*, but this spatial light gradient flattens as bilateral exports grow. The coefficients in the lower panel of Table 2.7 suggest that these effects are significantly stronger in advanced economies.[31]

We should however interpret this apparent difference with care, as the weaker effects measured in developing economies might at least partly be due to attenuation bias from measurement error in the export variable. We do not observe informal trade, which is considerably more important in developing than in advanced economies; and even formal trade may be recorded more accurately in the latter countries. Instrumenting with neighbor-country import tariffs likely cannot entirely solve this problem, as informal exports might to some extent be a substitute for formal exports and therefore react to tariffs in the opposite way. While cross-country differ-

[31] Advanced economies also exhibit stronger measured trade effects when we estimate interaction models with continuous measures of per-capita income, be it in 1990 terms (Table 2.23) or in 2015 terms (Table 2.24).

Table 2.7: Developing and advanced economies

Dependent variable: Average light intensity (logs)		
	(1)	(2)
	OLS	IV
	Effects in developing economies	
Distance from border (in 10km)	0.131***	0.118***
	(0.024)	(0.025)
Bilateral exports (in logs)	0.193	0.343
	(0.034)	(0.079)
Bilateral exports × Distance from border	-0.021***	-0.014***
	(0.005)	(0.004)
	Additional effects in advanced economies	
Advanced economy (dummy) × Distance from border	0.052***	0.078***
	(0.016)	(0.024)
Advanced economy × Bilateral exports	0.112***	0.200***
	(0.031)	(0.058)
Advanced economy (dummy) × Distance from border × Bilateral exports	-0.015***	-0.029***
	(0.004)	(0.009)
Off-road cells	NO	NO
Controls	ALL	ALL
Road FE	YES	YES
Region-Year FE	YES	YES
Neighbor country-Year FE	YES	YES
Kleibergen-Paap F statistic		13
# Clusters	812	812
# Observations	113,289	113,289

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

ences along the income dimension should therefore be interpreted with caution, our results strongly suggest that exports reduce border shadows in both advanced and developing economies.

Table 2.8: Effects by continent

Dependent variable: Average light intensity (logs)					
	(1)	(2)	(3)	(4)	(5)
	IV	IV	IV	IV	IV
Continent	Africa	Asia	Europe	Latin America	North America
Distance from border (in 10km)	0.134***	0.061*	0.194***	0.012	-0.108***
	(0.038)	(0.036)	(0.052)	(0.035)	(0.036)
Bilateral exports (in logs)	0.339*	0.342	0.441*	0.111	0.405
	(0.180)	(0.241)	(0.241)	(0.201)	(0.501)
Bilateral exports × Distance from border	-0.035***	-0.033***	-0.052***	-0.024	0.025***
	(0.010)	(0.011)	(0.014)	(0.021)	(0.011)
Off-road cells	YES	YES	YES	YES	YES
Controls	ALL	ALL	ALL	ALL	ALL
Road FE	YES	YES	YES	YES	YES
Region-Year FE	YES	YES	YES	YES	YES
Neighbor country-Year FE	YES	YES	YES	YES	YES
Kleibergen-Paap F statistic	12	11	10	3	13
# Countries	42	43	38	13	2
Share of cells with zero lights	0.86	0.63	0.16	0.67	0.39
# Clusters	291	132	527	447	242
# Observations	53,207	168,015	139,716	155,426	132,586

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

In Table 2.8, we subdivide the world further, showing estimates of the on-road baseline model (2.2) individually by continent. Our world-wide results turn out to be driven by African, Asian and European countries. Restricting the sample to each of these three continents in turn leads to qualitatively similar results. Given that these continents contain some 56 percent of our total number of observations and 123 of our 138 sample countries, it is unsurprising that they dominate our overall estimates.

Our central insights, however, do not seem to generalize readily to the Americas. In Latin America, the estimated border-shadow-reducing effects of exports are consistent with those we detect elsewhere, but they are not statistically significant. Moreover, there is no evidence of border shadows in a hypothetical zero-export scenario, the main effect of distance, $\hat{\beta}_1$, being indistinguishable from zero. In North America, border regions are brighter, *ceteris paribus* than interior regions. This is no doubt due to the continent's particular geography, with many urban centers clustered along the U.S. borders.

2.6.2 Cities

So far, we have controlled for the most important features of topography and for proximity to ports and airports, but we otherwise abstracted from within-country economic geography. Even conditional on these sources of spatial heterogeneity, however, locations may be unequally positioned to benefit from opportunities for cross-border trade. One evident source of heterogeneity is urbanization: cities are likely to be affected differently from rural locations. Such differences could arise for multiple reasons, including different sectoral specialization, different skill abundance, different availability of trading infrastructure, and agglomeration effects. Existing empirical studies seem to support the hypothesis that access to cross-border trade favors rural regions and smaller cities more than large cities, but these studies are all based on individual countries.[32]

In a first step towards investigating this issue, we distinguish "urban" roads from "rural" roads. Roads are defined as urban if anywhere within 200 kilometers from the border they reach a city with a population of at least 500,000.[33] For urban roads, we do not consider segments that lie between the first city reached when traveling inland from the border and the 200 kilometer cut-off, but the grid cells covering the city itself are included. According to this definition, 46% of sample grid cells belong to urban roads. Among urban roads, the average distance from the border to the nearest city is 47 kilometers.

We estimate on-road-cell models akin to equation (3.1), and we interact the distance and export variables with a dummy for urban roads.[34]

Table 2.9 reports our estimation results. Unsurprisingly, urban roads are found to be brighter on average than rural roads, and they have a steeper light gradient. Our main results holds both for urban and rural roads: exports reduce the gradient of lights. This effect is significantly more pronounced for urban roads, however. This suggests that border locations stand to benefit more from trade if they are close to a city in their own country. Consistent with this, the effect of exports on lights at the border is about twice as strong for urban roads as for rural roads. These results imply that urban border regions stand to benefit comparatively more from trade liberalization than rural border regions, a result somewhat at odds with previous empirical findings.

[32] Redding and Sturm (2008) show that population growth of smaller intra-German border towns suffered relatively more from Cold War partition than population growth of larger towns. Baum-Snow *et al.* (2018) find that population and GDP of non-primate Chinese prefectures grew more strongly than those of primate prefectures as a result of improved access to major sea ports. Studying Austria after the raising of the Iron Curtain, Brlhart, Carre and Robert-Nicoud (2018) find that below-average sized border towns experienced above-average employment growth.

[33] Appendix Table 2.25 shows results based on a population cut-off of 100,000. The qualitative results are very similar.

[34] This is equivalent to estimating equation (2.2) with a binary variable that is set to one for all grid cells that are located on urban roads instead of the binary variable Off_i.

Table 2.9: Urban and rural roads

Dependent variable: Average light intensity (logs)	(1) OLS	(2) IV
	Effects on rural roads	
Distance from border (in 10km)	0.121***	0.103***
	(0.030)	(0.035)
Bilateral exports (in logs)	0.274	0.360
	(0.189)	(0.263)
Bilateral exports × Distance from border	-0.015***	-0.017***
	(0.005)	(0.006)
	Additional effects on urban roads	
Road leading to city ¿500k (dummy) × Distance from border	0.079***	0.116**
	(0.033)	(0.054)
Road leading to city ¿500k × Bilateral exports	0.104**	0.166***
	(0.050)	(0.051)
Road leading to city ¿500k × Distance from border × Bilateral exports	-0.039***	-0.048***
	(0.010)	(0.016)
Off-road cells	NO	NO
Controls	ALL	ALL
Road FE	YES	YES
Region-Year FE	YES	YES
Neighbor country-Year FE	YES	YES
Kleibergen-Paap F statistic		12
# Clusters	776	776
# Observations	108,019	108,019

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

In a second step, we consider city locations on the opposite side of the border. This allows us to account for heterogeneity in treatment intensity along a given border segment: a given change in trade openness toward a neighbor country is likely to have a stronger impact in the vicinity of border crossings situated close to a center of economic activity in that neighbor country than in the vicinity of border crossings far away from any neighbor-country economic hub.[35] Our findings conform with expectations: the positive effect of export expansion is greater for border locations close to a city on the other side of the border, and this is true for both urban and rural roads (with respect to city location on the own side of the border).[36]

2.6.3 A mechanism: increased border-region production

Bilateral exports appear to favor the economic development of locations close to the relevant land border. A natural interpretation of this finding is that development takes the form of export-oriented production that is stimulated in border regions. However, other mechanisms are conceivable. It could be that increased activity observed near borders stems mainly from non-traded services that support trading activities, or that it is the result of redistributive policies aimed at spreading trade-related gains towards border regions through public spending.

In order to explore the mechanism behind the estimated trade effects, we focus on the link between agricultural exports and the development of agriculture-dependent border regions. The reason for focusing on agriculture is that there exists fine-grained spatial information on production in that sector of a kind that is not available for manufacturing or services. This allows us to relate localized production to product-level export

[35] This is consistent with the simulations reported by Redding and Rossi-Hansberg (2017, p. 42), which suggest that "(t)he areas that benefit the most are the ones close to but on the opposite side of the border from the large cities".

[36] See Tables 2.26 and 2.27.

data, which in turn makes it possible to explore whether trade expansion is particularly beneficial to border-region development if it occurs in a product the region is specialized in.

Specifically, we can draw on geo-referenced data on the cultivation of 25 different crops at a resolution of 10 × 10 km. This information allows us to establish the main agricultural product for each 10 × 10 km grid cell as the crop that occupies the biggest share of land.[37] Additionally, we compute the share of total land (not just arable land) that is used to grow the dominant crop of a given grid cell, which we then use to weight grid cells according to the importance of their main crop.[38]

Table 2.10: Trade in local crops

Dependent variable: Average light intensity (logs)	(1) OLS	(2) IV	(3) OLS	(4) IV
Distance from border (in 10km)	0.162***	0.199***	0.158***	0.202***
	(0.045)	(0.065)	(0.059)	(0.075)
	Effects of exporting main crop			
Bilateral crop exports (in logs)	0.404**	0.557**		
	(0.199)	(0.275)		
Bilateral crop exports × Distance from border	-0.055***	-0.091***		
	(0.024)	(0.035)		
	Effects of overall exports			
Bilateral exports (in logs)			0.152	0.236
			(0.119)	(0.175)
Bilateral exports × Distance from border			-0.021	-0.026
			(0.019)	(0.023)
Off-road cells	YES	YES	YES	YES
Controls	ALL	ALL	ALL	ALL
Road FE	YES	YES	YES	YES
Region-Year FE	YES	YES	YES	YES
Neighbor country-Year FE	YES	YES	YES	YES
Kleibergen-Paap F statistic		13		12
# Clusters	417	417	417	417
# Observations	159,982	159,982	159,982	159,982

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Note: Cells are weighted according to the share of land used to grow the major agricultural crop.

We estimate two variants of equation (2.1). In columns (1) and (2) of Table 2.10, we present estimates of equations (2.1) using exports of the major crop grown in cell i to neighbor country c' as the trade variable, using neighbor-country c''s tariff on this product and the world price of the respective crop as instruments. In columns (3) and (4), we report specifications with the trade variable defined as overall exports instrumented with average tariffs. Columns (3) and (4) of Table 2.10 therefore show our baseline specification estimated over the sample for which we have information on crops by way of a benchmark for comparison with the estimates for crop-specific exports.[39]

What we find further corroborates our finding that trade causes economic activity to attenuate border shadows, and, importantly, it suggests that the stimulation of local production is a significant mechanism behind

[37] For a list of crops in the sample, see Table 2.30 in the Appendix. See Appendix A for details on the data.

[38] This is not an essential procedure, our results turning out to be very similar in un-weighted regressions.

[39] The underlying dataset concerning crops has less than 20% of land classified as ”cropland” for the large majority of countries (a distinction is made between cropland and pasture). Since we use only information on crops in our estimation, it is not surprising that the number of observations in this exercise shrinks to about a quarter of our baseline sample.

that effect.[40] Table 2.10 shows that both our estimated main effect of export and the interaction effect with distance are noticeably larger when we focus on exports of border regions' dominant crops than when we consider overall trade.

2.7 Conclusion

Our estimates based on world-wide spatially disaggregated data suggest that trade facilitation encourages the economic development of border areas. Given that border regions on average are less developed than interior regions , this predominantly implies a spatially equalizing effect of international trade. The effect emerges very consistently irrespective of how we cut the data: it applies to both developing and advanced economies, and to both rural and urban regions. We also find that border regions benefit in both gross and per-capita terms, suggesting that trade expansion boosts both the populations and the incomes of border regions. Based on detailed information on agricultural production and trade, we moreover establish that trade-related development of border regions is significantly driven by local export-oriented production.

Our results also show that land borders are, in themselves, factors of remoteness. This is a striking result in view of the finding by Henderson *et al.* (2012), also based on night lights data, that inland areas in Sub-Saharan Africa have not, contrary to perceptions, grown more slowly than coastal areas. Combining their observation and ours suggests that it is not landlockedness that holds back economic development, but rather proximity to borders. Many borders in the developing world are, in spite of modernization efforts, still largely dysfunctional; moreover, some are the theater of conflicts between central governments and minorities and between neighboring countries, the two being sometimes linked. Bilateral trade liberalization might therefore represent an underappreciated tool for the appeasement of such conflicts.

Our analysis suggests trade liberalization between neighbor countries to be a potential channel to achieve a more balanced spatial distribution of economic activity within regions located in proximity of the affected border. However, our approach has its limitations. Night lights, although shown elsewhere to be a reliable proxy for local output, is an imperfect measure. Most importantly, as we do not observe wages and local prices and our approach is reduced form, we cannot make rigorously derived statements on local welfare, nor on distributional and incidence effects.[41] Another limitation is that our approach is essentially short-term: we observe a fifteen-year data span, and we take the urban geography and local production specialization patterns as exogenously given.

[40]It should also be noted that part of the increased effect when measuring trade through agricultural exports might be explained by the higher likelihood of local crops to be transported through roads instead of aerial or maritime trade.

[41]We note, though, that quantitative models of economic geography with imperfect intra-national labor mobility, local changes in population, real wages and welfare are strongly correlated (Redding, 2016).

2.8 Data sources

Data on *population* density are taken from the Gridded Population of the World dataset published by the Earth Institute of Columbia University. The dataset contains globally consistent population information by grid cell, drawn from national censuses for varying subnational units (municipalities, census tracts, etc.). The finest available grid-cell resolution is 2.5 arc-minutes, or around 5 km at the equator. The underlying census data cover more than 300,000 national and sub-national administrative units worlwide. Within each of these subnational units, population is distributed across grid cells using an algorithm that takes into account characteristics such as the location of cities and lakes, and the average population density of rural and urban areas in the relevant subnational units.[42]

Since the Gridded Population data are available for the years 1995, 2000, 2005 and 2010, we also use the lights data for those years only throughout this paper.

To measure *trade* liberalization, we draw on bilateral export volumes and simple average applied tariff rates between neighboring countries from the United Nations' UN Comtrade database and the UNCTAD Trade Analysis and Information System (TRAINS) database.

Georeferenced data on the location of national and state *borders* as well as the location of *urban areas* are taken from the Database of Global Administrative Areas hosted by the Hijmans Lab at UC Davis. Data on *roads* are obtained from the ESRI Roads and Highways dataset. Road locations are taken as recorded in the 1995 version of the dataset. Some additional roads appear in later versions of the dataset, but it is not clear whether these represent new roads, upgraded roads, or roads that simply were not recorded in the 1995 data. To avoid problems of attributing light properly to a road in different years, we focus our analysis on roads that were recorded at the beginning of our observation period and follow them through time.

We also consider a number of control variables.

The location of *ports* is taken from the World Port Index published by the US National Geospatial-Intelligence Agencys Maritime Safety Office, and the location of *airports* is taken from Natural Earth and the Emergency and Preparedness Geospatial Information Unit at the World Food Programme (WFP). Only airports that appear in both datasets are kept.

Information on *altitude* is available through the Scripps Institution of Oceanography at the University of California Sand Diego, whose SRTM30 Plus dataset combines sea floor and land elevation data for the entire planet.

Data on the location of *ethnic groups* are taken from the Geo-referencing of Ethnic Groups (GREG) dataset compiled by Weidmann, Rød and Cederman (2010). The GREG data contain global polygon information on the location of ethnic groups. The size of the polygons ranges from 0.6 to 6,954,564 square kilometers.The main source underlying this dataset is the *Atlas Narodov Mira* (Bruk and Apenchenko, 1964), consisting of 57 ethnographic maps drawn from (1) ethnographic and geographic maps assembled by the Institute of Ethnography at the USSR Academy of Sciences, (2) population census data, and (3) ethnographic publications of government agencies, covering all regions of the world at various scales. Despite the the ANM data having been collected in the early 1960s, Weidmann *et al.* (2010, p. 496) argue that ethnic settlement patterns exhibit a lot of inertia, so that it is plausible to also use the GREG data as the basis for measuring ethnic geography in recent times.

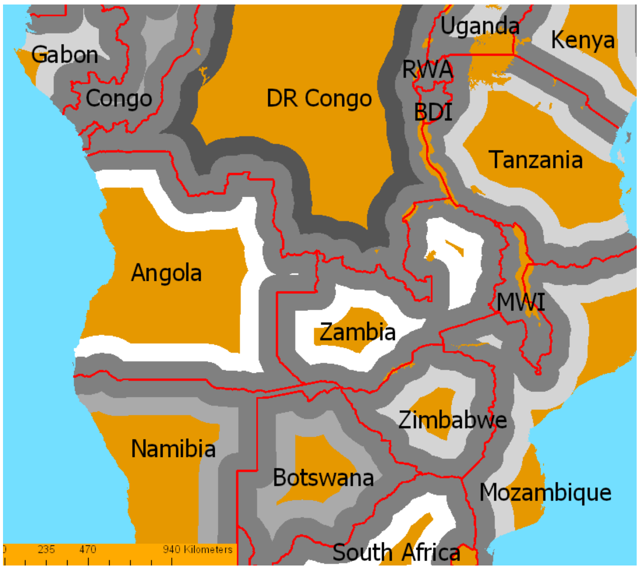
Finally, worldwide data on harvested areas of 175 *crops* are obtained from Ramankutty *et al.* (2008) in 10×10 km grid format. The authors satellite data from the Modetate Resolution Imaging Spectroradiometer (MODIS) and the Satellite Pour l'Observation de la Terre (SPOT) to produce a precise global dataset of agricul-

[42] For details on the computation of the population densities in the cells, see Balk, Deichmann and Yetman (2001).

tural landuse in the year 2000.[43] The dataset is constructed from two different satellite datasets on land-cover and then combined with data from agricultural censuses and FAO data.

2.9 Appendix figures and tables

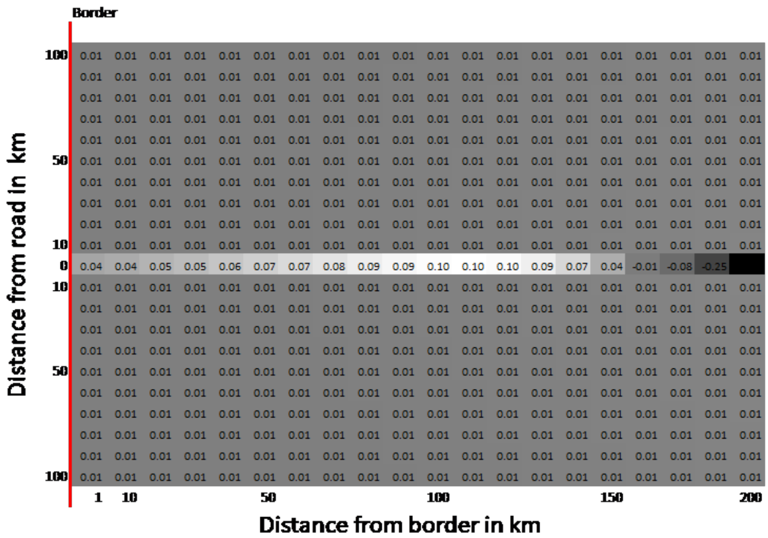
Figure 2.9: Border shadows in Sub-Saharan Africa



Note: The map illustrates average light intensity along roads within two bands of 100km from land borders. The shading of the 100-200km band represents relative light intensity with respect to light intensity in the 0-100km band. White or light gray 100-200km bands are consistent withour baseline definition of border shadows, not conditioned on any covariates.

[43]See Appendix Table 2.30 for a list and summary statistics of the covered crops.

Figure 2.10: Predicted absolute change in light intensity associated with a doubling of exports



Note: The graph shows predicted absolute changes in light intensity after a doubling of exports starting from a scenario with trade set to the value of the 25th percentile in our data (i.e. starting from the values presented in Figure 3.6, based on a specification featuring road fixed effects, year fixed effects, all control variables and exports instrumented with tariffs (Table 2.19, column 4). Darker colors symbolize lower light intensity.

Table 2.11: Summary statistics: Baseline sample (on-road and off-road grid cells)

Variable	Mean	Std. Dev.	Min.	Max.	N
Average light intensity	3.93	8.93	0	63	649,109
Distance from border	92.14	48.52	0.00	200	649,109
Total exports to neighbor country (in 100 mio US dollar)	324.79	652.93	0.00	2131.69	649,109
Simple average applied tariff rate	4.33	6.96	0	104.34	649,109
Population density (people/km^2)	32.51	255.11	0	31,735	649,109
Regional trade agreement dummy	0.67	0.47	0	1	649,109
Altitude	766.98	886.92	-405	6659	649,109
Same ethnicity dummy	0.06	0.20	0	1	649,109
Port dummy	0.03	0.18	0	1	649,109
Airport dummy	0.25	0.43	0	1	649,109
Distance from port (if port dummy = 1)	72.69	47.50	1.88	189.86	22,105
Distance from airport (if airport dummy = 1)	62.21	46.14	1.09	199.34	168,362
Dummy for road leading to city $> 100,000$ inhabitants	0.35	0.48	0	1	649,109
Dummy for road leading to city $> 500,000$ inhabitants	0.02	0.15	0	1	649,109
Dummy for light = 0	0.53	0.35	0	1	649,109
Average light intensity at 0 km distance	2.97				
Average light intensity at 100 km distance	3.41				
Average light intensity at 200 km distance	4.75				

Table 2.12: Summary statistics: On-road grid cells

Variable	Mean	Std. Dev.	Min.	Max.	N
Average light intensity	10.86	11.81	0	63	113,512
Distance from border	76.21	40.75	0.00	200	113,512
Total exports to neighbor country (in 100 mio US dollar)	378.21	681.84	0.00	2131.69	113,512
Simple average applied tariff rate	3.72	6.56	0	104.34	113,512
Population density (people/km^2)	130.45	471.89	0	31,735	113,512
Regional trade agreement dummy	0.69	0.46	0	1	113,512
Altitude	580.58	703.14	-405	5540	113,512
Same ethnicity dummy	0.03	0.17	0	1	113,512
Port dummy	0.04	0.19	0	1	113,512
Airport dummy	0.28	0.45	0	1	113,512
Distance from port (if port dummy = 1)	56.38	41.76	1.88	189.86	22,105
Distance from airport (if airport dummy = 1)	50.43	39.60	1.09	199.34	168,362
Dummy for road leading to city $> 100,000$ inhabitants	0.37	0.48	0	1	113,512
Dummy for road leading to city $> 500,000$ inhabitants	0.05	0.21	0	1	113,512
Dummy for light = 0	0.19	0.47	0	1	113,512
Average light intensity at 0 km distance	8.33				
Average light intensity at 100 km distance	10.06				
Average light intensity at 200 km distance	11.54				
Number of conflict events (20 $\times$ 20 km cells, Africa only)	0.53	4.28	0	109	9,981
Conflict fatalities (20 $\times$ 20 km cells, Africa only)	5.95	58.19	0	2769	9,981

Table 2.13: Summary statistics: Off-road grid cells

Variable	Mean	Std. Dev.	Min.	Max.	N
Average light intensity	1.23	3.70	0	37.8	535,597
Distance from border	95.52	43.88	0.00	200	535,597
Total exports to neighbor country (in 100 mio US dollar)	314.30	646.59	0.00	2131.69	535,597
Simple average applied tariff rate	4.46	7.04	0	104.34	535,597
Population density (people/km^2)	11.75	88.12	0	9,190	535,597
Regional trade agreement dummy	0.66	0.47	0	1	535,597
Altitude	806.49	916.32	-405	6659	535,597
Same ethnicity dummy	0.09	0.22	0	1	535,597
Port dummy	0.03	0.17	0	1	535,597
Airport dummy	0.25	0.43	0	1	535,597
Distance from port (if port dummy = 1)	76.15	47.99	1.88	189.86	17,881
Distance from airport (if airport dummy = 1)	64.96	47.11	1.09	199.34	136,521
Dummy for road leading to city $> 100,000$ inhabitants	0.34	0.48	0	1	535,597
Dummy for road leading to city $> 500,000$ inhabitants	0.02	0.13	0	1	535,597
Dummy for light = 0	0.60	0.36	0	1	535,597
Average light intensity at 0 km distance	0.89				
Average light intensity at 100 km distance	1.09				
Average light intensity at 200 km distance	1.68				

Table 2.14: Sensitivity to different border distance cutoffs

Dependent variable: Average light intensity (logs)				
	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Maximum distance from border (in km)	200	150	100	50
Distance from border (in 10km)	0.178***	0.195***	0.206**	0.489***
	(0.065)	(0.070)	(0.072)	(0.125)
Bilateral exports (in logs)	0.556	0.450	0.386*	0.334**
	(0.513)	(0.364)	(0.219)	(0.164)
Bilateral exports × Distance from border	-0.035***	-0.025***	-0.020**	-0.029*
	(0.013)	(0.008)	(0.010)	(0.016)
Off-road cells	YES	YES	YES	YES
Controls	ALL	ALL	ALL	ALL
Road FE	YES	YES	YES	YES
Region-Year FE	YES	YES	YES	YES
Neighbor country-Year FE	YES	YES	YES	YES
First-stage F statistic	13	13	13	12
# Clusters	1,639	1,639	1,639	1,608
# Observations	648,783	633,105	596,279	491,880

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.15: Baseline estimates - excluding border crossings

Dependent variable: Average light intensity (logs)	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
	Effects on grid cells along road			
Distance from border (in 10km)	0.179***	0.188***	0.185***	0.193***
	(0.029)	(0.055)	(0.028)	(0.063)
Bilateral exports (in logs)	0.060	0.475	0.075	0.541
	(0.138)	(0.503)	(0.142)	(0.500)
Bilateral exports × Distance from border	-0.029***	-0.031**	-0.031***	-0.033**
	(0.010)	(0.015)	(0.011)	(0.014)
	Additional effects on off-road grid cells			
Off-road × Distance from border			-0.157***	-0.147*
			(0.032)	(0.077)
Off-road × Distance from road			-0.010***	-0.009***
			(0.002)	(0.003)
Off-road × Bilateral exports			0.073	0.800
			(0.180)	(1.382)
Off-road × Bilateral exports × Distance from border			0.033***	0.032*
			(0.013)	(0.018)
Off-road × Bilateral exports × Distance from road			-0.002***	-0.002**
			(0.000)	(0.001)
Off-road cells	NO	NO	YES	YES
Controls	ALL	ALL	ALL	ALL
Road FE	YES	YES	YES	YES
Region-Year FE	YES	YES	YES	YES
Neighbor country-Year FE	YES	YES	YES	YES
First-stage F statistic		15		13
# Clusters	812	812	1,639	1,639
# Observations	103,674	103,674	598,113	598,113

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.16: Polynomial regression

Dependent variable: Average light intensity (logs)		
	(1)	(2)
	OLS	OLS
Distance from border (in 10km)	0.228***	0.221***
	(0.055)	(0.086)
Distance from border (in 10km) squared	-0.006	-0.003
	(0.005)	(0.006)
Bilateral exports (in logs)	0.066	0.433
	(0.142)	(0.513)
Bilateral exports $\times$ distance to border	-0.027***	-0.029***
	(0.008)	(0.011)
Bilateral exports $\times$ Distance from border squared	-0.001	-0.001
	(0.001)	(0.002)
Off-road cells	NO	NO
Controls	ALL	ALL
Road FE	YES	YES
Region-Year FE	YES	YES
Neighbor country-Year FE	YES	YES
First-stage F statistic		11
# Clusters	812	812
# Observations	113,289	113,289

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.17: Baseline estimates with different ways of clustering

Dependent variable: Average light intensity (logs)	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
	Effects on grid cells along road			
Distance from border (in 10km)	0.170	0.183	0.171	0.178
	(0.030)	(0.058)	(0.028)	(0.065)
	[0.026]	[0.044]	[0.023]	[0.057]
	{0.021}	{0.036}	{0.018}	{0.042}
Bilateral exports (in logs)	0.058	0.482	0.080	0.556
	(0.146)	(0.519)	(0.155)	(0.513)
	[0.121]	[0.402]	[0.134]	[0.382]
	{0.081}	{0.299}	{0.095}	{0.277}
Bilateral exports × Distance from border	-0.030	-0.033	-0.033	-0.035
	(0.010)	(0.016)	(0.010)	(0.016)
	[0.009]	[0.013]	[0.008]	[0.010]
	{0.006}	{0.008}	{0.006}	{0.009}
	Additional effects on off-road grid cells			
Off-road × Distance from border			-0.156	-0.144
			(0.032)	(0.076)
			[0.026]	[0.061]
			{0.020}	{0.050}
Off-road × Distance from road			-0.009	-0.008
			(0.001)	(0.002)
			[0.001]	[0.001]
			{0.001}	{0.001}
Off-road × Bilateral exports			0.069	0.805
			(0.187)	(1.405)
			[0.140]	[1.162]
			{0.106}	{0.769}
Off-road × Bilateral exports × Distance from border			0.032	0.030
			(0.012)	(0.018)
			[0.009]	[0.012]
			{0.007}	{0.010}
Off-road × Bilateral exports × Distance from road			-0.002	-0.002
			(0.000)	(0.001)
			[0.000]	[0.001]
			{0.000}	{0.001}
Off-road cells	NO	NO	YES	YES
Controls	ALL	ALL	ALL	ALL
Road FE	YES	YES	YES	YES
Region-Year FE	YES	YES	YES	YES
Neighbor country-Year FE	YES	YES	YES	YES
First-stage F statistic		(15)		(13)
		[16]		[15]
		{17}		{17}
# Observations	648,783	648,783	648,783	648,783

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses (1,639 clusters).

Standard errors clustered clustered at road level in brackets (5,250 clusters).

Spatially clustered standard errors in curly brackets.

Table 2.18: First stage results for within-road specification

Dependent variable:	Bilateral exports	Bilateral exports × Distance from border	Bilateral exports × Off-road	Bilateral exports × Off-road × Distance from border	Bilateral exports × Off-road × Distance from road
	(1)	(2)	(3)	(4)	(5)
Distance from border (in 10km)	0.001	5.567***	-0.000	0.144**	0.138**
	(0.001)	(0.605)	(0.000)	(0.062)	(0.059)
Tariffs imposed by neighbor country (in logs +1)	-0.124***	0.471***	0.188**	0.348**	0.200***
	(0.014)	(0.149)	(0.089)	(0.168)	(0.033)
Tariffs imposed by neighbor country × Distance from border	-0.000	-0.140***	-0.005	0.015**	0.005*
	(0.000)	(0.032)	(0.009)	(0.007)	(0.003)
Off-road × Distance from border	-0.000	-0.044	0.001	5.521***	2.368
	(0.001)	(0.917)	(0.001)	(0.689)	(1.882)
Off-road × Distance from road	-0.000*	0.002	-0.000*	0.002	2.108***
	(0.000)	(0.005)	(0.000)	(0.005)	(0.578)
Off-road × Tariffs imposed by neighbor country	0.009	0.010	-0.015	0.481***	3.833***
	(0.019)	(0.216)	(0.014)	(0.156)	(1.467)
Off-road × Tariffs imposed by neighbor country × Distance from border	-0.000	-0.002	-0.000	-0.141***	-0.060
	(0.000)	(0.047)	(0.000)	(0.034)	(0.069)
Off-road × Tariffs imposed by neighbor country × Distance from road	0.000	-0.000	0.000	-0.000	-0.133***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.038)
Off-road cells	YES	YES	YES	YES	YES
Controls	ALL	ALL	ALL	ALL	ALL
Road FE	YES	YES	YES	YES	YES
Region-Year FE	YES	YES	YES	YES	YES
Neighbor country-Year FE	YES	YES	YES	YES	YES
F statistic	22	26	16	24	17
Clusters	1,639	1,639	1,639	1,639	1,639
Observations	648,783	648,783	648,783	648,783	648,783

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.19: Baseline estimates - Different fixed effects

Dependent variable: light intensity per 10x10 km grid cell (logs)	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
			Effects on grid cells along road					
Distance from border (in 10km)	0.118***	0.127***	0.111***	0.116***	0.054**	0.089**	0.069**	0.118**
	(0.026)	(0.036)	(0.026)	(0.042)	(0.027)	(0.039)	(0.027)	(0.047)
Bilateral exports (in logs)	0.188***	0.383***	0.207***	0.454***	0.093	0.284	0.080	0.337*
	(0.063)	(0.115)	(0.049)	(0.107)	(0.078)	(0.189)	(0.074)	(0.179)
Bilateral exports × Distance from border	-0.016	-0.019	-0.018*	-0.019	-0.019**	-0.031**	-0.019**	-0.030**
	(0.010)	(0.012)	(0.010)	(0.012)	(0.008)	(0.013)	(0.008)	(0.012)
			Additional effects on off-road grid cells					
Off-road × Distance from border			-0.096***	-0.088*			-0.048***	-0.094***
			(0.030)	(0.052)			(0.019)	(0.029)
Off-road × Distance from road			-0.010***	-0.009***			-0.010***	-0.010***
			(0.001)	(0.002)			(0.001)	(0.002)
Off-road × Bilateral exports			-0.049	-0.014			0.049	-0.042
			(0.051)	(0.082)			(0.084)	(0.227)
Off-road × Bilateral exports × Distance from border			0.018*	0.016			0.015*	0.028**
			(0.011)	(0.014)			(0.008)	(0.013)
Off-road × Bilateral exports × Distance from road			-0.002***	-0.003***			-0.002***	-0.002**
			(0.000)	(0.001)			(0.000)	(0.001)
			Control Variables					
Altitude (in 100m)	-0.093***	-0.086***	-0.063***	-0.066***	-0.100***	-0.090***	-0.050***	-0.057***
	(0.008)	(0.010)	(0.005)	(0.007)	(0.007)	(0.009)	(0.008)	(0.011)
Slope	-0.000	-0.001*	-0.001	-0.001	-0.000	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.002)
Port closer than next land border (dummy)	1.559***	1.547***	1.052***	1.033***	1.757***	1.671***	1.588***	1.588***
	(0.402)	(0.391)	(0.399)	(0.383)	(0.377)	(0.359)	(0.277)	(0.273)
Port dummy × Distance from port	-0.141***	-0.136***	-0.124**	-0.117***	-0.153***	-0.118***	-0.178***	-0.177***
	(0.046)	(0.042)	(0.051)	(0.041)	(0.041)	(0.044)	(0.055)	(0.061)
Airport closer than next land border (dummy)	1.317***	1.306***	0.764***	0.759***	1.177***	1.122***	0.764***	0.761***
	(0.122)	(0.130)	(0.125)	(0.127)	(0.119)	(0.129)	(0.092)	(0.091)
Airport dummy × Distance from airport	-0.116***	-0.112***	-0.049***	-0.047***	-0.064***	-0.039	-0.035***	-0.034***
	(0.033)	(0.037)	(0.014)	(0.015)	(0.030)	(0.032)	(0.011)	(0.012)
Same ethnicity on both sides of border (dummy)	-0.141	-0.099	-0.113	-0.111	-0.060	-0.052	-0.059	-0.049
	(0.121)	(0.119)	(0.091)	(0.183)	(0.129)	(0.114)	(0.114)	(0.139)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	YES	YES	YES	YES	NO	NO	NO	NO
Year FE	YES	YES	YES	YES	NO	NO	NO	NO
Region-Year FE	NO	NO	NO	NO	YES	YES	YES	YES
Neighbor country-Year FE	NO	NO	NO	NO	YES	YES	YES	YES
Kleibergen-Paap F statistic		23		13		16		12
# Clusters	835	835	1,671	1,671	813	813	1,640	1,640
# Observations	113,512	113,512	649,109	649,109	113,292	113,292	648,787	648,787

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.20: Baseline estimates - including 2013

Dependent variable: Average light intensity (logs)	(1) OLS	(2) IV	(3) OLS	(4) IV
	Effects on grid cells along road			
Distance from border (in 10km)	0.181***	0.190***	0.185***	0.195***
	(0.028)	(0.054)	(0.022)	(0.060)
Bilateral exports (in logs)	0.070	0.441	0.095	0.502
	(0.130)	(0.430)	(0.109)	(0.427)
Bilateral exports × Distance from border	-0.032***	-0.036***	-0.035***	-0.036**
	(0.010)	(0.014)	(0.011)	(0.016)
	Additional effects on off-road grid cells			
Off-road × Distance from border			-0.160***	-0.152**
			(0.033)	(0.075)
Off-road × Distance from road			-0.009***	-0.009***
			(0.002)	(0.003)
Off-road × Bilateral exports			0.084	0.731
			(0.161)	(1.139)
Off-road × Bilateral exports × Distance from border			0.034***	0.033**
			(0.014)	(0.016)
Off-road × Bilateral exports × Distance from road			-0.002***	-0.002**
			(0.000)	(0.001)
Off-road cells	NO	NO	YES	YES
Controls	ALL	ALL	ALL	ALL
Road FE	YES	YES	YES	YES
Region-Year FE	YES	YES	YES	YES
Neighbor country-Year FE	YES	YES	YES	YES
First-stage F statistic		16		14
# Clusters	1,012	1,012	2,050	2,050
# Observations	150,680	150,680	813,871	813,871

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.21: Overglow, small countries and landlocked countries

Dependent variable: Average light intensity (logs)	(1) IV	(2) IV	(3) IV	(4) IV
Sample	Baseline	Overglow correction	No small countries	Landlocked countries
Distance from border (in 10km)	0.178***	0.174**	0.169***	0.359***
	(0.065)	(0.084)	(0.059)	(0.092)
Bilateral exports (in logs)	0.556	0.538	0.541	0.487
	(0.513)	(0.470)	(0.496)	(0.409)
Bilateral exports × Distance from border	-0.035**	-0.038**	-0.041***	-0.046***
	(0.016)	(0.019)	(0.020)	(0.020)
Controls	ALL	ALL	ALL	ALL
Road FE	YES	YES	YES	YES
Region-Year FE	YES	YES	YES	YES
Neighbor country-Year FE	YES	YES	YES	YES
First-stage F statistic	13	13	12	10
# Clusters	1,639	1,639	1,492	243
# Observations	648,783	648,783	586,661	66,292

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors clustered at road and country-pair-year level in parentheses.

Table 2.22: Baseline results: Grid cell FE

Dependent variable: Average light intensity (logs)	(1) OLS	(2) IV
Effect on grid cells along road		
Bilateral exports (in logs)	0.474***	3.127***
	(0.024)	(0.804)
Bilateral exports × Distance from border	-0.006	-0.215***
	(0.006)	(0.071)
Additional effect on off-road grid cells		
Off-road × Bilateral exports	-0.227***	-3.854***
	(0.014)	(1.018)
Off-road × Bilateral exports × Distance from road	-0.002	0.206***
	(0.004)	(0.084)
Off-road × Bilateral exports × Distance from border	-0.002***	0.027**
	(0.000)	(0.011)
Controls	ALL	ALL
Grid cell FE	YES	YES
First-stage F statistic		3
# Clusters	5,255	5,255
# Obervations	648,572	648,572

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.23: Heterogeneous effects with respect to differences in 1990 GNI

Dependent variable: Average light intensity (logs)	(1) OLS	(2) IV
Distance from border (in 10km)	0.161***	0.169***
	(0.025)	(0.027)
Bilateral exports (in logs)	0.050	0.405
	(0.113)	(0.417)
Bilateral exports × Distance from border	-0.026***	-0.030**
	(0.008)	(0.009)
GNI 1990 (in logs) × Distance from border	0.021***	0.023**
	(0.007)	(0.011)
GNI 1990 × Bilateral exports	0.040***	0.065***
	(0.012)	(0.021)
GNI 1990 × Distance from border × Bilateral exports	-0.006***	-0.004***
	(0.002)	(0.001)
Off-road cells	NO	NO
Controls	ALL	ALLL
Road FE	YES	YES
Region-Year FE	YES	YES
Neighbor country-Year FE	YES	YES
Kleibergen-Paap F statistic		12
# Clusters	812	812
# Observations	113,289	113,289

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.24: Heterogeneous effects with respect to differences in 2015 GNI

Dependent variable: Average light intensity (logs)	(1) OLS	(2) IV
Distance from border (in 10km)	0.146***	0.161***
	(0.026)	(0.031)
Bilateral exports (in logs)	0.048	0.392
	(0.094)	(0.364)
Bilateral exports × Distance from border	-0.023***	-0.025**
	(0.005)	(0.006)
GNI 2015 (in logs) × Distance from border	0.024***	0.028***
	(0.007)	(0.009)
GNI 2015 × Bilateral exports	0.042***	0.068***
	(0.014)	(0.022)
GNI 2015 × Distance from border × Bilateral exports	-0.007***	-0.004**
	(0.002)	(0.002)
Off-road cells	NO	NO
Controls	ALL	ALLL
Road FE	YES	YES
Region-Year FE	YES	YES
Neighbor country-Year FE	YES	YES
Kleibergen-Paap F statistic		12
# Clusters	812	812
# Observations	113,289	113,289

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.25: Border cities (100,000 cut-off)

Dependent variable: Average light intensity (logs)	(1) OLS	(2) IV
Effects on rural roads		
Distance from border (in 10km)	0.105***	0.092***
	(0.032)	(0.032)
Bilateral exports (in logs)	0.254	0.343
	(0.177)	(0.240)
Bilateral exports × Distance from border	-0.013***	-0.015***
	(0.004)	(0.005)
Additional effects on urban roads		
Road leading to city $> 100k$ (dummy) × Distance from border	0.087***	0.129**
	(0.038)	(0.059)
Road leading to city $> 100k$ × Bilateral exports	0.117***	0.176***
	(0.039)	(0.052)
Road leading to city $> 100k$ × Distance from border × Bilateral exports	-0.040***	-0.050***
	(0.013)	(0.018)
Off-road cells	NO	NO
Controls	ALL	ALL
Road FE	YES	YES
Region-Year FE	YES	YES
Neighbor country-Year FE	YES	YES
Kleibergen-Paap F statistic		12
# Clusters	776	776
# Observations	100,928	100,928

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.26: Cities across the border

Dependent variable: Average light intensity (logs)	(1) OLS	(2) IV
Effects on roads without a city across the border		
Distance from border (in 10km)	0.108***	0.091***
	(0.026)	(0.027)
Bilateral exports (in logs)	0.233	0.336
	(0.176)	(0.231)
Bilateral exports $\times$ Distance from border	-0.010**	-0.011**
	(0.005)	(0.005)
Additional effects on roads with a city across the border		
Road leading to foreign city $> 500k$ (dummy) $\times$ Distance from border	0.074**	0.099**
	(0.035)	(0.048)
Road leading to foreign city $> 500k$ $\times$ Bilateral exports	0.079*	0.121*
	(0.049)	(0.078)
Road leading to foreign city $> 500k$ $\times$ Distance from border $\times$ Bilateral exports	-0.024***	-0.031***
	(0.010)	(0.016)
Off-road cells	NO	NO
Controls	ALL	ALL
Road FE	YES	YES
Region-Year FE	YES	YES
Neighbor country-Year FE	YES	YES
Kleibergen-Paap F statistic		15
# Clusters	812	812
# Observations	113,289	113,289

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.27: Cities on either side of the border

Dependent variable: Average light intensity (logs)	(1) OLS	(2) IV
Effects on rural roads		
Distance from border (in 10km)	0.117***	0.109***
	(0.032)	(0.033)
Bilateral exports (in logs)	0.212	0.230
	(0.168)	(0.184)
Bilateral exports × Distance from border	-0.009**	-0.014**
	(0.004)	(0.006)
Additional effects on urban roads		
Road leading to city $> 500k$ × Distance from border	0.061**	0.072**
	(0.031)	(0.033)
Road leading to city $> 500k$ × Bilateral exports	0.120**	0.187**
	(0.052)	(0.085)
Road leading to city $> 500k$ × Distance from border × Bilateral exports	-0.028***	-0.039**
	(0.011)	(0.017)
Additional effects on roads with a city across the border		
Road leading to foreign city $> 500k$ × Distance from border	0.010	0.028
	(0.009)	(0.030)
Road leading to foreign city $> 500k$ × Bilateral exports	0.034	0.067*
	(0.031)	(0.037)
Road leading to foreign city $> 500k$ × Distance from border × Bilateral exports	-0.007*	-0.012
	(0.005)	(0.010)
Off-road cells	NO	NO
Controls	ALL	ALL
Road FE	YES	YES
Region-Year FE	YES	YES
Neighbor country-Year FE	YES	YES
Kleibergen-Paap F statistic		11
# Clusters	776	776
# Observations	108,019	108,019

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.28: Border cities - Industrialized countries

Dependent variable: Average light intensity (logs)	(1) OLS	(2) IV
Effects on rural roads		
Distance from border (in 10km)	0.173***	0.187***
	(0.049)	(0.061)
Bilateral exports (in logs)	0.197*	0.282*
	(0.127)	(0.160)
Bilateral exports × Distance from border	-0.016***	-0.022***
	(0.006)	(0.007)
Additional effects on urban roads		
Road leading to city $> 500k$ × Distance from border	0.069**	0.085**
	(0.032)	(0.038)
Road leading to city $> 500k$ × Bilateral exports	0.314*	0.392*
	(0.171)	(0.223)
Road leading to city $> 500k$ × Distance from border × Bilateral exports	-0.030***	-0.041***
	(0.011)	(0.012)
Off-road cells	NO	NO
Controls	ALL	ALL
Road FE	YES	YES
Region-Year FE	YES	YES
Neighbor country-Year FE	YES	YES
Kleibergen-Paap F statistic		8
# Clusters	207	207
# Observations	36,635	36,635

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.29: Border cities - Developing countries

Dependent variable: Average light intensity (logs)	(1) OLS	(2) IV
Effects on rural roads		
Distance from border (in 10km)	0.113**	0.125**
	(0.054)	(0.061)
Bilateral exports (in logs)	0.186	0.239
	(0.160)	(0.195)
Bilateral exports × Distance from border	-0.011**	-0.015**
	(0.005)	(0.007)
Additional effects on urban roads		
Road leading to city $> 500k$ × Distance from border	0.099**	0.108**
	(0.047)	(0.048)
Road leading to city $> 500k$ × Bilateral exports	0.370	0.444
	(0.286)	(0.295)
Road leading to city $> 500k$ × Distance from border × Bilateral exports	-0.036***	-0.055***
	(0.013)	(0.017)
Off-road cells	NO	NO
Controls	ALL	ALL
Road FE	YES	YES
Region-Year FE	YES	YES
Neighbor country-Year FE	YES	YES
Kleibergen-Paap F statistic		12
# Clusters	591	591
# Observations	71,298	71,298

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 2.30: Summary statistics: Main crops

Crop	# cells	Percent
Barley	29,170	3.39
Cassava	64,196	7.45
Cotton	32,487	3.77
Groundnut	1,644	0.19
Maize	179524	20.84
Millet	24,988	2.90
Oilpalm	11,616	1.35
Potato	24,670	2.86
Rice	76,834	8.92
Rye	48	0.01
Sorghum	49,576	5.76
Soybean	63,955	7.43
Sugarcane	7,356	0.85
Sunflower	5,308	0.62
Wheat	289,958	33.66
Total	861,330	100.00

Chapter 3

Trade Facilitation and Spatial Patterns of Economic Activity: Evidence from the Intensive Margin

ALEXANDER HIMBERT

This paper uses a multi-country panel of firm-level data tracked over seven years (the Orbis dataset) to explore the relationship between trade and the spatial dispersion of economic activity. I find that firms located closer to land borders are, on average, smaller in terms of employment and turnover than those located further away. I also show that exports of goods, instrumented by a measure of trade facilitation, reduce the agglomeration of activity near large urban centers typically away from land borders. My instrument is the product of the Logistics Performance Indices (LPIs) of the two countries straddling the land border closest to each firm. The use of Orbis data allows me to show that this pattern holds, qualitatively, for all sectors, i.e. that overland trade boosts not just directly trade-related services, but also other service activities as well as manufacturing, although the magnitude of the effect and the difference between export-driven and import-driven effects vary substantially across sectors. These cross-sectoral similarities and differences contribute to our understanding of the underlying causal channels.

3.1 Introduction

The quantification of the effects of trade facilitation on economic activity and employment has long been a central topic in the economic literature. Over the past two decades, both the political discourse about winners and losers of liberalized trade as well as the availability of more precise data have shifted the focus from analysing effects of trade on aggregate country wide outcomes towards approaches allowing for diverse reactions within a country to the same trade shock. Two possible dimensions of heterogeneity within economies have received particular attention in this regard: the spatial distribution of the gains of trade and the reaction of different sectors to increases in trade volumes. This paper aims to shed light on both of these dimensions and the interplay between them by estimating the effect of trade facilitation on individual firm performance at different locations in different sectors.

The issue of which regions and which geographic locations react more favourably to increases in trade has been investigated in the light of factors such as market access or economies of scale. Specifically, the question of whether trade spurs further concentration of economic activity in urban centres or whether it disperses activity towards previously remote areas has been of both academic as well as political interest. As noted in Michalopoulos and Papaioannou (2013) for African data and in Brülhart et al. (2018) for a world wide sample, land border regions tend to lag behind in economic development in the majority of countries, leading to a variety of undesirable outcomes. However, the theoretical literature is able to accommodate scenarios of concentration as well as dispersion of activity in reaction to increases in trade and has not yet found a consensus view.

Additionally, trade has been shown to have differing effects on different sectors within an economy. Papers like Dix-Carneiro (2014) or Caliendo, Dvorkin and Parro (2017) develop dynamic trade models with spatially distinct labour markets. In these models, factors like different degrees of exposure to foreign trade or sector-specific experience generate frictions that cause different regions within an economy to react differently to trade shocks based on their prior sectoral composition.

Regarding the spatial dimension of the impact of trade on economic activity, this paper provides new empirical evidence by analysing the effect of trade facilitation on firm level activity at different geographic locations. Instead of focusing on aggregate effects of concentration and dispersion as the literature has done previously, I analyse effects on the intensive margin of existing firms To this end, I construct a precisely geo-referenced dataset of over 1 million firms of different sectors in border regions of 24 countries . I measure firm level turnovers and employment depending on the distance to the nearest land border, and find that in the absence of trade, firm turnover is increased by about 14% for every ten additional kilometres to the border, while employment goes up by about 7% for every ten kilometres of distance to the border.[1] This initial "remoteness" of borders is significantly attenuated when trade flows to the neighbour country increase in reaction to improvements in the efficiency of the customs clearance process in the two countries straddling a border. More precisely, bilateral exports between two countries are instrumented with the product of the scores of the two countries in the "Customs" dimension of the World Bank's Logistics Performance Index. Doing so, I identify a causal effect of increases in trade attenuating the gradients of firm size and turnover along distance from border at the intensive margin. The effects prove robust to the inclusion of geographic control variables and region-sector-year fixed effects controlling for time-varying political changes at the local level. Taking into account firm location via individual firm addresses also allows for effects to vary on a very fine spatial resolution instead of being limited to aggregate administrative level data. The empirical approach in this paper also

[1] As coastal areas benefit from easy access to international markets through ports and therefore do not generally do not impede regional economic development , the term "border regions" in this paper only refers to land borders.

allows to study the link between the geographic distribution of the effects of trade and the underlying sectoral composition of an economy. I find that even within the same region-sector-year, distance to the nearest land border remains a significant factor influencing the reaction of otherwise similar firms to trade facilitation. Furthermore, the distance gradient of firm activity exists even within sectors.[2] However, the slope of this gradient is not identical, but both the initial gradient as well as the size of the dispersing effect of trade on economic activity varies across different sectors.

Exploiting the heterogeneity of firms from different sectors in the Orbis data, I show that the attenuating effect of trade on the gradients of economic activity with respect to the distance from borders holds, qualitatively, for all sectors. Overland trade boosts not just directly trade-related services, but also other service activities as well as manufacturing, although the magnitude of the effect and the difference between export-driven and import-driven effects vary substantially across sectors. When compared to other sectors, manufacturing is found to gain the least from increases in exports and even lose both turnovers as well as employment after increases in imports. Other sectors such as transportation and storage or wholesale and retail trade reap bigger benefits from increased trade flows. These cross-sectoral similarities and differences contribute to the understanding of the underlying causal channels of the impact of trade facilitation on the spatial distribution of economic activity.

The remainder of the paper is organized as follows. Section 2 puts the paper in the context of related literature, Section 3 describes the data, Section 4 discusses the estimation and identification strategy, Section 5 presents baseline results, Section 6 analyses heterogeneous effects across different sectors and different firm categories, and Section 7 concludes.

3.2 Background

3.2.1 Trade and the spatial distribution of economic activity

The empirical analysis in this paper aims to estimate the impact of a firm's location relative to the nearest national land border on its size (as measured by employment) and activity (as measured by turnover), and the impact of trade facilitation at the respective border on the relationship between firm location and firm activity.

The impact of trade on the spatial distribution of economic activity has been analysed by an active and growing strand of both theoretical as well as empirical literature over the past three decades.[3] From a theoretical point of view, it has proven possible to develop models that argue for a dispersing effect of trade on economic activity as well as models in which trade leads to a further spatial concentration of activity in previously existing major cities. Historically, these models developed from ”urban system models” and ”new economic geography” models. In the urban systems models (e.g. Henderson, 1982; Rauch, 1991), trade is often predicted to further concentrate economic activity due to economies of scale in large economic centres.[4] The models developed from the new economic geography setting (e.g. Krugman and Livas Elizondo, 1996; Monfort and Nicolini, 2000) on the other tend to predict activity to disperse away from economic centers into previously remote border regions, that reap the benefits of easier access to the foreign market after a liberalization of trade.[5]

---

[2] For the sake of easier interpretation of the results, distance is entered into all regressions as linear, while trade as well as turnover and employment are measured in logs. This means that the gradient of firm activity along distance is actually a semi-elasticity. The term gradient is used loosely throughout the paper for the sake of easier readability.

[3] For an extensive survey of the theoretical literature, see Redding and Rossi-Hansberg, 2017. For a survey of empirical studies, see Brülhart, 2011.

[4] With the assumption of the border region being specialised in sectors with comparative advantages, the models are however also compatible with a scenario in which trade disperses activity.

[5] Again, it is also possible to generate spatial concentration within these models, depending on the assumptions made about the strength of agglomeration economies in economic centres.

In recent years, theoretical models have become more complex to allow for heterogeneity of the effects across different border regions.[6] Multi-location models such as Redding (2016) and Redding and Rossi-Hansberg (2017) are able to simulate vast diversity in geography and initial economic conditions that generate increases in variables like employment and real wages for some border regions while predicting activity to be distributed away from others. Overall, the theoretical literature so far does not provide any clear cut answers about the impact of trade on spatial distribution of economic activity.

Figure 3.1 graphically represents the two scenarios that can be derived from the theoretical literature. In panel (a), the underlying assumption is that the benefits of increased trade flows across the border are reaped mostly by existing firms towards the center of of a country at locations that already hosted larger firms before trading. This would be the case if benefits of economies of scale and agglomeration in major economic centres outweigh the locational advantage of being located more closely to the foreign market that locations close to the border have. Panel (b) represents the opposite assumption, where firm level growth is bigger the closer one gets to the border as previously remote places further away from big domestic markets now reap the benefits of proximity to the foreign market. This paper uses firm level data to investigate which of the scenarios is prevalent at the intensive margin of activity of existing firms and whether a similar initial gradient can be observed at the firm level, in which case otherwise identical companies would be bigger and more active as one moves from the border towards the center of a country.

Figure 3.1 assumes in both scenarios that the general effects of trade on activity in the border region are positive. While this assumption is generally accepted whenever trade is measured by exports such as in the majority of the analysis of this paper, predictions are more diverse concerning the impact of increased imports on affected regions.[7] Sections 5.2. and 6.1. test this scenario both for overall firm activity as well as for different sectors.

Empirically, the literature is less ambiguous, with the majority of empirical studies pointing towards trade as a factor of spatial dispersion. Brülhart (2011) surveys the empirical literature in this field, finding that ten out of eleven cross-country analyses provide evidence for economic activity to be dispersed after a trade liberalization.[8] Most prominently, papers that used the fall of the iron curtain as natural experiment such as Redding and Sturm (2008) or Brülhart, Carre and Trionfetti (2012) find increases in population, employment and wages to be significantly larger the closer one gets to the border after the liberalization of trade.

Recently, Brülhart, Cadot and Himbert (2018) measure the intensity in satellite night lights around the borders of 138 countries and find that on average, trade significantly dispersed economic activity along border crossing road corridors.

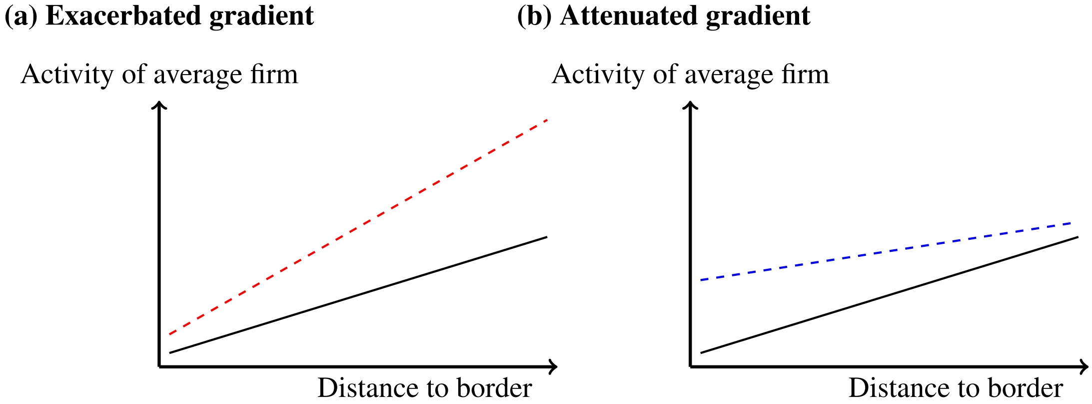
The analysis in this paper complements the empirical literature in general and the latter paper in particular by estimating the effect of trade facilitation on firm activity based on the location of a firm with respect to the nearest land border. By estimating elasticities with respect to changes in trade volumes within border regions, the paper quantifies the impact of trade on firms at different locations. While most of the literature measures aggregate effects of concentration and dispersion, the empirical analysis in this paper quantifies the effects of trade liberalization at different locations purely at the intensive margin of existing firms. The paper also adds evidence on the heterogeneity of reactions of different sectors to an increase in trade flows.

[6] See e.g. Allen and Arkolakis, 2014; Atkin and Donaldson, 2015; Cosar and Fajgelbaum, 2016; Fajgelbaum and Redding, 2014; Redding, 2016; Rossi-Hansberg, 2005

[7] See Autor, Dorn and Hanson (2013), Dix-Carneiro and Kovak (2017) or Caliendo, Dvorkin and Parro (2017).

[8] Some of the within-country studies surveyed find spatial divergence as a result of trade liberalization. This can however often be explained by idiosyncratic characteristics of the analysed country, such as Mexico where the US border region was already highly industrialized before trade liberalization.

Figure 3.1: Two scenarios of the impact of trade on firm activity



The solid lines illustrate the predicted activity for the average firm prior to trade liberalization. The dashed lines illustrate the predicted activity for the average firm after trade liberalization.

3.2.2 Heterogeneous effects of trade liberalization across sectors

While the literature reviewed above deals with the effect of trade on aggregate economic activity and at least implicitly assumes trade to have positive effects overall (with the discussion being which regions benefit to which extent from the gains of trade), a growing branch of literature also disentangles the effects of trade on a sectoral level.

In the theoretical literature, Caliendo, Dvorkin and Parro (2017) develop a dynamic trade model in which labour markets are spatially segmented and vary in their exposure to foreign trade. Frictions in the mobility of labour and goods in their model lead to largely heterogeneous responses of different regions to increased import competition. Taking their model to the data, they find that while the aggregate response of the U.S. economy to the increased trade with China, the effects vary across different regional labor markets. While employment in the manufacturing sector fell, workers were reallocated to other sectors benefiting from access to cheaper intermediate inputs such as construction and services. Dix-Carneiro (2014) develops a structural dynamic equilibrium model of the Brazilian labour market to model the impact of trade liberalization. In his model, an additional mobility friction in the form of sector-specific experience is assumed. This prolongs the transition phase in reaction to trade liberalization and causes heterogeneous outcomes across different regions based on initial sector of employment and on worker demographics such as age and education. The role of imperfect interregional mobility of workers and goods in generating heterogeneous effects of trade liberalization in different regions and sectors has also been highlighted by Artuc, Chaudhuri and McLaren (2010). Cosar, Guner and Tybout (2016) model the impact of globalization on heterogeneous firms and structurally estimate a model for worker transitions using Colombian firm data. They find that Colombia increased its national income at the cost of higher unemployment and larger wage disparities.

In the empirical literature, the impact of trade on the manufacturing sector has received special attention. Focusing on regional economies in a reduced form analysis, Autor, Dorn and Hanson (2013) find that the increased exposure of the U.S. particularly to imports from China increased unemployment and lowers wages in the U.S. manufacturing sector.[9] Dix-Carneiro and Kovak (2017) present similar results for local Brazilian labour markets.[10] They furthermore provide evidence for the negative effects on regions and sectors most

[9]Other papers confirming this effect on the U.S. manufacturing sector are e.g. Acemoglu et al., 2014; Pierce and Schott, 2016.
[10]Menezes-Filho and Muendler (2011) come to similar conclusions about the effect of Brazilian trade liberalization in the 1990s.

severely hit by cuts in tariffs and resulting import competition to increase over time. In their analysis, imperfect interregional labour mobility driven by slow regional capital adjustment and agglomeration economies counteract the impact of spatial arbitrage between regions as assumed in most spatial equilibrium models.[11]

This paper adds to the existing literature concerning the impact of trade on different sectors in the economy by disentangling the effects of increases in trade volumes on individual firms within border regions for different sectors of the economy. By doing so, it also sheds light on the question which sectors are predominantly contributing to patterns of in the spatial distribution of firm activity after trade facilitation, which could be used to better predict the impacts of trade on different regions based on their sectoral composition in the future. Taking into account the spatial dimension of firm location via individual firm addresses also provides within region variation that allows for effects to vary on a very fine spatial resolution instead of being limited to administrative level data.

3.3 Data

3.3.1 Data sources

The key data source for the empirical analysis in this paper is the Orbis firm database provided by Bureau van Dijk, which compiles data on private companies in over 90 countries worldwide.[12] The dataset has been widely used in the empirical literature due to several advantages such as the good coverage of small and medium sized private firms and the harmonized approach that allows cross country comparisons between firms. Data is collected from official business registers, annual reports, newswires, and webpages and contains information such as employment, turnovers, financial assets and detailed industry classification (4-digit SIC). Within the scope of this paper, this offers the opportunity to study the effects of trade facilitation on individual firm performance both on an aggregate level as well as on a sector level. Most importantly, the data mimics the official size distribution of firms as provided by Eurostat for at least most European countries, and covers more than 50 percent of the aggregate output in all countries and around 80-90 percent in most countries.[13]

Furthermore, at least in parts the data also records the address of a company's headquarter However, the coverage of headquarter addresses varies largely across countries. Figure 3.2 plots the share of firms that report their headquarter address in 2007 against their home countrys GDP per capita in 2007. The graph shows the large variation in the share of reported headquarter addresses in general and the link between GDP per capita and the reporting share, with richer countries tending to have a better coverage in terms of firm addresses. The share of reported addresses per country is important for two reasons: first, companies with missing information on their headquarter address are significantly smaller in terms of employment and have lower turnovers than companies that do report headquarter addresses.[14] Since the empirical analysis in this paper consists of regressing firm size and turnover on the distance to the closest land border (and therefore only firms with reported addresses enter the regression) countries with lower coverage of headquarter addresses would rely on a sample distorted towards bigger firms.

Secondly, the identification in this paper relies on within region-sector-year variation of distances to land borders. As shown in Figure 3.3, the share of companies that report their headquarter address is strongly

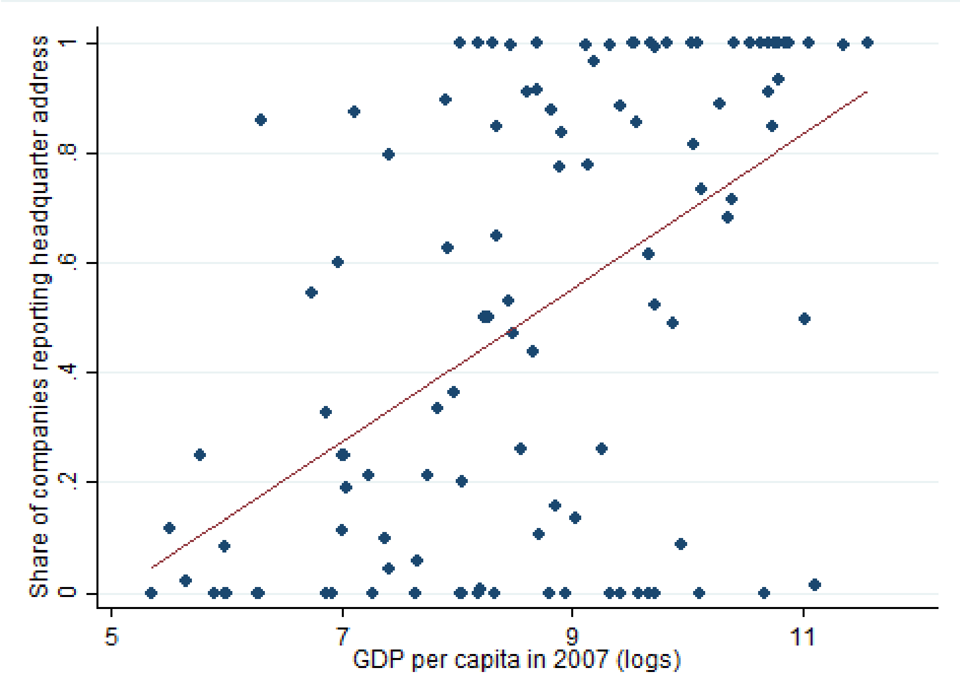
[11] See for example Blanchard and Katz, 1992; Bound and Holzer, 2000.

[12] For information about the potential obstacles and solutions in constructing a dataset from this database, see Kalemli-Ozcan, et al. (2015).

[13] See Kalemli-Ozcan et al. (2015).

[14] Firms that do not report the headquarter address are on average 22% smaller in terms of employment and report turnovers that are about 27% lower than firms with recorded headquarter addresses. This difference in means is statistically significant (P-value ¡ 0.01).

Figure 3.2: Share of firms reporting their headquarter address



Note: The y-axis plots the share of companies in the ORBIS dataset that report their head quarter address per country in 2007. The red line represents a linear prediction of the reported share as a function of the log GDP per capita in 2007.

correlated with the share of firms that report their address in a countrys capital. The worse the coverage of headquarter addresses, the higher the share of firms reported in the capital becomes. For the empirical analysis, high shares of firms reported in the capital pose a problem both because they decrease the variation in terms of distance to land borders needed to identify the main effects and because it increases the risk of disparities between reported firm addresses and the actual location of the firms main economic activity.[15] In order to minimise the prevalence of the problems described above, I restrict the sample to 24 European countries displayed in Figure 3.4. Countries are included in the sample based on two selection criteria: on one hand, at least 95% of the observed firms in 2007 report their headquarter address. On the other hand, less than 30% of the reported addresses are in the countrys capital.[16]

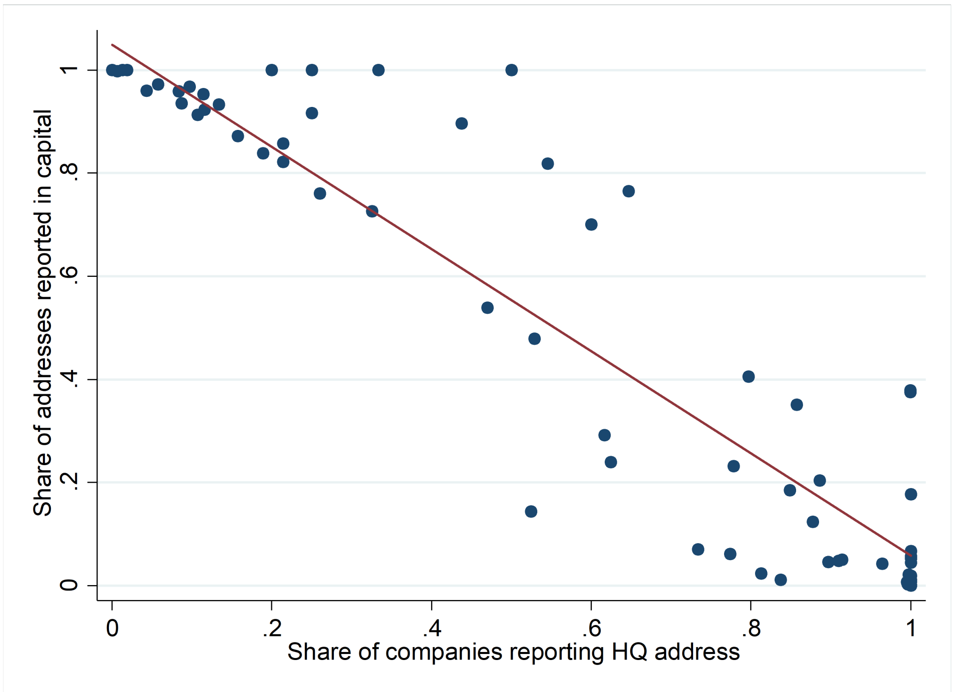
The Orbis dataset contains information on different categories of firms. Companies with a turnover of at least 100 mln EUR and at least $1,000$ employees are categorized as ”Very large companies”, companies with a turnover between 10 mln EUR and 100 mln EUR and a number of employees between 150 and $1,000$ are counted as ”Large companies”. Since these last two categories of companies are most likely to have their headquarter registered in high congestion cost urban areas while simultaneously having establishments in other locations and therefore pose the biggest risk of wrongly attributing economic activity to a location, the majority of estimations in this paper excludes them from the sample.[17] Most of the analysis in this paper focuses on small and medium sized companies. Companies are categorized as ”Medium sized” if their annual turnover is

[15] I.e. firms registering their headquarter in the capital in order to have easier access to the administration while their main activity takes place in a more remote location.

[16] Table 3.22 provides estimates for the baseline regressions with a sample of 39 countries. The selection criteria in this exercise were changed to $>=$ 60% share of reported addresses and $<=$ 45% share of addresses reported in the capital

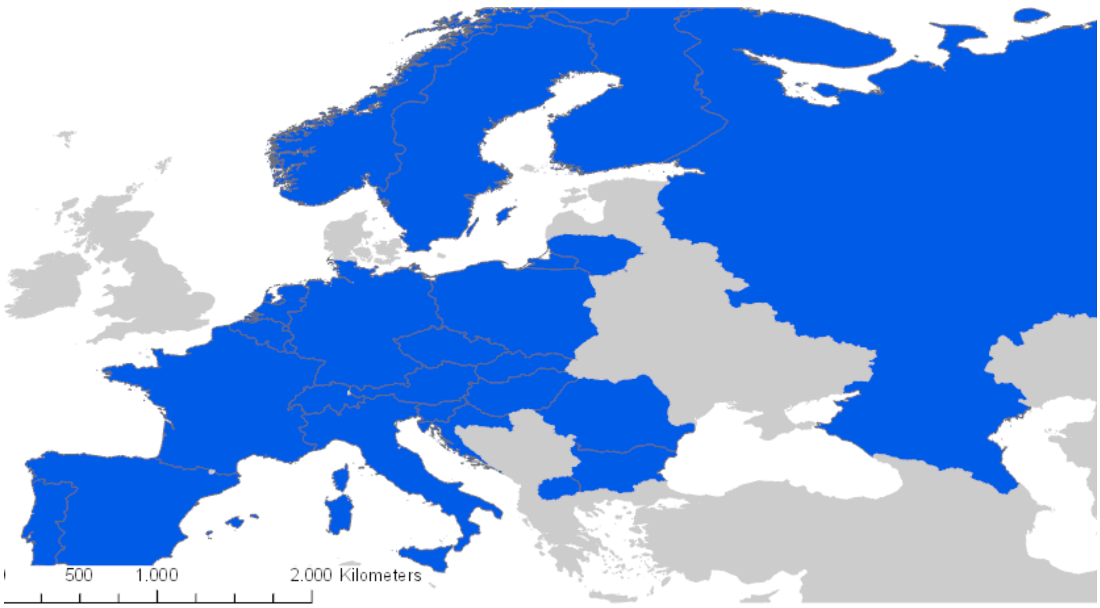
[17] Section 5.1 elaborates on this point and presents a robustness check allowing for large and very large companies.

Figure 3.3: Share of firms reporting their headquarter address



Note: The x-axis plots the share of companies in the ORBIS dataset that report their head quarter address per country in 2007. The y-axis plots the share of companies with a recorded head quarter address that report their address in the capital of a country. The red line represents a linear prediction of the relation between the two shares.

Figure 3.4: Sample countries



Note: Sample countries are represented in blue. Countries were selected based on availability of firm addresses for the entire sample period and the diversity of zipcodes in which firms were registered.

between 1 mln EUR and 10 mln EUR and have between 15 and 150 employees. Furthermore, they must not be related to a company categorized as ”Large” or ”Very large”. ”Small” enterprises are defined as any company with less than 15 employees and less than 1 mln EUR of annual turnover. Furthermore, the dataset is restricted to active firms that were not subject to mergers or acquisitions during the observation period.[18]

For the empirical analysis, two outcome variables are of interest: turnover and employment of individual firms. Turnover provides a proxy to measure gross economic activity, thus linking the paper to the majority of the theoretical literature studying the spatial effects of trade, and is measured in mln EUR. Employment as an outcome variable allows the paper to contribute empirical evidence on the relationship between trade and employment and is measured as the total number of employees in a given year, not accounting for differences in hours worked or full-time equivalents.[19][20]

The key goal in this paper is to document spatial gradients with respect to distance to the nearest land border for these two outcome variables. Table 3.1 presents summary statistics for turnover and employment within different distances to the nearest land border. Within the border regions informing the empirical analysis in this paper (i.e. up to 200 kilometres of distance), the interior band of $0 - 100$ kilometres of distance displays both lower turnovers as well as lower employment than the exterior band of $100 - 200$ kilometres of distance, with the differences between these two bands being statistically significant. This pattern of firms increasing in size as one moves away from the border does not only persist within border regions, but continues when comparing firms in border regions to firms located further than 200 kilometres away from the nearest land border. Again, firms level activity increases with distance from border, with the distances being statistically significant.

Table 3.1: Turnover and employment by distance to border

	Turnover (mln EUR)			Number of employees		
	Mean	Std. dev.	p-value	Mean	Std. dev.	p-value
$0 - 100$ km distance to border	1.27	2.06		11.19	38.71	
$100 - 200$ km distance to border	1.41	1.94		12.99	32.78	
Difference	-0.14		< 0.001	-1.80		< 0.001
$0 - 200$ km distance to border	1.34	1.95		11.93	36.42	
> 200 km distance to border	1.52	2.37		14.55	39.60	
Difference	-0.18		< 0.001	-2.62		< 0.001

Note: Statistics based on firms categorised as ”Small” or ”Medium Sized” in the Orbis database. All statistics for up to 200 km of distance from the border are based on firms of the baseline estimation sample. The statistics for $>$ 200 km of distance are based on all firms that can be geo-referenced in the observation period and countries. However, these firms do not inform any estimation presented in the paper and are presented only for illustrative purposes in this table.

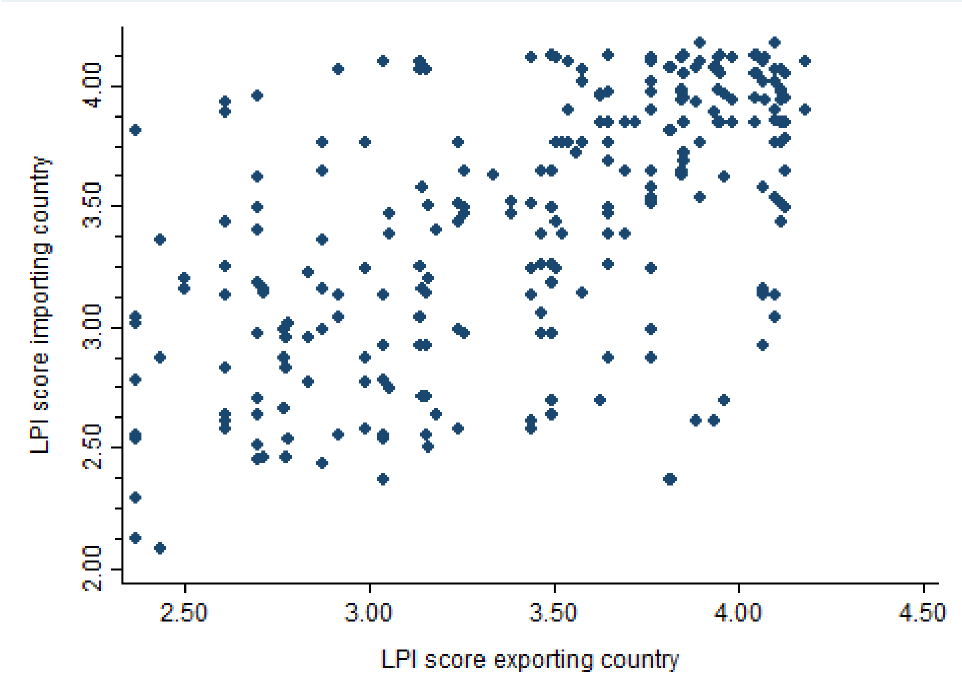
Trade is measured through bilateral export values between neighboring countries from the United Nations' UN Comtrade database. Information about trade facilitation is gathered from the category ”Customs” of the World Bank's *Logistics Performance Index (LPI)*, an index assessing the ability of countries to conduct international trade. Figure 3.5 shows the cross country variation of LPI scores at the borders considered in the estimation sample. Within countries, the average growth rate from 2007 to 2014 of the LPI scores in the sample is 0.23, showing a trend of improvements in trade facilitation for the considered sample period.

[18] Bureau van Dijk provides a separate product tracking all changes in ownership (ZEPHYR) that can be combined with the Orbis data. See Kalemli-zcan et al. (2015).

[19] Papers like Helpman, Itskhoki and Redding (2010) note the absence of a consensus on the link between trade and unemployment.

[20] Table 3.15 provides summary statistics. 1,432 companies with negative turnover in any year are excluded from the regression. Of the sample companies, only 0.4% of the observations record 0 turnover.

Figure 3.5: LPI scores for borders in sample



Georeferenced data on the location of national and state borders as well as the location of urban areas are taken from the Database of Global Administrative Areas hosted by the Hijmans Lab at UC Davis. Data on roads are obtained from the Global Roads Open Access Data Set (gROADS), version 1 (1980-2010). The location of ports is taken from the World Port Index published by the US National Geospatial-Intelligence Agencys Maritime Safety Office, and the location of airports is taken from Natural Earth and the Emergency and Preparedness Geospatial Information Unit at the World Food Programme (WFP).[21] Information on altitude is available through the Scripps Institution of Oceanography at the University of California Sand Diego, whose SRTM30 Plus dataset combines sea floor and land elevation data for the entire planet.

3.3.2 Construction of the dataset

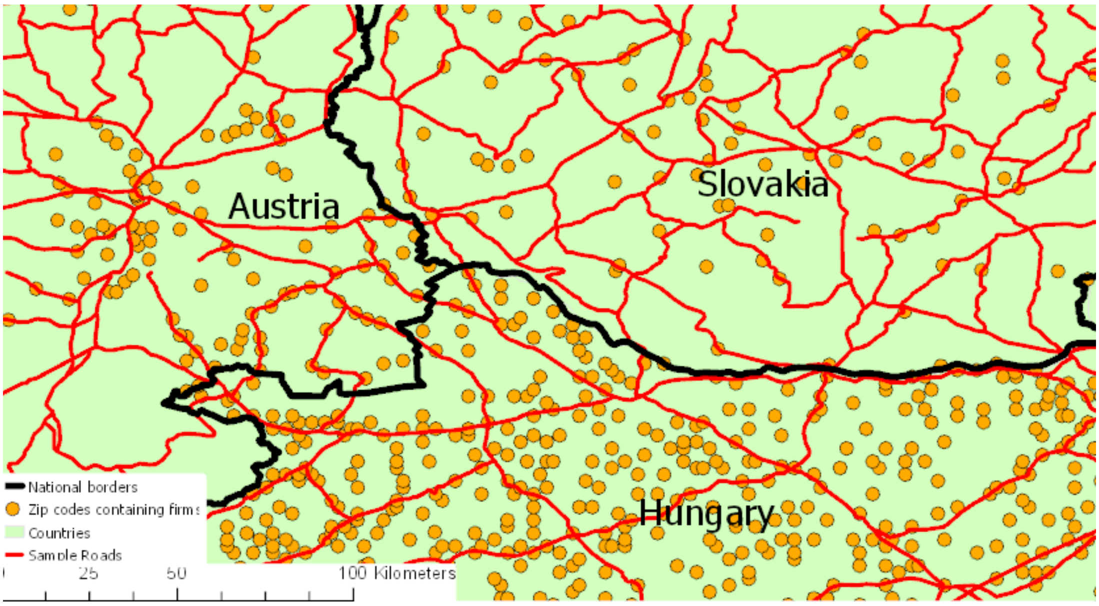
In order to be able to analyse the impact of trade on the spatial distribution of economic activity at the firm level, I construct a dataset of firms within the border regions of 24 countries.[22] In order to do this, the firm headquarter addresses in the Orbis dataset are used to georeference the location of firms at the zip code level using GIS software.

The key determinant of geographic location in this paper is the road distance of a firm to the nearest land border. Importantly, within this paper only land borders are considered in the analysis while coastal areas are not. This is done for two reasons: on one hand, coastal regions are often more developed than the rest of the country due to historical advantages arising from easy access to international markets and are not impeded in their development due to remoteness in the way land borders tend to be (which often times run through inhospitable regions such as mountain ranges or deserts). From a development perspective, it is therefore more relevant to study potential mechanisms that can help previously underdeveloped regions around land borders

[21] Only airports that appear in both datasets are kept.

[22] Throughout this paper, the border region is defined as the corridor of 200 km to each side of the border, following the definition in Brülhart, Cadot and Himbert (2018). Table 3.17 also provides estimations for the baseline results with different cut-off distances, showing that the obviously arbitrary choice of the size of the border region does not affect the results in a meaningful way.

Figure 3.6: Roads, borders and firms



Note: Zip codes containing firms are displayed as yellow dots. National borders are displayed in black, roads in red. In order to inform the regressions, a zipcode must be within 20 km of the next road and within 200 km road distance from the nearest land border.

to catch up. Furthermore, focussing on land borders offers the technical advantage of providing a clearer identification of the effects of trade on a bilateral basis. This bilateral identification would be impossible to argue for coastal areas, where any port can potentially be used to conduct trade with any other port in the world. Within this paper, border regions are therefore defined as 200 km buffers to either side of a land border between two countries. To construct the dataset, I first construct a network of border crossing roads in the sample countries, which contains all road corridors that cross any land border to a neighbouring country, following each road up to 200 km inwards into the country. In the next step, zip codes are located at these border crossing roads by spatially joining them with a matching distance of at most 20 km. From the location of a zip code, I then compute the shortest road distance to the nearest land border. Furthermore, I compute the distances to the closest airports and sea ports and the altitude for each zip code in order to be able to control for locations that benefit both from access to the international market across the border as well as access to world markets through ports and airports.

In total, this yields a dataset of 1,346,472 individual firms in 24 countries for the years 2007, 2010 and 2014. The firms are distributed over 103,117 distinct zipcodes in 45,094 cities, showing that the pooling of firms at the zip code level still leaves a very high spatial resolution of observations with a high degree of variation in distance to the nearest border even within sub-national administrative regions. Figure 3.6 illustrates this using the border region between Austria, Hungary and Slovakia. While the degree of spatial diversity varies across countries depending on administrative (i.e. geographic size of zipcodes) as well as economic idiosyncrasies (i.e. share of zip codes that contains firms recorded in the Orbis dataset), all observed countries provide firms location that vary on a fine spatial resolution. Concerning trade, I observe 88 borders, providing ample variation in the degree of trade facilitation and changes in trade over the years.

3.4 Estimation

3.4.1 Baseline estimation: road and country-sector-year fixed effects

The baseline empirical strategy entails estimating the elasticity of economic activity of firms depending on their distance to the closest land border.

As measures of firm level activity, the analysis focusses on employment and turnover of individual firms in the considered border regions. Let $y^b_{ijrscc't} = y^b_{it}$ be the activity of firm i specialized in sector j located on road r in sub-national region s leading from country c to country c' in year t. The superscript b denotes the different outcome variables and can be equal to either turnover or employment[23]. The subscripts r, s, c and c' are implied by i, as every firm is uniquely assigned to a country, region and nearest road. d_i denotes firm is distance to the closest land border along road r, while $T_{cc't}$ signifies the volume of exports from country c to neighbouring country c' across that border.

The baseline estimation equation is then as follows:

$$\ln y^b_{it} = \alpha + \beta_1 d_i + \beta_2 \ln T_{cc't} + \beta_3 (d_i \times \ln T_{cc't}) + \theta \mathbf{x}_i + \gamma_r + \gamma_{cjt} + v_{irt}, \quad (3.1)$$

where $\mathbf{x}_i$ is a vector of firm level controls that includes log altitude, dummy variables for whether firm i is located more closely to a sea port or airport than to the closest land border, interactions of those two dummies with the road distance to the port or airport in question, and a dummy for whether the shortest route to the next border crossing runs predominantly along roads other than highways and the interaction of this dummy with the road distance to the border.[24] γ_r is a road fixed effect, and γ_{cjt} is a country-sector-year fixed effect.

In order to facilitate the interpretation of coefficients, distance is entered linearly in units of 10 kilometres, while trade is measured in logarithms in order to interpret the estimated coefficients for $\hat{\beta}_2$ as elasticities.[25]

Country-sector-year fixed effects provide a very sharp channel of identification. Not only do they sum up to country, sector and year fixed effects respectively but also to country-year fixed effects, filtering out any common time-varying characteristics for firms in a given country that are subject to national legislation. Additionally, they also sum up to sector-year fixed effects, controlling for common time-varying determinants for firms in any country, thus filtering out technological progress. Furthermore country-sector-year fixed effects are also able to control for sectoral business cycle movements in every country. Standard errors are twoway-clustered at the road and the country-pair-year level.

3.4.2 Identification and instrumentation

The set of fixed effects used in equation 3.1 implies that the identification of the effects of borders and trade on firm level activity runs within-road. The coefficients of main interest are β_1, β_2 and β_3. Whenever the estimated coefficient β_1 is significantly positive firm level activity increases towards the center of country c in a theoretical scenario of a completely closed border (i.e. at zero trade), i.e. the the turnover and employment of otherwise identical firms along a road varies according to their distance to the nearest border.

[23] Employment is measured as the number of employees, turnover in million .

[24] The minor roads dummy is constructed by computing the share of the shortest route to the nearest land border that runs along highways. If this share is below 50%, the road is classified as ”minor road”.

[25] Technically, the ”gradient” of firm level activity and employment along distance are therefore actual semi-elasticities.

The overall effect of increased exports from country c to country c' on activity of firms at any location in the border region is measured by β_2 and expected to be significantly positive. The interaction term between distance and trade (β_3) allows for the effect of trade on firm activity to vary over distance and to affect the initial gradient. Whenever β_3 is of the same sign as β_1, it provides evidence for attenuated gradient in which locations that already hosted bigger and more active firms before trade also reap the biggest benefits of trade facilitation. In the contrary case where β_3 has the opposite sign of β_1 previously smaller and less active firms grow faster due to an increase in exports.

A possible threat to establishing a causal link between changes in trade volumes and the activity of firms within the border region is the potential endogeneity of trade to firm activity in the border region. While trade can clearly be expected to affect the activity of firms, changes in employment decisions and turnovers of border region firms can as well have an impact on the volume of trade between two neighbouring countries. In order to address this issue, equation (3.1) is therefore estimated using an instrumental variable approach where bilateral exports $T_{cc't}$ are instrumented with the product of the scores of the two countries c and c' in the "Customs" dimension of the World Bank's *Logistics Performance Index (LPI)* in year t. This acknowledges the fact that exports depend both on the efficiency of the export customs clearance process of the exporting country as well as on the import customs clearance process of the importing country. Using the product of the scores of two neighbouring countries provides variation at the country-pair-year level and is therefore not absorbed by the country-sector-year fixed effects. Furthermore, constructing the instrument as the product of the two scores has the advantage of penalizing country-pairs in which one country has a much lower score (which taking the sum would not) without making the extreme assumption of perfect complementarity between the two countries (which taking the minimum would imply).[26]

Two features of this approach should be noted: Firstly, the instrument is constructed using only the scores of two neighbouring countries in the category "Customs" of the LPI. This category measures the efficiency of the customs clearance process by border control agencies (i.e., speed, simplicity and predictability of formalities). This provides a good measure of trade facilitation and avoids possible endogeneity between firm level activity and factors like technology and infrastructure in the border region which could have been present if the overall score in all categories of the LPI had been used as instrument.[27] Following from this, it is plausible that the imposed exclusion restriction holds, which requires the LPI score in customs clearance of countries c and c' to affect firm activity in country c only through changes in the volume of exports from country c to c'. The identifying assumption of the instrument is that employment and turnovers of firm i do not excel a direct impact on the efficiency of the customs clearance process. Especially given the fact that all baseline estimations in this paper exclude large and very large companies (which might be argued to have sufficient bargaining power to influence public spending), this assumption seems plausible as well.

Secondly, it needs to be noted that the LPI gives country wide scores that are not specific to particular border posts. Combining the scores of the two countries straddling a border therefore gives a unique score to each border, but does not allow for within border variation of the instrument. In the same way, the trade variable is measured by the overall volume of exports from country c to country c', but does not permit to further distinguish how much exports crossed a particular border posts. Both the instrument as well as the trade variable vary on the country-pair-year level.

[26]It should also be noted that running the baseline estimations with exports instrumented by the sum or the minimum of the two scores has no discernible impact on the results, while providing weaker instruments and lower F statistics.

[27]Other categories measured in the LPI include for example the quality of trade and transport related infrastructure (e.g., ports, railroads, roads, information technology) or the competence and quality of logistics services (e.g., transport operators, customs brokers), which could be argued to be endogenous to firm activity in the border region.

3.4.3 Alternative estimation: region-sector-year fixed effects

Another potential threat to the identification strategy implied by equation (3.1) consists in unobserved political or other regional events that would affect firm level activity in the border region over time as well as national trade and trade policy. This threat could be prevalent in cases where well-connected local politicians in border areas can increase public funding for their regions (for roads, electrification etc.) and at the same time use their political influence to reform administrative procedures and improve efficiency of border posts in their constituencies (see Hodler and Raschky, 2014).

In order to filter out such cases, the fixed effects estimation approach of equation (3.1) is altered to control for region-sector-year fixed effects. Again, these fixed effects sum up to region-year fixed effects that control for general time-varying developments of public spending or legislation at a regional level. Furthermore, the sectoral component of the fixed effects even allows to absorb time-varying political or infrastructural changes put forward by interest groups of certain sectors.

The alternative estimation equation then looks as follows:

$$\ln y_{it}^{b} = \alpha + \beta_1 d_i + \beta_2 \ln T_{cc't} + \beta_3 (d_i \times \ln T_{cc't}) + \theta \mathbf{x}_i + \gamma_{jst} + \gamma_{c't} + v_{irt}, \tag{3.2}$$

where γ_{jst} is a vector of time-variant sub-national region fixed effects specific to sector j and $\gamma_{c't}$ is a vector of neighbour-country fixed effects, allowing for effects originating in the foreign market.[28] Since identification of β_2 and β_3 now relies on within region-year variation, it is driven by regions from which at least two different neighbour countries can be reached within 200 km. This is the case for 48.7% of the regions in the baseline sample.

3.5 Baseline results

3.5.1 Exports

Table 3.2 presents the baseline results when using either firm turnover (columns (1) to (4)) or employment (columns (5) to (8)) as outcome variable, with firms of all categories in the Orbis data included in the estimation sample. Among the specifications using turnover as outcome variable, columns (1) and (2) estimate equation (3.1) using fixed effects at the road and the country-sector-year level while columns (3) and (4) rely on the region-sector-year and neighbour country-year fixed effects as presented in equation (3.2). In both cases, estimates are presented both as OLS as well as 2SLS instrumenting for bilateral exports. Columns (5) to (8) are organized in the same fashion for employment as outcome variable.

Among the coefficients of interest $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$, results are in line with the theoretical scenario of trade both increasing economic activity within the border region and attenuating the gradient of firm activity along distance from border. $\hat{\beta}_1$ is consistently positive and significant at the 1% confidence level across all specifications. I find that firm turnover is increased by about 14% for every ten additional kilometres to the border. For employment, the same result is found qualitatively, yet the gradient along distance is much flatter, ranging from 5.9% to 6.9% for every ten kilometres of distance to the border when instrumenting trade.

Exogenous changes in bilateral exports caused by changes in customs clearance efficiency exert a positive effect on firms turnover and employment, with $\hat{\beta}_2$ being consistently positive and statistically significant. Again

[28] Regions are defined at the highest sub-national administrative level, e.g. at the state level in the United States (hence the notation s).

the coefficients tend to be bigger in magnitude on turnover than on employment. Even more importantly, exogenous changes in trade significantly attenuate the gradients of firm size and turnover along distance from the nearest border, with firms located closer to the border generating higher increases in turnover and employment than firms located towards the center of the country. Again, the coefficients are slightly bigger in magnitude for turnover than for employment. Figure 3.7a and Figure 3.7b put these coefficients into context with the initial gradients of turnover and employment along distance. Here we see that the attenuating ~~dispersing~~ impact of trade is relatively more important for turnover than for employment.[29] An increase in trade from the value of the 25th percentile to the median value (which is equivalent to increasing Austria's exports to Hungary to the level of exports Austria has towards Italy) generates a 46% increase in turnover at the border crossing point but only a 13% increase at 200 kilometres distance from the border. Furthermore, the difference between the two gradients for turnover stops being statistically significant from 160 kilometres of distance from the border onwards. Compared to that, the same increase in trade generates a 50% increase in employment at the border crossing point, which is almost identical to the effect of trade on turnover directly at the border. Contrary to the case of turnover, the difference between the two gradients remains statistically significant for the entire sample, with an increase in employment of 26% at 200 kilometres distance from the border (i.e. twice the size as the increase in turnover at the same distance).

The main difference between the within road and the within region-sector-year results occurs concerning $\hat{\beta}_2$, with the coefficients being significantly smaller when estimated within region-sector-year. This can be intuitively explained due to the potential spillovers within a region: while an increase in trade over time cannot affect trade in the earlier period, the identification with region-sector-years is less sharp. Within a region, gains of an increase in trade with one neighbour country c' can be incorporated partly by firms located within the same region that are located closer to neighbour country c'', which biases $\hat{\beta}_2$ downwards.

Control variables affect firms according to intuition, with firms in higher altitudes generating less turnover and employment while closeness to ports and airports helps firms to generate higher turnovers and employ more people.[30] Along minor roads, employment and turnovers are significantly lower and the gradient of turnover along distance is minimally steeper.[31]

Instrumenting generally does not affect the qualitative findings, and the product of the two LPI scores of the two countries straddling the border provides a strong instrument with first stage F statistics being consistently between 20 and 30. Table 3.18 presents first stage results.

[29] For the remainder of the paper, "trade" always refers to the predicted exogenous changes in bilateral trade caused by changes in customs clearance efficiency.

[30] It should be noted that aerial and maritime trade potentially bias the estimated coefficients concerning β_2 and β_3 towards zero, since the effect of an increase in exports to the neighbor country would not be measured through changes in light intensities along border crossing roads if trade is carried out by other means of transport.

[31] Since the underlying road network is time invariant, a road is either classified as major or as minor road in all observed years. For this reason, the coefficient on the pure effect of minor roads drops out in the specification using road fixed effects, since there is no within-road variation in this variable. Furthermore, it should be noted that this potentially biases the estimated coefficients concerning β_2 and β_3 towards zero. This is due to the fact that in the case of new roads being constructed in the vicinity of our sample roads, trade would be likely to flow along new and more modern roads, decreasing the potential effect measured at the initial sample roads.

Table 3.2: Baseline estimates - All firms

Dependent variable:	Turnover (logs)				Employment (logs)			
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
				Effects of distance and trade				
Distance to border (in 10km)	0.115***	0.143***	0.100***	0.142***	0.040***	0.059***	0.039***	0.069***
	(0.008)	(0.020)	(0.011)	(0.030)	(0.007)	(0.013)	(0.008)	(0.013)
Bilateral exports (in logs)	0.201***	0.442***	0.080***	0.138*	0.135***	0.463***	0.042***	0.803***
	(0.065)	(0.086)	(0.040)	(0.074)	(0.041)	(0.097)	(0.024)	(0.029)
Bilateral exports × Distance to border	-0.011*	-0.014*	-0.011***	-0.020*	-0.011***	-0.010**	-0.010***	-0.013***
	(0.006)	(0.008)	(0.003)	(0.013)	(0.003)	(0.005)	(0.003)	(0.004)
				Control Variables				
Minor road (dummy)	0.000	0.000	-0.100*	-0.025*	0.000	0.000	-0.035**	-0.020**
	(.)	(.)	(0.055)	(0.014)	(.)	(.)	(0.019)	(0.010)
Minor road × Distance to border	0.005	0.008	0.001*	0.003*	0.003	0.005	0.002	0.001
	(0.009)	(0.009)	(0.001)	(0.002)	(0.006)	(0.006)	(0.004)	(0.007)
Altitude (in 100m)	-0.048***	-0.033*	-0.018*	-0.026*	-0.005**	-0.009**	-0.010**	-0.010**
	(0.018)	(0.018)	(0.010)	(0.015)	(0.002)	(0.004)	(0.005)	(0.005)
Port closer than next land border (dummy)	0.183**	0.159**	0.263**	0.265**	0.062**	0.073**	0.144**	0.195**
	(0.077)	(0.065)	(0.124)	(0.119)	(0.028)	(0.035)	(0.072)	(0.078)
Port dummy × Distance to port	-0.010	-0.009	-0.035**	-0.037**	-0.006	-0.005	-0.016	-0.016
	(0.007)	(0.008)	(0.017)	(0.015)	(0.005)	(0.005)	(0.012)	(0.011)
Airport closer than next land border (dummy)	0.021***	0.038***	0.145*	0.209*	0.078***	0.091***	0.018***	0.036***
	(0.005)	(0.013)	(0.074)	(0.111)	(0.030)	(0.040)	(0.005)	(0.013)
Airport dummy × Distance to airport	-0.010	-0.009	-0.009	-0.015	-0.006**	-0.005**	-0.000	-0.002
	(0.006)	(0.006)	(0.010)	(0.012)	(0.003)	(0.002)	(0.004)	(0.004)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓	✓			✓	✓		
Country-Sector-Year FE	✓	✓			✓	✓		
Region-Sector-Year FE			✓	✓			✓	✓
Neighbor country-Year FE			✓	✓			✓	✓
Kleibergen-Paap F statistic		31		22		29		24
# Clusters	249	249	239	239	249	249	239	239
# Observations	2,410,178	2,410,178	2,399,226	2,399,226	2,410,178	2,410,178	2,399,226	2,399,226

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Employment is measured as the number of employees, turnover in million .

Table 3.2 relies on a sample including both small and medium sized enterprises as well as large and very large firms from the Orbis dataset. Since the Orbis dataset only identifies headquarter and not establishment addresses, including large and very large firms potentially introduces noise to the estimations since those firms have the highest risk of disparities between establishment and headquarter locations due to their capacity to entertain establishments at different locations while affording their headquarter to be located in high-congestion-cost areas. Table 3.3 reproduces the baseline estimations in a sample limited to only small and medium sized enterprises. All results remain virtually unaffected, which is partly due to the low percentage of large and very large firms in the dataset. For the sake of easier comparison, Table 3.20 in the appendix provides results when reproducing the results of Table 3.3 using a sample of only large and very large firms within 200 km distance from land borders. Results are qualitatively similar to the results for SMEs, with firm level activity increasing with distance to the border and trade both increasing activity within the border region and attenuating the gradients of firm size and turnover along distance. However, the estimated coefficient for $\hat{\beta}_3$ is smaller in size than for the baseline sample of SMEs. Together with the slightly steeper gradient indicated by the estimated coefficient of $\hat{\beta}_1$ being bigger in size than the gradient for SMEs this could go in line with the assumption that larger firms tend to have headquarters in central locations (hence the bigger $\hat{\beta}_1$) and establishments closer to the border. In this case, the Orbis dataset would record increases in employment that occur in establishments at the border at the headquarter address, which would bias the coefficient of $\hat{\beta}_3$ downwards. This exercise shows that the problem of the Orbis dataset not recording establishment addresses can at least be partly addressed by excluding large and very large firms from the analysis, and that a potential bias arising from the discrepancy in headquarter addresses and actual economic activity would be downwards, hence diminishing the estimated coefficient for $\hat{\beta}_3$. For this reason, all following estimations are conducted on the truncated sample of only SMEs.

The fact that the results for large and very large firms are qualitatively identical to those of small and medium sized firms and the good coverage of total output provided by the firms in the Orbis dataset in European countries (see section 3) suggests that the firm level results obtained in this paper also give a good indication of the spatial effects of trade on an aggregate level.

Table 3.3: Baseline estimates - Excluding large and very large firms

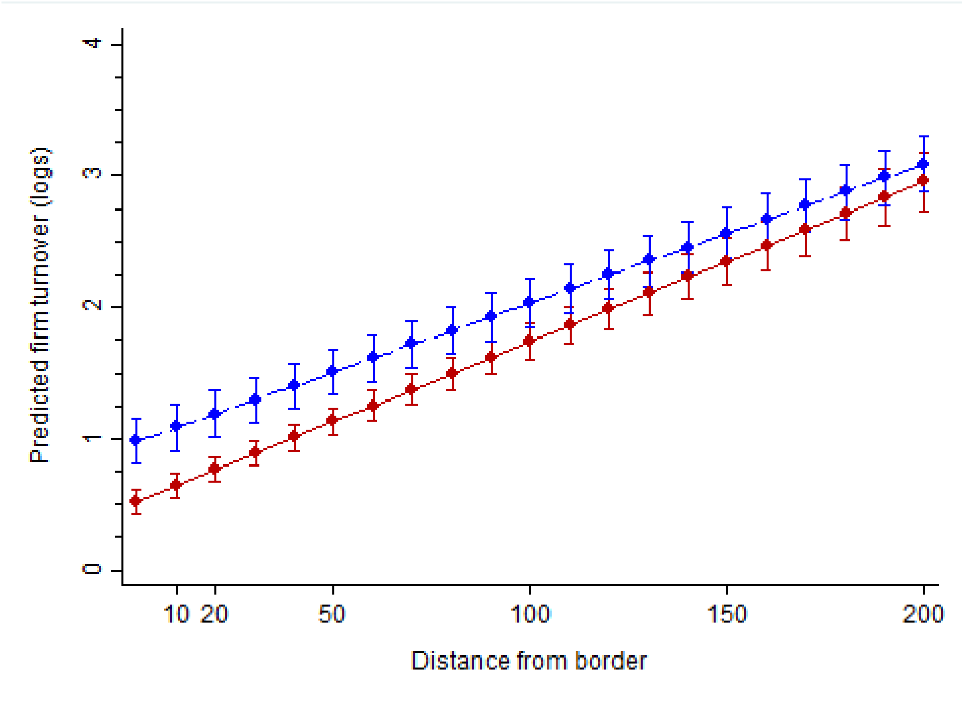
Dependent variable:	Turnover (logs)				Employment (logs)			
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
				Effects of distance and trade				
Distance to border (in 10km)	0.115***	0.140***	0.102***	0.144***	0.039***	0.058***	0.036***	0.067***
	(0.008)	(0.021)	(0.011)	(0.030)	(0.007)	(0.013)	(0.008)	(0.013)
Bilateral exports (in logs)	0.203***	0.460***	0.086***	0.140*	0.135***	0.491***	0.040***	0.079***
	(0.065)	(0.091)	(0.026)	(0.072)	(0.040)	(0.104)	(0.013)	(0.026)
Bilateral exports × Distance to border	-0.013*	-0.016*	-0.012***	-0.020*	-0.012***	-0.012**	-0.010***	-0.015***
	(0.007)	(0.009)	(0.004)	(0.012)	(0.004)	(0.005)	(0.002)	(0.005)
				Control Variables				
Minor road (dummy)	0.000	0.000	-0.099*	-0.025*	0.000	0.000	-0.035**	-0.019**
	(.)	(.)	(0.052)	(0.014)	(.)	(.)	(0.017)	(0.009)
Minor road × Distance to border	0.006	0.010	0.001*	0.003*	0.003	0.006	0.002	0.001
	(0.009)	(0.009)	(0.001)	(0.002)	(0.006)	(0.006)	(0.004)	(0.007)
Altitude (in 100m)	-0.048***	-0.037**	-0.021*	-0.026*	-0.004**	-0.009**	-0.007**	-0.010**
	(0.018)	(0.018)	(0.012)	(0.015)	(0.002)	(0.004)	(0.003)	(0.005)
Port closer than next land border (dummy)	0.182**	0.159**	0.261**	0.264**	0.062**	0.073**	0.145**	0.195**
	(0.077)	(0.065)	(0.123)	(0.118)	(0.028)	(0.035)	(0.072)	(0.078)
Port dummy × Distance to port	-0.010	-0.008	-0.034**	-0.037**	-0.006	-0.005	-0.015	-0.015
	(0.007)	(0.007)	(0.015)	(0.015)	(0.005)	(0.005)	(0.010)	(0.010)
Airport closer than next land border (dummy)	0.021***	0.036***	0.141*	0.209*	0.078***	0.091***	0.018***	0.036***
	(0.005)	(0.011)	(0.073)	(0.111)	(0.030)	(0.040)	(0.005)	(0.013)
Airport dummy × Distance to airport	-0.010	-0.009	-0.009	-0.015	-0.006**	-0.005**	-0.000	-0.002
	(0.006)	(0.006)	(0.010)	(0.012)	(0.003)	(0.002)	(0.004)	(0.004)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓	✓			✓	✓		
Country-Sector-Year FE	✓	✓			✓	✓		
Region-Sector-Year FE			✓	✓			✓	✓
Neighbor country-Year FE			✓	✓			✓	✓
Kleibergen-Paap F statistic		31		22		29		24
# Clusters	249	249	239	239	249	249	239	239
# Observations	2,298,459	2,298,459	2,288,788	2,288,788	2,298,459	2,298,459	2,288,788	2,288,788

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

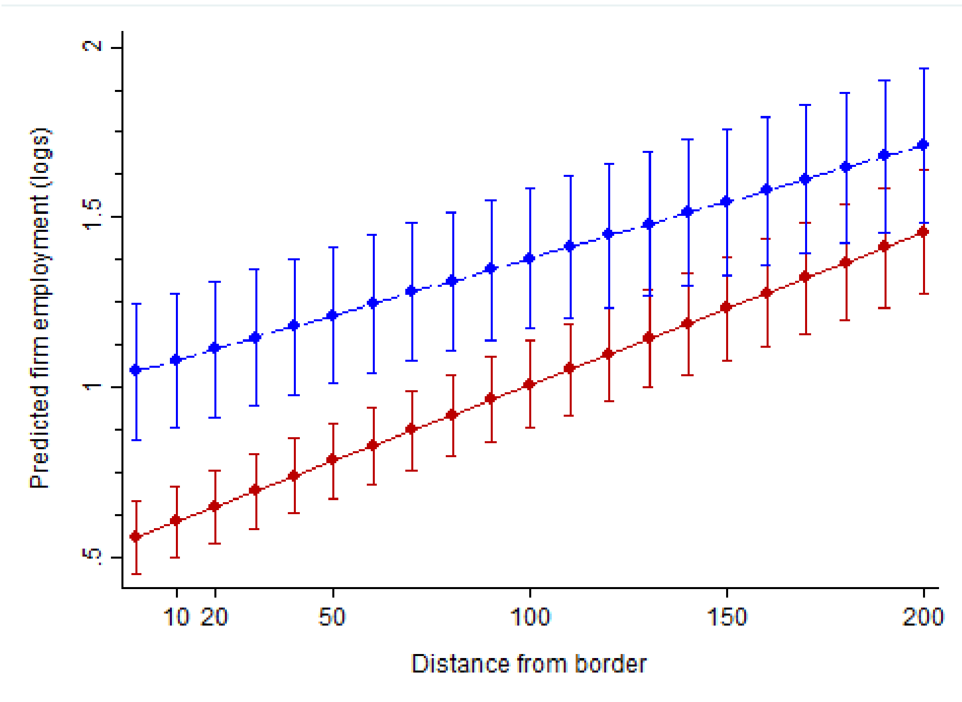
Two-way clustered standard errors at road and country-pair-year level in parentheses.

Employment is measured as the number of employees, turnover in million .

Figure 3.7: Graphical representation of baseline results



(a) Turnover



(b) Employment

Note: The figures visualize predicted turnover and employment based on baseline specifications featuring road fixed effects, country-sector-year fixed effects, all control variables and exports instrumented with LPI scores (3.3 columns (2) and (6)). Red lines represent simulations with trade set to the value of the 25th percentile of the baseline sample, blue lines set trade to the median value of the baseline sample.

3.5.2 Imports

After having measured the impact of exports to the neighbour country on firm level activity in border regions so far, the analysis now investigates the potentially asymmetric nature of trade facilitation. In order to do so, trade intensity is measured as the volume of imports from country c' into country c while still being instrumented with the product of LPI scores of the two countries in Table 3.4.

We observe a nearly identical pattern of employment and turnover of the average firm increasing with the distance to the border as in Table 3.3. However, the impact of trade on turnover and employment on firms at any distance from the border is significantly lower. The estimated coefficients for $\hat{\beta}_2$ are on average only between one third and half the size that they were for the case of exports to the neighbor country. All coefficients are however positive.[32] The coefficients are also less precisely measured and in the specifications with road and country-sector-year fixed effects not statistically significant when instrumented.

Furthermore, the effects are less dependent on geographical proximity to the border, as coefficients for $\hat{\beta}_3$ are much smaller in size. Overall, imports change the spatial pattern of firm activity by much less than exports and generate lesser gains for the affected border region overall.

The seemingly surprising result of imports increasing both employment and turnover of firms in border regions on average might be due to the predominance of service sector firms in the analysed sample. For this reason, section 6.1.2 analyses the effect of increased imports on different sectors.

Table 3.4: Baseline estimates with imports as measure of trade

Dependent variable:	Turnover (logs)				Employment (logs)			
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Distance to border (in 10km)	0.101***	0.126***	0.096***	0.112***	0.027***	0.057***	0.030***	0.052***
	(0.006)	(0.018)	(0.008)	(0.017)	(0.006)	(0.018)	(0.006)	(0.017)
Bilateral imports (in logs)	0.111**	0.253	0.025	0.058**	0.025**	0.076	0.027**	0.036***
	(0.050)	(0.183)	(0.012)	(0.028)	(0.042)	(0.062)	(0.013)	(0.016)
Bilateral imports × Distance to border	-0.006	-0.009*	-0.001*	-0.003*	0.002	-0.004**	0.003	-0.001*
	(0.005)	(0.005)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓	✓			✓	✓		
Country-Sector-Year FE	✓	✓			✓	✓		
Region-Sector-Year FE			✓	✓			✓	✓
Neighbor country-Year FE			✓	✓			✓	✓
Kleibergen-Paap F statistic		31		22		29		24
# Clusters	249	249	239	239	249	249	239	239
# Observations	2,298,459	2,298,459	2,288,788	2,288,788	2,2984,59	2,298,459	2,288,788	2,288,788

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Two-way clustered standard errors at road and country-pair-year level in parentheses.
Employment is measured as the number of employees, turnover in million .

3.6 Heterogeneous effects and mechanism

3.6.1 Effects by sectors

Exports

So far, the analysis has focused on estimating gradients of firm level activity along distance to the nearest land border on average for all sectors of the economy. Making further use of the heterogeneity of firms in

[32]This finding is in part driven by the correlation between imports and exports since exports are not controlled for in this specification. Table 3.9 addresses this issue.

the dataset, Table 3.5 presents the results when estimating equations (3.1) and (3.2) separately for the four biggest sectors in the data using turnover as the outcome variable. This adds nuance to the empirical analysis by allowing the investigation of three additional questions: first, it permits an analysis of the link of between sectoral composition and the steepness of the gradient and by this to identify which sectors contribute most to the average gradient and whether the phenomenon of firm activity increasing with distance to the closest border exists within all sectors. Secondly, it allows to disentangle the impact trade has on economic activity by sectors in order to determine which sectors stand to gain least or most from trade facilitation. Thirdly and most importantly, an analysis on a sectoral level provides insight into the question which sectors mainly drive the higher average growth of firms located more closely to the border observed after an increase in trade. In this regard, the analysis further investigates the question of whether the attenuating effect of trade on the gradients of firm size and turnover along distance is mainly a reflection of activity directly related to the border crossing (i.e. construction activity at the border crossing) and services related to the transportation of goods across the border or whether it is also driven by an increased production of goods in the border region. [33] The empirical analysis of firm level data in this paper allows for a clean identification of the heterogeneous impact of trade on different sectors.

Results are qualitatively similar across sectors but vary significantly in magnitude. As a first observation, all sectors display a positively sloped gradient of turnover along distance. While this confirms that the increase of firm level activity along distance to the border is a general pattern and not just driven by some specific sectors, the steepness of the gradient varies considerably across sectors. For example, firms increase their turnover depending on the empirical specification by about 1.5 to 2 times as much for every ten kilometres of distance in the construction sector than in the transporting and storage sector. Intuitively, this can be understood as a sign of smaller locational disadvantages for transporting firms that locate close to a border (due to the potential of cross-border transporting in the case of bilateral trade) when compared to construction firms, for which a location close to the border brings more disadvantages (since it might be harder for them to receive contracts across the border due to factors like different languages). Furthermore, it has to be kept in mind that the estimated coefficients for $\hat{\beta}_1$ measure the increase in size of individual firms along distance, not the total employment or turnover at a given distance. Due to this, the gradient also reflects differences in market entry barriers, which could partly explain the steepness of the sector in the construction sector. Assuming high market entry barriers in the construction sector, the increase in activity towards the center of a country would be predominantly reflected by larger firm sizes, not by an increase in the number of firms.

A second dimension of heterogeneity across sectors is the impact of trade on economic activity. As should be expected, the effects of increasing bilateral exports to neighbour countries are most pronounced for transporting and storage as well as for wholesale and retail firms, with wholesale and retail generating even bigger increases in turnover than transporting and storage. Furthermore, the estimated coefficient for $\hat{\beta}_3$ is very small in size for transporting and storage firms compared to the three other sectors, indicating that within the border region, geographic closeness to the border is of least importance for the transport sector. In comparison, manufacturing and construction firms only increase their turnover by about half as much as transporting and wholesale and retail trade firms when exports to neighbour countries are increased. Interestingly, $\hat{\beta}_3$ is by far largest in size for the construction sector. This could partly indicate improvements in infrastructure in the direct vicinity of border crossings.

[33] Brülhart, Cadot and Himbert (2018) take a first step into analysing this question by focusing on the effect of agricultural trade depending on crop specific productivities in border regions.

A very similar picture arises when estimating the effects on employment across sectors in Table 3.6. Again, transporting and storage firms as well as wholesale and retail firms increase their employment more strongly than manufacturing and construction firms, and again geographic proximity to the border is most important for firms in the construction sector.

Table 3.5: Effects of exports on different sectors - Turnover

Dependent variable: Turnover (logs)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV
Sector:	Full Sample		Transporting and storage		Wholesale and retail trade		Manufacturing		Construction	
Distance to border (in 10km)	0.140***	0.144***	0.109***	0.092***	0.127***	0.139***	0.157***	0.125***	0.161***	0.199***
	(0.021)	(0.030)	(0.024)	(0.032)	(0.022)	(0.034)	(0.019)	(0.025)	(0.034)	(0.068)
Bilateral exports (in logs)	0.460***	0.140*	0.456***	0.126*	0.569***	0.179**	0.305***	0.065**	0.319***	0.257
	(0.091)	(0.072)	(0.106)	(0.068)	(0.134)	(0.081)	(0.067)	(0.029)	(0.070)	(0.210)
Bilateral exports × Distance to border	-0.016*	-0.020*	-0.003***	-0.002***	-0.009**	-0.015***	-0.019**	-0.009*	-0.024*	-0.036*
	(0.009)	(0.012)	(0.001)	(0.001)	(0.005)	(0.005)	(0.011)	(0.005)	(0.014)	(0.018)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓		✓		✓		✓		✓	
Country-Sector-Year FE	✓		✓		✓		✓		✓	
Region-Sector-Year FE		✓		✓		✓		✓		✓
Neighbor country-Year FE		✓		✓		✓		✓		✓
Kleibergen-Paap F statistic	29	24	27	24	26	25	28	23	25	24
# Clusters	249	239	243	231	249	239	247	236	246	236
# Observations	2,298,459	2,288,788	175,394	174,721	873,163	870,865	479,772	478,157	419,157	418,520

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 3.6: Effects of exports on different sectors - Employment

Dependent variable: Employment (logs)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV
Sector:	Full Sample		Transporting and storage		Wholesale and retail trade		Manufacturing		Construction	
Distance to border (in 10km)	0.058***	0.067***	0.053***	0.047***	0.054***	0.066***	0.049***	0.051***	0.090***	0.091***
	(0.013)	(0.013)	(0.014)	(0.011)	(0.013)	(0.013)	(0.011)	(0.011)	(0.018)	(0.017)
Bilateral exports (in logs)	0.491***	0.079***	0.505***	0.062***	0.622***	0.098***	0.229***	0.032**	0.315***	0.066***
	(0.104)	(0.027)	(0.119)	(0.022)	(0.160)	(0.032)	(0.048)	(0.016)	(0.086)	(0.022)
Bilateral exports × Distance to border	-0.012**	-0.015***	-0.012*	-0.010*	-0.015***	-0.013**	-0.009**	-0.007	-0.023***	-0.023***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.004)	(0.004)	(0.007)	(0.007)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓		✓		✓		✓		✓	
Country-Sector-Year FE	✓		✓		✓		✓		✓	
Region-Sector-Year FE		✓		✓		✓		✓		✓
Neighbor country-Year FE		✓		✓		✓		✓		✓
Kleibergen-Paap F statistic	29	24	27	24	26	25	28	23	25	24
# Clusters	249	239	243	231	249	239	247	236	246	236
# Observations	2,298,459	2,288,788	175,394	174,721	873,163	870,865	479,772	478,157	419,157	418,520

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Imports

To complement the analysis on the sectoral level, Tables 3.7 and 3.8 split the sample in the same way as previously, but focus on the impact of increased imports from the neighbour country on different sectors. The results however differ both qualitatively as well as quantitatively from the effects of exports. The gains in turnover and employment are smaller in size after an increase in imports than after an increase in exports, underlining the results of Table 3.4. Contrary to before, the biggest gains from increasing imports are reaped in the construction sector both in terms of turnover as well as in terms of employment. The relatively small difference in the magnitude of $\hat{\beta}_2$ in the construction sector suggests that this sector is relatively independent from the direction of trade flows: increased trade between two neighbouring countries primes construction firms in the border region to increase turnover and employment, whether the home country exports or imports. As construction is essentially non-tradable, this result corresponds to economic intuition. At the same time, as we observe again the size of $\hat{\beta}_3$ to be biggest in size in this sector, a similar mechanism of improvements in infrastructure in the direct vicinity of the border seems to be at play as when increasing exports.

For the transporting and storage and the wholesale and retail trade sectors, results remain qualitatively identical as in the export case, but all coefficients decrease significantly in size. The effect of trade boosting firm level activity in the border region is cut by about a third across specifications when imports are used to measure trade.

The most important change when looking at imports instead of exports occurs however in the manufacturing sector. Here, the effect of imports on overall manufacturing firm level activity in the border region is negative and statistically significant in all specifications. Furthermore, there is no strong evidence for an effect of imports attenuating the gradients of firm activity along distance as $\hat{\beta}_3$ remains statistically insignificant in three out of four specifications.

The results in this section draw a more nuanced picture of the effects of trade on border regions. While it can generally be said that trade facilitation leads to a more balanced pattern of firm level activity within the border region with firms being located more closely to the border reaping the biggest part of the gains from trade, the exact effects depend both on the sectoral composition of the border region as well as on the relative trade power compared to the neighbouring country. Border regions with a high share of manufacturing firms tend to gain least from increasing exports to the neighbour country and even risk losing employment and turnover in cases where the increase in imports due to trade facilitation outweigh the increase in exports. In any case, trade facilitation is unsurprisingly related to improvements in the infrastructure of border regions as reflected by the increased activity of construction firms.

Table 3.7: Effects of imports on different sectors - Turnover

Dependent variable: Turnover (logs)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV
Sector:	Full Sample		Transporting and storage		Wholesale and retail trade		Manufacturing		Construction	
Distance to border (in 10km)	0.126***	0.112***	0.110***	0.104***	0.120***	0.111***	0.127***	0.114***	0.144***	0.128***
	(0.018)	(0.017)	(0.013)	(0.015)	(0.018)	(0.015)	(0.017)	(0.016)	(0.030)	(0.022)
Bilateral imports (in logs)	0.253	0.058**	0.294**	0.067**	0.292	0.075**	-0.197**	-0.032***	0.296*	0.108*
	(0.132)	(0.028)	(0.145)	(0.033)	(0.304)	(0.037)	(0.096)	(0.011)	(0.152)	(0.064)
Bilateral imports × Distance to border	-0.009*	-0.003*	-0.010	-0.002*	-0.003**	-0.002*	-0.005**	-0.003	-0.015*	-0.006***
	(0.005)	(0.002)	(0.006)	(0.001)	(0.002)	(0.001)	(0.002)	(0.004)	(0.007)	(0.004)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓		✓		✓		✓		✓	
Country-Sector-Year FE	✓		✓		✓		✓		✓	
Region-Sector-Year FE		✓		✓		✓		✓		✓
Neighbor country-Year FE		✓		✓		✓		✓		✓
Kleibergen-Paap F statistic	29	24	27	24	26	25	28	23	25	24
# Clusters	249	239	243	231	249	239	247	236	246	236
# Observations	2,298,459	2,288,788	175,394	174,721	873,163	870,865	479,772	478,157	419,157	418,520

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 3.8: Effects of imports on different sectors - Employment

Dependent variable: Employment (logs)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV
Sector:	Full Sample		Transporting and storage		Wholesale and retail trade		Manufacturing		Construction	
Distance to border (in 10km)	0.057***	0.052***	0.045***	0.042**	0.049***	0.049**	0.037***	0.039***	0.084***	0.068***
	(0.018)	(0.017)	(0.014)	(0.017)	(0.018)	(0.020)	(0.013)	(0.013)	(0.026)	(0.021)
Bilateral imports (in logs)	0.076	0.036***	0.205***	0.045***	0.085	0.027***	-0.193*	-0.141***	0.291**	0.063***
	(0.062)	(0.016)	(0.075)	(0.015)	(0.069)	(0.010)	(0.100)	(0.053)	(0.143)	(0.024)
Bilateral imports × Distance to border	-0.004**	-0.001*	-0.011**	-0.010**	-0.011*	-0.010*	-0.006	-0.004	-0.023***	-0.017***
	(0.002)	(0.001)	(0.005)	(0.004)	(0.006)	(0.005)	(0.005)	(0.003)	(0.009)	(0.005)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓		✓		✓		✓		✓	
Country-Year FE	✓		✓		✓		✓		✓	
Region-Year FE		✓		✓		✓		✓		✓
Neighbor country-Year FE		✓		✓		✓		✓		✓
Kleibergen-Paap F statistic	29	24	27	24	26	25	28	23	25	24
# Clusters	249	239	243	231	249	239	247	236	246	236
# Observations	2,298,459	2,288,788	175,394	174,721	873,163	870,865	479,772	478,157	419,157	418,520

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 3.9 and Table 3.10 show results split by sectors when the regressions are run on bilateral exports and imports simultaneously.

The most important finding of this exercise is that the attenuating effect of trade on the gradient of firm activity is driven entirely by increasing exports to the neighbour country. For both outcome variables, the impact of the interaction term between exports and distance is significantly negative. Increasing imports however are consistently insignificant in explaining turnovers when interacted with distance, and exert even an effect of spatial concentration on firm level employment. The positive coefficient on the interaction term between imports and distance in Table 3.10 means that at a constant level of exports, increasing imports relatively not only decreases the firm activity in the border region overall, but also reinforces spatial concentration since most of the losses in employment are incurred at locations close to the border.

Furthermore, the impact of trade on activity of all firms in the border region is confirmed to be much stronger for exports than for imports. Bilateral imports increase firm turnovers on by about half compared to bilateral exports and even consistently decrease firm level employment in the border region once exports are held constant. The differences between sectors is similar to previous findings, with the manufacturing sector both gaining least from increased exports and losing most from increasing imports from the neighbour country.

Table 3.9: Effects of exports and imports on different sectors - Turnover

Dependent variable: Turnover (logs)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Sector:	Full Sample		Transporting and storage		Wholesale and retail trade		Manufacturing		Construction	
Distance to border (in 10km)	0.105***	0.096***	0.100***	0.086***	0.105***	0.098***	0.109***	0.105***	0.105***	0.109***
	(0.007)	(0.008)	(0.011)	(0.011)	(0.006)	(0.009)	(0.007)	(0.011)	(0.011)	(0.014)
					Effects of bilateral exports					
Bilateral exports (in logs)	0.105***	0.051***	0.098**	0.045*	0.104***	0.027**	0.044*	0.024**	0.184***	0.088***
	(0.032)	(0.017)	(0.049)	(0.025)	(0.039)	(0.013)	(0.024)	(0.012)	(0.057)	(0.021)
Bilateral exports × Distance to border	-0.012***	-0.010**	-0.011***	-0.005**	-0.006**	-0.004**	-0.011**	-0.008*	-0.021***	-0.013***
	(0.003)	(0.005)	(0.004)	(0.002)	(0.003)	(0.002)	(0.005)	(0.005)	(0.004)	(0.005)
					Effects of bilateral imports					
Bilateral imports (in logs)	0.042**	0.025	0.056**	0.026	0.047*	0.018	-0.027**	-0.006	0.031	0.018
	(0.021)	(0.019)	(0.017)	(0.029)	(0.024)	(0.027)	(0.040)	(0.013)	(0.033)	(0.028)
Bilateral imports × Distance to border	-0.001	-0.001	-0.000	0.000	-0.001	0.000	-0.001	0.002	-0.003	-0.001
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓		✓		✓		✓		✓	
Country-Year FE	✓		✓		✓		✓		✓	
Region-Year FE		✓		✓		✓		✓		✓
Neighbor country-Year FE		✓		✓		✓		✓		✓
# Clusters	249	239	243	231	249	239	247	236	246	236
# Observations	2,298,459	2,288,788	175,394	174,721	873,163	870,865	479,772	478,157	419,157	418,520

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 3.10: Effects of exports and imports on different sectors - Employment

Dependent variable: Employment (logs)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Sector:	Full Sample		Transporting and storage		Wholesale and retail trade		Manufacturing		Construction	
Distance to border (in 10km)	0.033***	0.030***	0.031***	0.027***	0.028***	0.031***	0.025***	0.031***	0.048***	0.048***
	(0.008)	(0.006)	(0.007)	(0.009)	(0.007)	(0.008)	(0.006)	(0.009)	(0.013)	(0.011)
					Effects of bilateral exports					
Bilateral exports (in logs)	0.081***	0.054***	0.102***	0.041***	0.080***	0.051***	0.049**	0.025***	0.145***	0.083***
	(0.024)	(0.015)	(0.029)	(0.010)	(0.027)	(0.014)	(0.023)	(0.010)	(0.042)	(0.023)
Bilateral exports × Distance to border	-0.014***	-0.016***	-0.015***	-0.012***	-0.013***	-0.012***	-0.014***	-0.014***	-0.025***	-0.025***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)	(0.004)
					Effects of bilateral imports					
Bilateral imports (in logs)	-0.012	-0.027**	-0.016	-0.026	-0.036	-0.035**	-0.058**	-0.091**	-0.012	-0.030
	(0.042)	(0.013)	(0.038)	(0.017)	(0.058)	(0.016)	(0.031)	(0.045)	(0.039)	(0.020)
Bilateral imports × Distance to border	0.003*	0.003	0.005***	0.004**	0.005***	0.004***	0.004***	0.005**	0.001	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)	(0.003)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓		✓		✓		✓		✓	
Country-Sector-Year FE	✓		✓		✓		✓		✓	
Region-Sector-Year FE		✓		✓		✓		✓		✓
Neighbor country-Year FE		✓		✓		✓		✓		✓
# Clusters	249	239	243	231	249	239	247	236	246	236
# Observations	2,298,459	2,288,788	175,394	174,721	873,163	870,865	479,772	478,157	419,157	418,520

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

The effect of imports on firm activity can be further disentangled by taking into account the difference between goods imported to be used as inputs (and therefore potentially contributing to the production of exported goods) and final goods. Table 3.11 and Table 3.12 show results that decompose imports into inputs and other imports. Doing this shows a clear difference between the two categories particularly in their effect on firm level employment: while imports excluding inputs show significantly decrease employment in all sectors (with the biggest losses incurred in the manufacturing sector) and exacerbate the gradient of firm activity along distance, imports of input goods increase employment and exert an attenuating effect on the gradient in a similar pattern as exports.

Table 3.11: Effects of exports and imports on different sectors - Turnover

Dependent variable: Turnover (logs)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Sector:	Full Sample		Transporting and storage		Wholesale and retail trade		Manufacturing		Construction	
Distance to border (in 10km)	0.105***	0.096***	0.100***	0.086***	0.105***	0.098***	0.109***	0.105***	0.105***	0.109***
	(0.007)	(0.008)	(0.011)	(0.011)	(0.006)	(0.009)	(0.007)	(0.011)	(0.011)	(0.014)
Effects of bilateral exports										
Bilateral exports (in logs)	0.105***	0.051***	0.098**	0.045*	0.104***	0.027**	0.044*	0.024**	0.184***	0.088***
	(0.032)	(0.017)	(0.049)	(0.025)	(0.039)	(0.013)	(0.024)	(0.012)	(0.057)	(0.021)
Bilateral exports × Distance to border	-0.012***	-0.010**	-0.011***	-0.005**	-0.006**	-0.004**	-0.011**	-0.008*	-0.021***	-0.013***
	(0.003)	(0.005)	(0.004)	(0.002)	(0.003)	(0.002)	(0.005)	(0.005)	(0.004)	(0.005)
Effects of bilateral imports (excluding inputs)										
Bilateral imports excluding inputs (in logs)	0.030	0.015	0.028**	0.013	0.040	0.010	-0.038**	-0.016**	0.010	0.018
	(0.032)	(0.013)	(0.014)	(0.011)	(0.029)	(0.008)	(0.018)	(0.008)	(0.008)	(0.014)
Bilateral imports excluding inputs × Distance to border	-0.003	-0.004	0.001	-0.001	-0.001	0.003	-0.002	-0.000	-0.007	-0.008
	(0.007)	(0.004)	(0.008)	(0.009)	(0.006)	(0.006)	(0.006)	(0.008)	(0.010)	(0.009)
Effects of bilateral imports (inputs only)										
Bilateral imports of inputs (in logs)	0.047*	0.028	0.054**	0.021	0.072*	0.010	0.046**	0.018	0.014	0.040*
	(0.024)	(0.025)	(0.027)	(0.019)	(0.037)	(0.030)	(0.019)	(0.032)	(0.033)	(0.024)
Bilateral imports × Distance to border	-0.001	0.001	-0.001	0.000	-0.000	-0.001	-0.000	0.002	-0.001	0.001
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓		✓		✓		✓		✓	
Country-Sector-Year FE	✓		✓		✓		✓		✓	
Region-Sector-Year FE		✓		✓		✓		✓		✓
Neighbor country-Year FE		✓		✓		✓		✓		✓
# Clusters	249	239	243	231	249	239	247	236	246	236
# Observations	2,298,459	2,288,788	175,394	174,721	873,163	870,865	479,772	478,157	419,157	418,520

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 3.12: Effects of exports and imports on different sectors - Employment

Dependent variable: Employment (logs)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Sector:	Full Sample		Transporting and storage		Wholesale and retail trade		Manufacturing		Construction	
Distance to border (in 10km)	0.033***	0.030***	0.031***	0.027***	0.028***	0.031***	0.025***	0.031***	0.048***	0.048***
	(0.008)	(0.006)	(0.007)	(0.009)	(0.007)	(0.008)	(0.006)	(0.009)	(0.013)	(0.011)
Effects of bilateral exports										
Bilateral exports (in logs)	0.081***	0.054***	0.102***	0.041***	0.080***	0.051***	0.049**	0.025***	0.145***	0.083***
	(0.024)	(0.015)	(0.029)	(0.010)	(0.027)	(0.014)	(0.023)	(0.010)	(0.042)	(0.023)
Bilateral exports × Distance to border	-0.014***	-0.016***	-0.015***	-0.012***	-0.013***	-0.012***	-0.014***	-0.014***	-0.025***	-0.025***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)	(0.004)
Effects of bilateral imports (excluding only)										
Bilateral imports excluding inputs (in logs)	-0.017	-0.046***	-0.023	-0.044**	-0.039	-0.044**	-0.063***	-0.109***	-0.015	-0.033
	(0.042)	(0.015)	(0.040)	(0.018)	(0.038)	(0.021)	(0.021)	(0.019)	(0.028)	(0.025)
Bilateral imports excluding inputs × Distance to border	0.004*	0.005***	0.004**	0.006***	0.004**	0.005***	0.004**	0.006***	0.004	0.006**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.005)	(0.003)
Effects of bilateral imports (inputs only)										
Bilateral imports of inputs (in logs)	0.048**	0.029**	0.032**	0.050**	0.051	0.016	0.056***	0.040***	0.021**	0.016**
	(0.023)	(0.014)	(0.015)	(0.023)	(0.041)	(0.012)	(0.009)	(0.014)	(0.020)	(0.007)
Bilateral imports of inputs × Distance to border	-0.002	-0.007**	0.004	-0.008	0.000	-0.005	0.003	-0.004	-0.011	-0.014*
	(0.007)	(0.003)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)	(0.012)	(0.008)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓		✓		✓		✓		✓	
Country-Sector-Year FE	✓		✓		✓		✓		✓	
Region-Sector-Year FE		✓		✓		✓		✓		✓
Neighbor country-Year FE		✓		✓		✓		✓		✓
# Clusters	249	239	243	231	249	239	247	236	246	236
# Observations	2,298,459	2,288,788	175,394	174,721	873,163	870,865	479,772	478,157	419,157	418,520

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

3.6.2 Effects by firm size

Another dimension along Making further use of the heterogeneity between firms, effects are compared for small and for medium sized firms separately. Table 3.13 presents the results of this comparison when turnover is used as outcome variable, Table 3.14 repeats the same exercise with employment as dependent variable. Again, splitting the sample reveals interesting quantitative differences between the two groups. On one hand, the effect of trade on all the turnover of all firms within the border region is significantly higher for small firms than for medium sized firms, as indicated by the differences in the coefficients on $\hat{\beta}_2$ in Table 3.13. On the other hand however, when looking at the impact of trade on employment, with the estimates for $\hat{\beta}_2$ being significantly larger in size for medium sized firms in Table 3.14. While small firms increase their turnovers disproportionally much in reaction to an increase in exports to the neighbour country, they lag behind in generating additional employment. This provides evidence for heterogeneity in hiring costs for different categories of firms.[34] However, the large effects on small firms' turnover indicate that small firms are very well able to benefit from trade facilitation. Furthermore, the attenuating effect of trade on the gradients of firm activity along distance affects both small and medium sized firms in sizeable magnitude. However, the effect is bigger for medium sized firms, regardless of whether turnover or employment is analysed.

Table 3.13: Differences in firmsize - Turnover

Dependent variable:	Turnover (logs)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Firm size:	Medium Sized		Small		Medium Sized		Small	
Distance to border (in 10km)	0.117***	0.160***	0.112***	0.132***	0.109***	0.174***	0.101***	0.137***
	(0.009)	(0.020)	(0.008)	(0.022)	(0.012)	(0.039)	(0.011)	(0.029)
Bilateral exports (in logs)	0.114**	0.278***	0.229***	0.551***	0.086**	0.127**	0.106*	0.143**
	(0.050)	(0.054)	(0.070)	(0.115)	(0.041)	(0.063)	(0.055)	(0.0.066)
Bilateral exports × Distance to border	-0.011**	-0.021***	-0.007***	-0.014***	-0.020**	-0.027**	-0.012*	-0.019*
	(0.005)	(0.008)	(0.002)	(0.004)	(0.010)	(0.013)	(0.007)	(0.010)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓	✓	✓	✓				
Country-Sector-Year FE	✓	✓	✓	✓				
Region-Sector-Year FE					✓	✓	✓	✓
Neighbor country-Year FE					✓	✓	✓	✓
Kleibergen-Paap F statistic		31		27		21		25
# Clusters	249	249	248	248	207	238	239	239
# Observations	522,411	522,411	1,775,689	1,775,689	326,789	514,596	1,768,023	1,768,023

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

[34] In the theoretical literature, Cosar, Guner and Tybout (2016) use non-linear hiring costs to model the reaction of the Colombian labour market to reductions in trade frictions, tariffs and firing costs.

Table 3.14: Differences in firmsize - Employment

Dependent variable: Employment (logs)	Employment (logs)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Firm size:	Medium Sized		Small		Medium Sized		Small	
Distance to border (in 10km)	0.045***	0.064***	0.037***	0.057***	0.110***	0.108***	0.034***	0.058***
	(0.007)	(0.016)	(0.007)	(0.013)	(0.013)	(0.034)	(0.007)	(0.011)
Bilateral exports (in logs)	0.248***	0.670***	0.193***	0.437***	0.147***	0.302*	0.103***	0.171***
	(0.075)	(0.128)	(0.029)	(0.107)	(0.047)	(0.176)	(0.033)	(0.058)
Bilateral exports × Distance to border	-0.022**	-0.024*	-0.011**	-0.012**	-0.013**	-0.025**	0.010**	-0.013***
	(0.011)	(0.013)	(0.006)	(0.005)	(0.006)	(0.012)	(0.005)	(0.005)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓	✓	✓	✓				
Country-Sector-Year FE	✓	✓	✓	✓				
Region-Sector-Year FE					✓	✓	✓	✓
Neighbor country-Year FE					✓	✓	✓	✓
Kleibergen-Paap F statistic		31		27		21		25
# Clusters	249	249	248	248	207	238	239	239
# Observations	522,411	522,411	1,775,689	1,775,689	326,789	514,596	1,768,023	1,768,023

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

3.6.3 Robustness

A potential issue with the use of the Orbis data in the context of this paper is that the data records firm addresses as the addresses of the headquarter of a firm and firm activity can therefore not be georeferenced at the establishment level. This could possibly distort the estimated effects if for example benefits from trade occurring in establishments close to the border are recorded as changes in employment at the headquarter address in a city towards the center of a country. However, such a mechanism should be expected to bias the estimated coefficients for $\hat{\beta}_3$ downwards in absolute terms.

As discussed earlier, one way to address this problem is to exclude large and very large firms from the estimation sample, since those companies are most likely to entertain establishments at different locations than their headquarter. Another robustness check regarding this issue consists of excluding bigger cities from the sample, where firms locate their headquarter in order to have access to the financial and public sector even if their main activity lies in more remote places. Table 3.19 reproduces the baseline estimations on a panel excluding major cities.[35] While the results are qualitatively no different from the baseline estimations, some quantitative differences need to be highlighted. Firstly, the gradient of firm level activity becomes flatter both for turnovers as well as for employment when major cities are excluded, indicating that firms in major cities are larger than comparable firms in smaller towns at the same distance to the border and that major cities tend to be located towards the center of a country. Secondly, the effect of increasing exports on activity on all firms within the border region is smaller in size when excluding major cities from the estimation. Lastly, $\hat{\beta}_3$ is larger in size when major cities are excluded from the estimation. This would support the assumption that major cities benefit more from trade facilitation than would be expected based solely on their distance to the nearest border, but also with the described scenario of headquarters being located in major cities even if firm activity that benefits from trade facilitation takes place in smaller towns closer to the border (in which case $\hat{\beta}_3$ would be downwards biased).

One additional very arbitrary choice made in the baseline analysis concerns the choice of the width of the considered border regions. Table 3.17 presents the results of sensitivity checks using different cut-off distances

[35] For the construction of this panel, I eliminate all firms located in cities that are in the top 20th percentile in terms of recorded firms within their country.

for the border crossing roads. Qualitatively, all results remain intact, with activity increasing in distance to the nearest border and trade increasing activity in general, but disproportionately so at locations close to land borders. Quantitatively, $\hat{\beta}_2$ increases in size the smaller the considered border regions are, while $\hat{\beta}_3$ diminishes in size.

As an additional robustness check, Table 3.21 tests two alternative compositions of fixed effects. Columns (1) and (2) and (5) and (6) estimate the effect of borders and exports on employment and turnover using firm level fixed effects. Results are qualitatively in line with the previous findings, but the coefficients are estimated with less precision and remain largely insignificant. In this specification, the instrument in the first stage is also weak. Columns (3) and (4) and (7) and (8) estimate an alternative version of equation (3.1), in which the country-sector-year fixed effects are decomposed into country-year and sector-year fixed effects. As this specification is slightly less demanding, results are slightly stronger and more significant.

3.7 Conclusion

The results in this paper provide further evidence for a dispersing effect of trade on economic activity in border regions. Using data on the location and performance of more than 1 million firms, I find that an increase in bilateral exports to a neighbour country due to improvements in customs clearance time at the considered border leads to an increase in firm level activity in the border region. The magnitude of this effect depends on the distance of a firm from the nearest land border, with firms located closest to border increasing their activity the most. This attenuates the gradient of firm activity along distance that exists at low levels of trade, in which otherwise identical firms are larger in terms of employment and turnover the further they are away from the border. Furthermore, when disentangling the results on a sectoral level, the data shows that manufacturing firms stand to gain the least from increased exports when compared to other sectors, and even lose both turnover as well as employment after increases in imports from the neighbor country.

The results in this paper complement the results found in Brülhart, Cadot and Himbert (2018) by focusing on the performance of already existing firms after trade liberalization and by looking deeper into the sectoral decomposition of the effects of trade on economic activity.

Appendix tables and figures

Table 3.15: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Employment (no. of employees)	11.93	36.42	1	149	2,298,459
Turnover (mln EUR)	1.34	1.95	0	100	2,298,459
Distance to border (in km)	92.13	53.84	0	199.96	2,298,459
Total exports to neighbor country (in 100 mio US dollar)	17.283	20.128	0	104.383	2,298,459
Minor road (dummy)	0.35	0.47	0	1	2,298,459
Altitude (meters)	236.19	263.14	1	3091	2,298,459
Port closer than next land border (dummy)	0.06	0.232	0	1	2,298,459
Distance to port (in km, if port dummy = 1)	45.68	35.40	0	17.378	137,997
Airport closer than next land border (dummy)	0.24	0.428	0	1	2,298,459
Distance to airport (in km, if airport dummy = 1)	74.24	38.79	0	19.59	552,116

Table 3.16: Summary statistics - Types of firms

Category	Firms	Observations
Overall	1,346,472	2,298,459
Transportation and storage	102,952	175,394
Wholesale and retail trade	480,102	873,163
Manufacturing	266,384	479,772
Construction	238,627	419,157
Medium sized firms	325,209	522,411
Small firms	1,021,263	1,775,689

Table 3.17: Baseline estimates for different cut-off distances

Dependent variable:	Turnover (logs)				Employment (logs)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IV	IV	IV	IV	IV	IV	IV	IV
Cut-off distance	200	150	100	50	200	150	100	50
Distance to border (in 10km)	0.140***	0.142***	0.155***	0.172***	0.058***	0.065***	0.071***	0.080***
	(0.021)	(0.024)	(0.017)	(0.027)	(0.030)	(0.036)	(0.038)	(0.033)
Bilateral exports (in logs)	0.460***	0.522***	0.621***	0.671***	0.491***	0.601***	0.643***	0.699***
	(0.091)	(0.162)	(0.199)	(0.196)	(0.104)	(0.187)	(0.181)	(0.231)
Bilateral exports × Distance to border	-0.016*	-0.010*	-0.005*	-0.003	-0.012**	-0.006**	-0.005*	0.002*
	(0.009)	(0.006)	(0.003)	(0.002)	(0.005)	(0.003)	(0.003)	(0.001)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓	✓	✓	✓	✓	✓	✓	✓
Country-Sector-Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Region-Sector-Year FE								
Neighbor country-Year FE								
Kleibergen-Paap F statistic	31	31	33	33	29	29	29	30
# Clusters	249	249	239	239	249	249	239	239
# Observations	2,298,459	2,298,459	2,288,788	2,288,788	2,298,459	2,298,459	2,288,788	2,288,788

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Employment is measured as the number of employees, turnover in million .

Table 3.18: First stage results

Dependent variable:	Bilateral exports (in logs)	Distance (in 10km) × Bilateral exports	Bilateral exports (in logs)	Distance (in 10km) × Bilateral exports
	(1)	(2)	(3)	(4)
Distance to border (in 10km)	0.012**	-0.216***	0.013**	-0.228***
	(0.005)	(0.056)	(0.006)	(0.055)
Product of LPI scores (in logs)	0.111***	-0.244***	0.106***	-0.240***
	(0.022)	(0.046)	(0.020)	(0.044)
Distance to border × Product of LPI scores	-0.001	0.314***	-0.000	0.324***
	(0.000)	(0.033)	(0.000)	(0.025)
Controls	ALL	ALL	ALL	ALL
Road FE	✓	✓		
Country-Sector-Year FE	✓	✓		
Region-Sector-Year FE			✓	✓
Neighbor country-Year FE			✓	✓
# Clusters	249	249	239	239
# Observations	2,410,178	2,410,178	2,399,226	2,399,226

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.19: Baseline estimates excluding major cities

Dependent variable:	Turnover (logs)				Employment (logs)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Distance to border (in 10km)	0.094***	0.110***	0.089***	0.112***	0.031***	0.044***	0.025***	0.054***
	(0.010)	(0.025)	(0.016)	(0.034)	(0.014)	(0.015)	(0.012)	(0.019)
Bilateral exports (in logs)	0.185***	0.435***	0.061	0.122**	0.124***	0.447***	0.013	0.070***
	(0.057)	(0.855)	(0.039)	(0.571)	(0.036)	(0.985)	(0.023)	(0.222)
Bilateral exports × Distance to border	-0.014***	-0.018*	-0.014***	-0.024**	-0.015***	-0.019**	-0.013***	-0.020***
	(0.003)	(0.009)	(0.003)	(0.012)	(0.005)	(0.010)	(0.004)	(0.009)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓	✓			✓	✓		
Country-Sector-Year FE	✓	✓			✓	✓		
Region-Sector-Year FE			✓	✓			✓	✓
Neighbor country-Year FE			✓	✓			✓	✓
Kleibergen-Paap F statistic		30		20		28		22
# Clusters	248	248	239	239	248	248	239	239
# Observations	1,691,265	1,691,265	1,679,243	1,679,243	1,691,265	1,691,265	1,679,243	1,679,243

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 3.20: Baseline estimates for large and very large companies

Dependent variable:	Turnover (logs)				Employment (logs)			
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Distance to border (in 10km)	0.112***	0.170***	0.107***	0.180***	0.053***	0.094***	0.046***	0.100***
	(0.011)	(0.021)	(0.014)	(0.030)	(0.010)	(0.015)	(0.013)	(0.029)
Bilateral exports (in logs)	0.160**	0.293***	0.006	0.097	0.169***	0.264***	0.053	0.126
	(0.069)	(0.060)	(0.051)	(0.059)	(0.065)	(0.057)	(0.054)	(0.098)
Bilateral exports × Distance to border	-0.007**	-0.008***	-0.008**	-0.010***	-0.004***	-0.007***	-0.005***	-0.007***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓	✓			✓	✓		
Country-Sector-Year FE	✓	✓			✓	✓		
Region-Sector-Year FE			✓	✓			✓	✓
Neighbor country-Year FE			✓	✓			✓	✓
Kleibergen-Paap F statistic		30		21		30		26
# Clusters	220	220	205	205	242	242	227	227
# Observations	114,489	114,489	107,911	107,911	114,489	114,489	107,911	107,911

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 3.21: Baseline estimates with alternative sets of fixed effects

Dependent variable:	Turnover (logs)				Employment (logs)			
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Distance to border (in 10km)	0.000	0.000	0.133***	0.169***	0.000	0.000	0.049***	0.069***
	(.)	(.)	(0.006)	(0.014)	(.)	(.)	(0.005)	(0.011)
Bilateral exports (in logs)	0.094	0.658	0.240***	0.502***	0.186	0.565***	0.167***	0.515***
	(0.108)	(0.436)	(0.066)	(0.086)	(0.169)	(0.130)	(0.038)	(0.097)
Bilateral exports × Distance to border	-0.015	-0.028	-0.014**	-0.018**	-0.007	-0.036**	-0.013***	-0.015***
	(0.014)	(0.052)	(0.007)	(0.009)	(0.012)	(0.015)	(0.003)	(0.005)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Firm FE	✓	✓			✓	✓		
Road FE			✓	✓			✓	✓
Country-Year FE			✓	✓			✓	✓
Sector-Year FE			✓	✓			✓	✓
Kleibergen-Paap F statistic		7		31		7		30
# Clusters	242	242	249	249	242	242	249	249
# Observations	1,695,045	1,695,045	2,298,167	2,298,167	1,695,045	1,695,045	2,298,167	2,298,167

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Table 3.22: Baseline estimates - Increased sample (39 countries)

Dependent variable:	Turnover (logs)				Employment (logs)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
				Effects of distance and trade				
Distance to border (in 10km)	0.147***	0.182***	0.121***	0.160***	0.057***	0.074***	0.051***	0.084***
	(0.022)	(0.033)	(0.025)	(0.039)	(0.019)	(0.024)	(0.014)	(0.023)
Bilateral exports (in logs)	0.188**	0.302**	0.061**	0.100*	0.119***	0.327**	0.035***	0.199**
	(0.090)	(0.139)	(0.029)	(0.060)	(0.043)	(0.175)	(0.011)	(0.108)
Bilateral exports × Distance to border	-0.008*	-0.011*	-0.008**	-0.010*	-0.009***	-0.009**	-0.009***	-0.011**
	(0.005)	(0.006)	(0.003)	(0.006)	(0.003)	(0.005)	(0.004)	(0.005)
Controls	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Road FE	✓	✓			✓	✓		
Country-Sector-Year FE	✓	✓			✓	✓		
Region-Sector-Year FE			✓	✓			✓	✓
Neighbor country-Year FE			✓	✓			✓	✓
Kleibergen-Paap F statistic		34		25		33		28
# Clusters	475	475	463	463	475	475	463	463
# Observations	4,512,788	4,512,788	4,486,200	4,486,200	4,512,788	4,512,788	4,486,200	4,486,200

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-way clustered standard errors at road and country-pair-year level in parentheses.

Employment is measured as the number of employees, turnover in million .

Bibliography

[1] Ades, A. F. and Glaeser, E. L. (1995) "Trade and circuses: explaining urban giants." *Quarterly Journal of Economics*, 110(1): 195–227.

[2] Ahlfeldt, G. M. and McMillen, D. P. (2015) "The vertical city: The price of land and the height of buildings in Chicago 1870-2010."

[3] Ahlfeldt, G. M. and Wendland, N. (2013) "How polycentric is a monocentric city? Centers, spillovers and hysteresis." *Journal of Economic Geography*, 13(1): 5383.

[4] Allen, T. and Arkolakis, C. (2014) "Trade and the topography of the spatial economy." *Quarterly Journal of Economics*, 129(3): 1085–1140.

[5] Alonso, W. (1964), "Location and land use. Toward a general theory of land rent." *Harvard University Press*, Cambridge, MA .

[6] Artuc, E.; Chaudhuri, S. and McLaren, J. (2010) "Trade Shocks and Labor Adjustment: A Structural Empirical Approach." *American Economic Review*, 100(3): 1008-1045.

[7] Atkin, D. and Donaldson, D. (2015) "Who's getting globalized? The size and implications of intra-national trade costs." *Mimeo*, MIT and Stanford.

[8] Autor, D. H.; Dorn, D. and Hanson, G. H. (2013) "The China syndrome: Local labor market effects of import competition in the United States." *American Economic Review*, 103(6): 2121–2168.

[9] Balk, D.; Deichmann, U. and Yetman, G. (2001) "Transforming population data for interdisciplinary usages: From census to grid." *Mimeo*, World Bank and Columbia University.

[10] Baum-Snow, N.; Henderson, J. V.; Turner, M. A.; Zhang, Q., and Brandt, L. (2017) "Does Investment in National Highways Help or Hurt Hinterland City Growth?" *Mimeo*, University of Toronto, London School of Economics and Peking University.

[11] Bayer, P., Ferreira, F. and McMillan, R. (2007) "A unified framework for measuring preferences for schools and neighborhoods." Journal of Political Economy, 115(4), 588 638.

[12] Black, S. E. (1999) "Do better schools matter? Parental valuation of elementary education." *Quarterly Journal of Economics*, pp. 577599.

[13] Blanchard, O. and Katz, L. F. (1992) "Regional Evolutions." *Brookings Papers on Economic Activity* (1): 161.

[14] Bound, J. and Holzer, H. J. (2000) “Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s.” *Journal of Labor Economics* 18 (1): 2054.

[15] Brueckner, J. K. (1987) “The structure of urban equilibria: A unified treatment of the Muth-Mills model.” *Handbook of Regional and Urban Economics*, 2, 821845.

[16] Brueckner, J. K. and Rosenthal, S.S.(2009) “Gentrification and neighborhood housing cycles: Will Americas future downtowns be rich?” *The Review of Economics and Statistics*,91(4), 725743.

[17] Bruederle, A. and Hodler, R. (2017) “Nighttime lights as a proxy for human development at the local level.” *CESifo Working Papers* #6555.

[18] Bruk, S. I. I., and Apenchenko, V. S. (1964) “Atlas Narodov Mira.” *Moscow: Glavnoye Upravleniye Geodezii i Kartografii Gosudarstvennogo Geologicheskogo Komiteta SSSR and Institut Etnografii im NN Miklukho-Maklaya Akademii Nauk SSSR*.

[19] Brülhart, M. (2011) “The spatial effects of trade openness: A survey.” *Review of World Economics*, 147(1): 59–83.

[20] Brülhart, M., Bucovetsky, S. and Schmidheiny, K. (2015) “Taxes in cities.” *Handbook of Regional and Urban Economics*, 5, 11231196.

[21] Brülhart, M.; Cadot, O., and Himbert, A. (2018) “Let There Be Light : Trade and the Development of Border Regions.” *Mimeo*.

[22] Brülhart, M.; Carrère, C., and Robert-Nicoud, F. (2018) “Trade and towns: Heterogeneous adjustment to a border shock.” *Journal of Urban Economics*, forthcoming.

[23] Brülhart, Marius, Carrère, C., and Trionfetti, F. (2012): “How wages and employment adjust to trade liberalization: Quasi-experimental evidence from Austria.” *Journal of International Economics*, 86(1): 68–81.

[24] Caliendo, L.; Dvorkin, M. and Parro, F. (2017) “Trade and labor market dynamics: General equilibrium analysis of the China trade shock.” *Mimeo*, Yale, Federal Reserve and Johns Hopkins.

[25] Chen, X. and Nordhaus, W. (2015) “A Test of the New VIIRS Lights Data Set: Population and Economic Output in Africa.” *Remote Sensing*,7(4): 4937-4947.

[26] Cheptea, A. (2013) “Border effects and European integration.” *CESifo Economic Studies*, 59(2): 277–305.

[27] Colella, F., Lalive, R., Sakalli, S. O. and Thoenig, M. (2018) “Inference with arbitrary clustering” *Mimeo*, University of Lausanne.

[28] Cosar, A. K. and Fajgelbaum, P. D. (2016) “Internal geography, international trade, and regional specialization.” *American Economic Journal: Microeconomics*, 8(1): 24–56.

[29] Cosar, A. K. and Guner, N. and Tybout, J. (2016) “Firm Dynamics, Job Turnover, and Wage Distributions in an Open Economy.” *American Economic Review*, 106(3): 625-63.

[30] Dix Carneiro, R. (2014) “Trade Liberalization and Labor Market Dynamics.” *Econometrica*, 82(3): 825–885.

[31] Dix Carneiro, R. and Kovak, B. K. (2017) "Trade Liberalization and Regional Dynamics." *American Economic Review*, 107(10): 2908–2946.

[32] Donaldson, D. and Storeygard, A. (2016) "The view from above: Applications of satellite data in economics." *Journal of Economic Perspectives*, 3(4): 171–198.

[33] Duranton, G. and Puga, D. (2015) "Urban land use." *Handbook of Regional and Urban Economics*, 5, 467560.

[34] Fack, G. and Grenet, J. (2010) "When do better schools raise housing prices? Evidence from Paris public and private schools." *Journal of Public Economics*, 94(1), 5977.

[35] Fajgelbaum, P. D. and Redding, S. J. (2014) "External integration, structural transformation and economic development: Evidence from Argentina 1870-1914." *NBER Working Paper*, #20217.

[36] Gibbons, S. and Machin, S. (2005) "Valuing rail access using transport innovations." *Journal of Urban Economics*, 57(1), 148169.

[37] Gibbons, S., Machin, S. and Silva, O. (2013) "Valuing school quality using boundary discontinuities." *Journal of Urban Economics*, 75, 1528.

[38] Guerrieri, V., Hartley, D. and Hurst, E. (2013) "Endogenous gentrification and housing price dynamics." *Journal of Public Economics*, 100, 4560.

[39] Hanson, G. H. (1998) "Regional adjustment to trade liberalization." *Regional Science and Urban Economics*, 28(4): 419444.

[40] Helpman, E.; Itskhoki, O. and Redding, S. J. (2010) "Inequality and unemployment in a global economy." *Econometrica*, 78.4: 1239-1283.

[41] Helsley, R. W. and Strange, W. C. (2008) "A game-theoretic analysis of skyscrapers." *Journal of Urban Economics*, 64(1), 4964.

[42] Henderson, J. V. (2003) "The urbanization process and economic growth: The so-what question." *Journal of Economic Growth*, 8(1): 47–71.

[43] Henderson, J. V. and Becker, R. (2000) "Political economy of city sizes and formation." *Journal of Urban Economics*, 48(3): 453–484.

[44] Henderson, J: V.; Storeygard, A. and Weil, D. N. (2012) "Measuring economic growth from outer space." *American Economic Review*, 102(2): 994–1028.

[45] Hirte, G.; Lessmann, C. and Seidel, A. (2018) "International trade, geographic heterogeneity and interregional inequality." *Mimeo*, TU Dresden and TU Braunschweig.

[46] Hodler, R. and Raschky, P. A. (2014) "Regional favoritism." *Quarterly Journal of Economics*, 129(2): 995–1093.

[47] Kanbur, R. and Zhang, X. (2005) "Fifty years of regional inequality in China: A journey through central planning, reform, and openness." *Review of Development Economics*, 9(1): 87–106.

[48] König, M. D.; Rohner, D.; Thoenig, M. and Zilibotti, F. (2017) "Networks in conflict: Theory and evidence from the Great War of Africa." *Econometrica*, 85(4): 1093–1132.

[49] Krugman, P. and Livas Elizondo, R. (1996) "Trade policy and the third world metropolis." *Journal of Development Economics*, 49(1): 137–150.

[50] Liu, C. H., Rosenthal, S. S. and Strange, W. C. (2015)"The vertical city: Rent gradients and spatial structure." *Under Review*.

[51] Maloutas, T. and Spyrellis, S. N. (2016) "Vertical segregation: Mapping the vertical social stratification of residents in Athenian apartment buildings." *Journal of Mediterranean Geography*, (127), 2736.

[52] Manski, C. F. (1993) "Identification of endogenous social effects: The reflection problem." *The Review of Economic Studies*, 60(3), 531542.

[53] McMillen, D. P. (1996) "One hundred fifty years of land values in Chicago: A nonparametric approach." *Journal of Urban Economics*, 40(1), 100124.

[54] Menezes-Filho, N. A. and Muendler, M. (2011) "Labor Reallocation in Response to Trade Reform." *NBER Working Paper*, #17372.

[55] Michalopoulos, S. and Papaioannou, E. (2013) "Pre-colonial ethnic institutions and contemporary African development." *Econometrica*, 81(1): 113–152.

[56] Michalopoulos, S. and Papaioannou, E. (2016) "The long-run effects of the Scramble for Africa." *American Economic Review*, 106(7): 1802–1848.

[57] Mills, E. S. (1967)"An aggregative model of resource allocation in a metropolitan area." *The American Economic Review*, 57(2), 197210.

[58] Monfort, P. and Nicolini, R. (2000) "Regional convergence and international integration." *Journal of Urban Economics*, 48(2): 286–306.

[59] Muth, R. F. (1969) "Cities and housing: the spatial pattern of urban land use."

[60] Ng, C. F. (2008) "Commuting distances in a household location choice model with amenities." *Journal of Urban Economics*, 63(1), 116129.

[61] Oates, W. E. (1969) "The effects of property taxes and local public spending on property values: An empirical study of tax capitalization and the Tiebout hypothesis." *Journal of Political Economy*, 77(6), 957971.

[62] Parchet, R. (2014) "Are local tax rates strategic complements or strategic substitutes?" *Technical report*, Universit della Svizzera italiana.

[63] Pinkovskiy, M. L. (2017) "Growth discontinuities at borders." *Journal of Economic Growth*, 22(2): 145192.

[64] Pinkovskiy, M. L. and Sala-i-Martin, X. (2017) "Lights, camera ... income! Illuminating the national accounts-household surveys debate." *Quarterly Journal of Economics*, 131 (2): 579-631.

[65] Ramankutty, N.; Evan, A. T.; Monfreda, C. and Foley, J. A. (2008) "Farming the planet: Geographic distribution of global agricultural lands in the year 2000." *Global Biogeochemical Cycles* 22(1): 1–19.

[66] Rauch, J. E. (1991) "Comparative advantage, geographic advantage and the volume of trade." *Economic Journal*, 101(408): 1230–1244.

[67] Redding, S. J. (2016) "Goods trade, factor mobility and welfare." *Journal of International Economics*, 101: 148–167.

[68] Redding, S. J. and Rossi-Hansberg, E. (2017) "Quantitative spatial economics." *Annual Review of Economics*, 9: 21–58.

[69] Redding, S. J. and Sturm, D. M. (2008) "The costs of remoteness: Evidence from German division and reunification." *American Economic Review*, 98(5): 1766–1797.

[70] Rossi-Hansberg, E. (2005) "A spatial theory of trade." *American Economic Review*, 95(5): 1464–1491.

[71] Schmidheiny,K.(2006) "Income segregation and local progressive taxation: Empirical evidence from Switzerland." *Journal of Public Economics*, 90(3), 429458.

[72] Storeygard, A. (2016) "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa." *Review of Economic Studies*, 83(3): 1263-1295.

[73] Sutton, P.; Elvidge, C.; and Ghosh, T. (2007) "Estimation of gross domestic product at sub-national scales using nighttime satellite imagery." *International Journal of Ecological Economics & Statistics*, 8(S07): 5–21.

[74] Weidmann, N.; Rød, J. and Cederman, L. (2010) "Representing ethnic groups in space: A new dataset." *Journal of Peace Research*, 47(4): 491499.

[75] Wong, S. K., Chau, K. W., Yau, Y. and Cheung, A. K. C. (2011) "Property price gradients: The vertical dimension." *Journal of Housing and the Built Environment*, 26(1), 33 45.

[76] Wright, C. (1971) "Residential location in a three-dimensional city." *Journal of Political Economy*, 79(6), 13781387.

[77] Zuk, M., Bierbaum, A. H., Chapple, K., Gorska, K., Loukaitou-Sideris, A., Ong, P. and Thomas, T. (2015) "Gentrification, displacement and the role of public investment: A literature review." *Federal Reserve Bank of San Francisco*.