ESSAYS IN PUBLIC AND URBAN ECONOMICS

BY

AMRITA KULKA

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Economics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2020

Date of final oral examination: 6/22/2020

The dissertation is approved by the following members of the Final Oral Committee:

Christopher R. Taber, Professor, Economics

Jesse M. Gregory, Associate Professor, Economics

Matthew J. Wiswall, Professor, Economics

J. Michael Collins, Professor, Consumer & Personal Finance

Dedication

To my grandparents whose never-ending encouragement and unconditional support always guide me.

Acknowledgments

First and foremost I want to thank my advisors Chris Taber and Jesse Gregory and my committee members Matt Wiswall and Michael Collins for their permanently open doors and patience. I feel incredibly fortunate to have been advised by them and to have had the opportunity to learn from them. I especially cherish my memories of joint meetings with Chris and Jesse during which I would hope that a tiny part of their brilliance could transfer to me. I thank Jenna Nobles for being an incredible mentor and the best role model of an interdisciplinary researcher that one could hope for. I am grateful to the research community at the Center for Demography and Ecology for supporting me and helping me become a better economist by widening my horizon outside of economics.

The first chapter of my dissertation was made possible by the Wake County Geographic Information System team. I thank them for providing parcel information and being extremely helpful and patient in answering my many questions about zoning in Wake County. In addition, I am grateful to the Wake County Public School System for providing boundary maps and institutional information. Further, I also thank Jeff Smith, Naoki Aizawa, Chao Fu, J. Michael Collins, Andra Ghent, Alina Arefeva, Timothy Riddiough, Yongheng Deng, Devin Michelle Bunten, Steve Billings, Raven Molloy, Sarena Goodman, Brigitte Roth-Tran, Eric Nielsen, Dan Ringo, Elliot Anenberg, Ralf Meisenzahl, Keren Horn, Francesco Berlingieri and seminar participants at the Federal Reserve Board, the University of Wisconsin-Madison, UW Milwaukee, the Urban Economics Association Meetings and APPAM for their helpful comments and advice.

I want to thank Dennis McWeeny, for enlightening me about life in rural Iowa as much as about BLP and for making coauthoring so enjoyable and productive. For additional comments on the second chapter, we thank Gary Baker, Victor Couture, Sarah Johnston, John Kennan, Raven Molloy, Elan Segarra, Christopher Taber, Emily Walden, and seminar participants at the University of Wisconsin-Madison, the 2018 Urban Economics Association Annual Meeting, and the 2018 Population Association of America Annual Meeting.

Going through the Ph.D. program would not have been possible without the support of my parents and brother who dove with me into the unknown world of graduate school in economics and have helped me maintain perspective and keep in mind the bigger picture. You gave me fresh energy during every visit and call and continued to sing the praises of Madison in all seasons. I also want to thank my uncle for continuing to believe in me.

Thanks to my wonderful office mates Elise Marifian, Andy Lehrer, Natalie Duncombe, Jonathan Becker and Federico de Andrea for making the office a second home, keeping it stocked with snacks and letting me turn it into a greenhouse one spring.

Finally, I want to thank the people that became my family in Madison for making grad school life fun and for cheering me on, for keeping me sane and sticking with me through the difficult times. Thank you Gary Baker, Moshi Ul Alam, Dennis McWeeny, Suchi Akmanchi, Arpita Patnaik, Elan Segarra, Rebeca Fernandez, Bre Begley and Ramneek Button for countless study sessions in memorial library, conversations about research and life, tech support, lunches, dinners, brunches, coffee breaks, walks, owl prowls and runs (both in quarantine and otherwise). I would have not made it to the end without all of you.

Abstract

This dissertation studies different ways in which government policies affect the location decisions that individuals and households make and the outcomes of movers in their destination location.

The first chapter examines how land use regulations induce households to sort by income into differently regulated neighborhoods using the example of minimum lot sizes. By imposing a floor on housing consumption, high minimum lot sizes play a role in limiting access to neighborhoods that may have more amenities. First, I quantify the effect of minimum lot size regulations on neighborhood composition using a boundary discontinuity design. I find that at boundaries with the average level of regulation, average household income increases by 4.5% for a decrease in the density by one dwelling unit per acre. Next, I develop a neighborhood choice model in which minimum lot sizes affect the trade-off between neighborhood amenities and consumption. I use the model to study the effects of relaxing minimum lot size restrictions on sorting and welfare. Average neighborhood income is 30% lower in a high-amenity neighborhood with relaxed minimum lot size. The price of vacant land that is regulated in different ways depends on the degree to which the regulation imposes a constraint on demand given the available quantity of land.

The second chapter (joint with Dennis B. McWeeny) analyzes the effects of government incentive programs intended to eliminate physician shortages in rural areas. Using a differences-in-differences approach, we estimate that student loan forgiveness programs lead to an increase of three physicians on average per rural county. We then estimate a structural model of physician specialty and location decisions to simulate counterfactual incentive policies. Physicians are relatively unresponsive to differences in salaries across locations and prefer to practice medicine in their home state. Consequently, current incentive payments are too small to eliminate shortages.

The third chapter evaluates the effects of Germany's "Integration course" - a 600 hour language class - on language learning and investigates the causal effect of language proficiency on labor market outcomes for immigrants. I use data from the German Socio Economic Panel to estimate a fuzzy regression discontinuity model. I find large and robust increases in the probability of full-time employment and in monthly income for immigrants that stem from increases in proficiency of spoken German. For refugees, I find that chances of being employed either part-time or full-time double for cohorts that are more exposed to the introductory language class.

Contents

Chapter 1

Sorting Into Neighborhoods: The Role of Minimum Lot Sizes

Chapter Summary

This paper examines how land use regulations induce households to sort by income into differently regulated neighborhoods using the example of minimum lot sizes. By imposing a floor on housing consumption, high minimum lot sizes play a role in limiting access to neighborhoods that may have more amenities. First, I quantify the effect of minimum lot size regulations on neighborhood composition using a boundary discontinuity design. I find that at boundaries with the average level of regulation, average household income increases by 4.5% for a decrease in the density by one dwelling unit per acre. This is a considerable effect compared to the sorting induced by boundaries with sharp changes in public goods provision. Next, I develop a neighborhood choice model in which minimum lot sizes affect the trade-off between neighborhood amenities and consumption. I use the model to study the effects of relaxing minimum lot size restrictions on sorting and welfare. Average neighborhood income is 30% lower in a high-amenity neighborhood with relaxed minimum lot size. I find that households with at least the area's median income benefit from this policy. The price of differently regulated vacant land depends on the degree to which the regulation imposes a constraint on demand given the available quantity of land.

1.1 Introduction

As state legislatures and municipalities across the United States debate relaxing land use regulations as a solution to the affordable housing crisis, residents of low-density neighborhoods continue to fight against higher density housing in their backyard.[1] Minimum lot size restrictions are one such regulation. They are ubiquitous within areas zoned exclusively for single-family homes that make up the large majority of residential land in the US. By imposing a lower bound on the size of allowable lots, the minimum lot size regulation introduces a minimum amount of land consumption.[2] Advocates of minimum lot sizes argue that they exist for two main reasons. First, they protect property values from an alternative in which any density of housing can be built anywhere (Fischel, 2001). Second, theoretical arguments in public finance suggest that they ensure the efficient provision of local public goods such as education or local police stations (Hamilton, 1975). When local public goods are financed through property taxes on the square footage of land, households may prefer their neighbors abide by a large minimum lot size to keep the fiscal burden of smaller-than-average properties low. Critics of land use regulations argue that they are a means for rich households to exclude others from their neighborhood by making housing more costly.

Minimum lot sizes may impact residential decisions and thereby access to neighborhoods and related amenities and opportunities.[3] Nevertheless, minimum lot sizes (and land use regulations more broadly) are typically omitted from the study of neighborhood choice. By imposing different housing consumption floors across neighborhoods, minimum lot sizes can potentially distort the location decisions of households. They could also be contributing to

[1] New York Times, June 18 2019. https://www.nytimes.com/interactive/2019/06/18/upshot/cities-across-america-question-single-family-zoning.html

[2] For example, if an area is regulated with a minimum lot size of one acre, then households seeking to live there must purchase at least one acre of land. For a given price of a square foot of land, an increase in the lot size raises the minimum expenditure on housing.

[3] For the purposes of this paper, the amenities of a neighborhood encompass all factors that households value which are not related to either the characteristics of a particular house or the individual's idiosyncratic situation (i.e. individual's distance to work). Neighborhood amenities are the sum of exogenous characteristics of a location such as geographical features and local public goods (e.g. public libraries, police stations), public education and neighbor composition (e.g. income, race and education of neighbors).

spatial segregation on income by sorting high-income households into low-density neighborhoods and low-income households into high-density neighborhoods. Since neighborhoods fundamentally shape the resources available to households, these types of sorting patterns may also lead to inequality of opportunities.[4] There is suggestive evidence that the strictness of land use regulation is correlated both with attractive neighborhood amenities like good public schools and higher housing expenditure (Rothwell, 2012). Understanding the impact of minimum lot sizes on location choice is important both to study this overlooked channel for neighborhood composition as well as to answer a question of current policy relevance.

In this paper, I aim to fill that gap in our knowledge by studying the effect that minimum lot size regulations have on households' location choices. To what extent do minimum lot sizes induce households to sort by income into different neighborhoods? Would a high-amenity neighborhood with relaxed minimum lot sizes change sorting on income? How much are households willing to pay for the availability of such a neighborhood?

The paper has two main parts. In the first part, I focus on quantifying the causal effect of the minimum lot size on neighborhood composition, mainly in terms of income (question 1). In the second part, I consider counterfactual scenarios in which I change the minimum lot size in either existing or vacant neighborhoods with high or low amenities. I concentrate on households' trade-off between neighborhood amenities and consumption which is affected by the minimum lot size. Using the counterfactual policies, I then ask how sorting into neighborhoods changes and which households benefit (questions 2 and 3).

To answer the first question, I create a novel data set of lot-level data on minimum lot sizes for Wake County in North Carolina.[5] I combine this with data on property transactions merged with information on mortgages that contain income and race of the borrower. While

[4] There are several examples in the literature that study how residential segregation affects educational outcomes of children. Card and Rothstein (2007) and Cutler and Glaeser (1997) examine the black-white achievement gap in school as a function of city-level segregation. Guryan (2004), Baum-Snow and Lutz (2011) and Angrist and Lang (2004) study school outcomes after the introduction of busing experiments for the purposes of desegregation.

[5] The second part of the paper will combine important information from this data set with information on the location choices of households with elementary school age children to estimate a structural model.

much of the previous literature has focused on studying the strictness of land use regulations across broad geographic areas (Glaeser and Gyourko, 2002; Ganong and Shoag, 2017; Hsieh and Moretti, 2015; Banzhaf and Mangum, 2018), I study the regulation and its implications for neighborhood sorting at a very local level.

To quantify the effect of minimum lot sizes on neighborhood composition, I employ a boundary discontinuity design at the minimum lot size boundary. By comparing houses within a narrow distance around the regulation boundary, this design addresses the endogeneity of the regulation, that is, the fact that places that are intrinsically more attractive for reasons that the researcher cannot observe have historically implemented a stricter regulation.

I find that the minimum lot size regulation has a considerable effect on income sorting into neighborhoods. At a boundary that is regulated with the average density regulation in Wake County, a decrease in the allowed density by one dwelling unit per acre implies an increase in average neighborhood income by 4.5%. The effects are driven by boundaries where the regulation is salient in the sense that a sizable fraction of lots are exactly of the minimum lot size and not larger. Compared to other sorting estimates in the literature (Bayer, Ferreira and McMillan, 2007; Black, 1999), this value implies that sorting on income into differently regulated areas is larger than sorting on income into different school attendance areas. While my findings stem from comparing areas with the same level of public goods, it stands to reason that if more tightly regulated areas on average also provide a higher quality of public goods, then large minimum lot sizes make those neighborhoods inaccessible to some households.

The mechanisms underlying this sorting could be twofold. First, the regulation increases housing expenditure. I find that on average lots are 19% larger and houses 7.5% more expensive on the more regulated side of a minimum lot size boundary. Second, if the density regulation itself creates an amenity, then households might be sorting due to their preference for that amenity. For example, households might have a high willingness to pay for low density or to live near the type of households that live on tightly regulated lots. I test for the importance of this effect by comparing areas near the boundary to areas on the interior of the boundary that are regulated in the same way but are surrounded by neighbors who are all subject to the same level of the regulation. I find that the preference for regulated neighbors is small compared to the primary effect that leads to larger, more valuable houses when the regulation is strict.

In the second part of the paper, I develop and structurally estimate a static discrete model of neighborhood choice to contextualize the empirical finding that minimum lot sizes bind and are an important component of neighborhood sorting. Using only the estimates from the boundary discontinuity design, it is not possible to elicit the willingness to pay of households with different observable characteristics for counterfactual scenarios. I am particularly interested in studying counterfactuals in which the minimum lot size is relaxed in a neighborhood with an estimated high level of the amenity value. Current policy debates center around two questions: How can vacant land be regulated to provide a larger variety of housing options once it is developed? Would existing neighborhoods become more affordable if land use regulations were relaxed? The model provides an answer to the first question and addresses the second question for small changes in the regulation.

In this model, households have preferences over land consumption, non-land consumption and neighborhood amenities. The minimum lot size is an exogenous feature of each neighborhood. Rather than taking a stance on why the regulation exists, the model focuses on the constraint it imposes on location choices of households that are sensitive to the constraint. Sorting on income arises naturally as a consequence of the land use regulation. The minimum lot size imposes a constraint on land consumption and affects the price of land. To study a policy that changes land use regulations, it is important to understand how households trade off consumption and neighborhood amenities. If households have a high preference for consumption or large lots, then reducing the minimum lot sizes in areas with desirable amenities is not likely to have a large effect. Households will prefer to buy cheaper and larger houses in less expensive areas. On the other hand, if households care strongly about neighborhood amenities, they will prefer to spend a large fraction of their income on housing but live in a nicer neighborhood once it becomes affordable after the land use regulation has been relaxed.

I estimate the model using the location choices of households with elementary school age children from administrative school records in North Carolina. I combine the location information of children with information from the property data on the available lot options in each neighborhood. This information is key to identify the parameters related to the marginal utility of consumption. More specifically, there are locations in which the minimum lot size regulation binds for all lots (i.e. all lots are exactly of the minimum lot size). In these neighborhoods, the marginal utility of consumption can be estimated by exploiting the variation in income of households that choose to locate there.

Having estimated the parameters of the model, I study two counterfactual scenarios: In the first counterfactual, I consider an area equivalent to my sample in all ways except for one neighborhood with amenities at the 75th percentile in which the minimum lot size is relaxed from half an acre to a tenth of an acre. I study sorting on income into this neighborhood and find that there is a drop in average income of 30% stemming from the fact that this neighborhood is now more affordable. I calculate the compensating variation required to make households in the counterfactual scenario indifferent to the true scenario. I find that this particular change in policy does not benefit households with household income less than $45,000. For households with the median income of Wake County, $75,000, I compute a compensating variation of $900 annually and for households with household income of $85,000, I find that the compensating variation is roughly $2,500 annually. These findings suggest that relaxing land use regulations at this high level of the amenity value benefits households of median income but is not attractive to households at the left tail of the income distribution.

Developing and zoning vacant land in cities is one of the key aspects of urban planning for local governments. In the second counterfactual, I study prices of and sorting into vacant land that is regulated in different ways. I consider a new vacant neighborhood that is endowed with either high or low neighborhood amenities and a high or low minimum lot size. I also vary the size of this vacant neighborhood. Then, I study the ensuing land price and average neighborhood income in the new neighborhood for different combinations of the minimum lot size, amenity value and size. I find that the affordability of vacant land depends on the sum of two opposing effects on the land price: as the minimum lot size decreases in a neighborhood, demand for square feet of land in that neighborhood increases, putting an upward pressure on the price. At the same time, some households that were previously purchasing lots of exactly the minimum lot size, now switch to demanding a smaller amount of land which puts a downward pressure on the price. In particular, it is possible that the new neighborhood is very expensive even if the minimum lot size is relatively low. When the vacant neighborhood is large enough that the upward pressure on the price does not dominate the overall price effect, reducing the minimum lot size from 0.2 acres to 0.125 acres induces sorting that decreases average neighborhood income by 15%. The model therefore captures the behavior of households that are constrained by the minimum lot size regulation but have a high valuation for the amenities in tightly regulated areas. Relaxing the regulation in a neighborhood with attractive amenities leads to a location that households of median income will choose with a high probability.

Finally, I use the predictions for neighborhood income stemming from the boundary discontinuity design to validate the predictions of the model. I also use the comparison to discuss the limitations of the boundary discontinuity design. When general equilibrium effects on land prices are small, both methods predict roughly similar results. However, when a change in the regulation leads to large changes in the price as is the case in the second counterfactual, the boundary discontinuity diverges from the model in terms of the size and sometimes the sign of the effect.

This paper contributes to three strands of literature: the literature on neighborhood sorting models, the literature on land use regulations more broadly, and the literature on multi-community models with local jurisdictions.

First, prior work on neighborhood choice has not considered minimum lot sizes or other land use regulations as a constraint for households choosing their location and focuses instead on designing optimal housing vouchers or on using location choices to estimate willingness to pay for neighborhood amenities (Bayer, Ferreira and McMillan, 2007; Black, 1999, Davis et al., 2017; Galiani, Murphy and Pantano, 2015). Researchers employ boundary discontinuity designs to elicit preferences for neighborhood characteristics like school quality or neighbors' race (Black, 1999; Bayer, Ferreira and McMillan, 2007). By contrast, I exploit variation in minimum lot sizes that occurs within areas where households have already sorted on their preferences for local public goods. Importantly, I exclude regulation boundaries that overlap with boundaries that determine school attendance since income sorting at those boundaries cannot be distinguished from willingness to pay for school quality. Turner, Haughwout and van der Klaauw (2014) also use a boundary discontinuity approach at a land use regulation boundary but estimate the effect on land prices.

Second, there is a large body of research considering land use regulations and the housing market. The effect of land use regulations on the elasticity of housing supply and house prices is fairly well understood: stricter land use regulations lead to a reduction in housing supply and hence to higher property values and sales prices (Glaeser and Gyourko, 2002; Gyourko, Saiz and Summers, 2008; Glaeser, Gyourko and Saks, 2005; Glaeser and Ward, 2009).[6] Instead of studying the effect of land use regulations on house prices, I study how they affect income sorting. Though land use regulations have been shown to increase house prices, the extent to which they lead to neighborhoods that are sorted on income is not clear given that households also sort on their preferences for different neighborhood amenities. Banzhaf and Mangum (2018) consider that land use regulations lead to a two-price tariff where owners pay a ticket price for a neighborhood as well as a marginal price

[6]From a more macroeconomic perspective, Ganong and Shoag (2017) and Hsieh and Moretti (2015) show that regional income convergence and productivity have decreased because land use regulations raise the cost of housing.

per square foot of land.[7] It is possible that minimum lot size regulations contribute little in addition to sorting on neighborhood amenities or that they sort households of similar incomes on their preferences for different types of housing (e.g., larger lots and houses).

Some papers have studied counterfactuals in which zoning regulations are relaxed. Anenberg and Kung (2018) find that even though regulations might lead to reduced housing supply, increasing housing supply may not be effective in reducing rents because rent elasticity is low. Bunten (2017) calculates that in a counterfactual without zoning, GDP growth due to a more efficiently allocated workforce is counteracted by similar declines in welfare caused by congestion externalities from more crowding. I contribute to this literature by considering how minimum lot sizes affect neighborhood demand.

Finally, previous models of neighborhood sorting that do consider zoning regulations have been mostly theoretical (Calabrese, Epple and Romano, 2007; Epple and Sieg, 1999; Fernandez and Rogerson, 1997*a*; Hanushek and Yilmaz, 2015) and focused on the provision of local public goods through voting mechanisms in local jurisdictions (Fernandez and Rogerson, 2003; Hoxby, 1996; Hoxby, 2001; Kotera and Seshadri, 2017). I provide empirical results for a model in which the zoning regulation is exogenous to households making their location decision, but the minimum housing expenditure is explicitly taken into account. I provide evidence of the bindingness of the minimum lot size by studying sorting on income in a counterfactual with relaxed regulations.

The paper proceeds as follows. Section 1.2 discusses institutional details of land use regulations in the context of Wake County. Section 1.3 introduces the different sources of data used in this study. Section 1.4 quantifies the effect of minimum lot size regulation on neighborhood composition using a boundary discontinuity design. Section 1.5 develops a model of neighborhood choice that takes into account minimum lot sizes and discusses identification of the model as well as estimation results. Section 1.6 considers two types of counterfactuals in which minimum lot size regulations are relaxed. Section 1.7 concludes.

[7] This ticket price in some sense simulates communities in Tiebout (1956) that are created based on a head tax.

1.2 Wake County and institutional background on land use regulations

Land use regulations can take on many forms including height restrictions, density regulations, minimum lot size, maximum house size, bans on multi family houses and apartments, setback rules, minimum parking requirements etc. In this paper I focus on minimum lot sizes that are one of the most predominant forms of land use regulations in the US. They have the advantage of affecting the decision of households seeking to purchase a home in a tangible way. Usually, a given municipality will not use all possible regulations simultaneously, e.g. some areas have minimum lot size regulations while others have maximum house sizes or prescribe a ratio of house size to lot size. Most regulations are ultimately intended to reduce dwelling density in one way or the other, and different types of regulations are substitutes for each other in achieving this goal. Therefore, even though considering further regulations is an important task in the future, it seems that one can learn about the fundamental way in which land use regulations operate to restrict density from examining the effect of minimum lot size regulations.

Land use regulations have been a part of US cities since the early 20th century and today almost all land with very few exceptions underlies a zoning regulation. They were introduced for the purposes of regulating where growth of towns and cities would occur as well as to protect home owners from property devaluations through changes in the neighborhood character through higher density or adjacent location of industrial uses (Fischel, 2001). As highways were constructed and jobs decentralized from inner cities to suburbs in the 1970s, the threat of a potential change in character for low-density suburban neighborhoods lead to the establishment of increasingly stricter land use regulations. Part of the neighborhood character seemingly is the composition of the neighbors themselves: The fact that even as school funding was increasingly decentralized and shifted away from local property taxes zoning regulations did nevertheless not change, implies that residents of the suburbs were also attempting to exclude households based on income (Fischel, 2001). In response to increasing civil rights law suits arguing that zoning lead to segregation on race and income, zoning regulations in some areas evolved into a complete ban on new construction which implies a ban of high density and low density regulation both.

In this paper I study minimum lot size regulations in Wake County, North Carolina. Raleigh, the main city in Wake County, adopted its first zoning ordinance in 1923. The city and surrounding suburbs experienced high growth since then and the zoning code developed simultaneously. According to the Community Inventory for the City of Raleigh, “The predominant pattern of development since 1950, representing the vast majority of the City’s built environment, has been one of low density residential development...”. Other municipalities in Wake County followed suit. The basis for today’s zoning code in Wake County was laid between 1950 and 1970. Decisions on new zoning plans are made by and appeals for rezoning are heard by the Planning Boards/Commissions of each municipality in Wake County. Unincorporated areas are served by the Planning Board of Wake County.

A caveat of this paper is that results are estimated using data from and the institutional background of Wake County. How valid are these results outside of Wake County? The answer to this question boils down to how representative Wake County is of other localities within the US. Wake County is the second most populous county in North Carolina and looks representative of the US average in terms of racial and age composition. It deviates from the national average in terms of higher educational attainment of its population as well as a higher population growth rate. Wake County is one unified school district and follows a neighborhood based school system like the vast majority of the United States.

To understand the extent to which results from Wake County generalize, it is crucially important to know how comparable Wake County is to other locations in the United States in terms of its land use regulations. While this paper uses lot-level data on land use regulations, previous papers have mostly made use of the Wharton Residential Land Use regulation index (Gyourko, Saiz and Summers, 2008) to study variation in the tightness of zoning regulations.

Table 1.1: Wharton Residential Land Use Index Comparison

	Raleigh-Durham MSA		Full WRLURI sample		Means comparison test	
	Mean	SD	Mean	SD	Difference	SE
Min lot size requirement, 0=no, 1=yes	0.71	0.49	0.84	0.36	0.13	(0.18)
≤0.5 Acre minlotsize requirement, 0=no, 1=yes	0.67	0.52	0.66	0.47	-0.01	(0.21)
>0.5 Acre minlotsize requirement, 0=no, 1=yes	0.50	0.58	0.39	0.49	-0.11	(0.29)
>1 Acre minlotsize requirement, 0=no, 1=yes	0.25	0.50	0.29	0.45	0.04	(0.25)
>2 Acres minlotsize requirement, 0=no, 1=yes	0.25	0.50	0.24	0.43	-0.01	(0.25)
Affordable housing requirement, 0=no, 1=yes	0.00	0.00	0.19	0.39	0.19***	(0.01)
# Applications for zoning changes approved (last year)	23.86	22.18	8.51	27.33	-15.39	(8.40)
Density restrictions importance, 1-not at all, 5-very	2.86	0.69	3.01	1.39	0.15	(0.26)
School crowding importance, 1-not at all, 5-very	2.86	1.07	2.23	1.28	-0.63	(0.40)
Wharton Residential Land Use Regulation Index	0.72	0.45	-0.00	1.00	-0.72**	(0.17)
Observations	7		2729		2729	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ *** $p < 0.001$

This table compares Wake County to the rest of the country using the ten questions from the Wharton Land Use Survey that ask about the strictness and importance of density regulations. Columns 1 and 2 show the mean and standard deviations for the respondents in Wake County, columns 3 and 4 show the mean and standard deviations for the entire Wharton sample and columns 5 and 6 report the difference in means and standard errors from a means test.

The Wharton survey currently provides the easiest comparison of land use regulations across municipality and county boundaries. Table 1.1 compares the 7 municipalities in the Raleigh-Durham-Chapelhill metropolitan area that answered the Wharton Land Use Survey with the other respondents from across the country. In terms of the overall index (Wharton Residential Land Use Regulation Index), the Raleigh-Durham MSA is statistically significantly more regulated than the average respondent. Gyourko, Saiz and Summers (2008) classify municipalities as moderately regulated if the the WRLURI is between -0.55 and 0.74. The average in the Raleigh MSA is 0.72 situating it at the upper end of the moderately regulated category (Table 1.1, lines 10). The regulations addressed in this paper concern the minimum lot size. In terms of this category Raleigh MSA is not statistically significantly different than the average MSA (Table 1.1, lines 1-5). 70% of respondents have minimum lot size requirements in place, 25% have particularly restrictive lot sizes such as 1 or 2 acre lot requirements. No municipality in this MSA has affordable housing requirements (Table 1.1, line 6) and on average this MSA rates the school crowding and density restrictions as fairly important for regulating the rate of residential development in their communities (line 8 and 9). Judging from this evidence, it seems that Wake County can be classified as being representative of a moderately regulated metropolitan area with a central city and surrounding suburbs.

1.3 Data

I compile data in a novel way from various sources. The first half of my paper uses lot level information from Zillow matched with data from the Home Mortgage Disclosure Act (HMDA) to study the effect of minimum lot sizes on neighborhood composition. The second half of the paper focuses on the location choices of households with children in elementary school. It uses the observed location decisions for that population from administrative school data for North Carolina. Both parts of the paper make use of minimum lot size regulation data that I standardize across Wake County at a lot level. I now briefly describe the different sources of data that I use.

Lot level data

The main data set that I use for the boundary discontinuity design in the first part of the paper is the property data company Zillow's Assessor and Real Estate Database (ZTRAX)[8]. Zillow standardizes information on property transactions, mortgages and property taxes that counties collect. For Wake County, ZTRAX contains information on deed transfers since 1997. While the Zillow data contains property characteristics and sale price as well as information about the deed, it does not contain any information about the buyer of the property. This information is crucial for me to study the effect of minimum lot sizes on neighborhood composition in terms of income and race instead of the more frequently studied effect of land use regulation on property values.

To solve this data problem I perform a Home Mortgage Disclosure Act (HMDA) match following the literature in urban economics (for example Bayer, Ferreira and McMillan, 2007; Bayer et al., 2016 and Diamond and McQuade, 2019). HMDA collects information on race and household income of individuals applying for home mortgages. By merging the transactions data from Zillow for Wake County with data from mortgage applicants from the Home Mortgage Disclosure Act, I am able to get individual lot-level data on income and

[8]https://www.zillow.com/research/ztrax/

race. The matching is done on the basis of loan amount, property census tract, transaction date and name of the lender. I follow the approach taken by Bayer et al. (2016) and Billings (2018) to conduct the merge. Bayer et al. (2016) show that this matching procedure leads to a representative sample in terms of income and race for home owners compared to Census data. The merge results in roughly 55% of matched property transactions from ZTRAX which corresponds to the matched fraction of of transactions in Billings (2018) for North Carolina. After the matching procedure I end up with a sample of 48,679 matched property transactions from 1997 to 2017 for whom the closest boundary is also an admissable boundary (discussed further in Section 4). The matched transactions are dispersed throughout the entire county.

I further use the Zillow data to obtain the set of available lot sizes in a given block group. This set is a crucial input into the neighborhood choice model. In particular, the variation in available lot sizes across neighborhoods is important for the identification of the marginal utility of consumption.

Table 1.2: Summary statistics for HMDA-matched Zillow data

	dwelling units per acre (du/a)					
	1 du/a		10 du/a		All	
	Mean	SD	Mean	SD	Mean	SD
Sale price (in $1,000)	362	2185	200	137	261	186
Min. lot size (acres)	1.00		0.10		0.22	0.25
Property acres	1.05	2.01	0.15	0.17	0.30	0.69
Realized lot size/Min lot size	1.07	2.01	1.49	1.69	1.42	2.82
Building age (years)	14.53	16.32	32.66	32.63	24.78	22.56
Household income (in $1,000)	136	936	99	155	110	146
Loan amount (in $1,000)	320	170	1771	112	229	181
White	0.78	0.42	0.82	0.38	0.77	0.42
Black	0.09	0.28	0.10	0.30	0.12	0.32
Observations	2,949		9,530		48,679	

Notes: This table shows summary statistics for the sample of Zillow transactions data in Wake County that I merge with data from the Home Mortgage Disclosure Act (HMDA). Row 4 shows summary statistics for the ratio of actual lot size to minimum lot size. When this ratio is 1, the regulation can be considered to be binding.

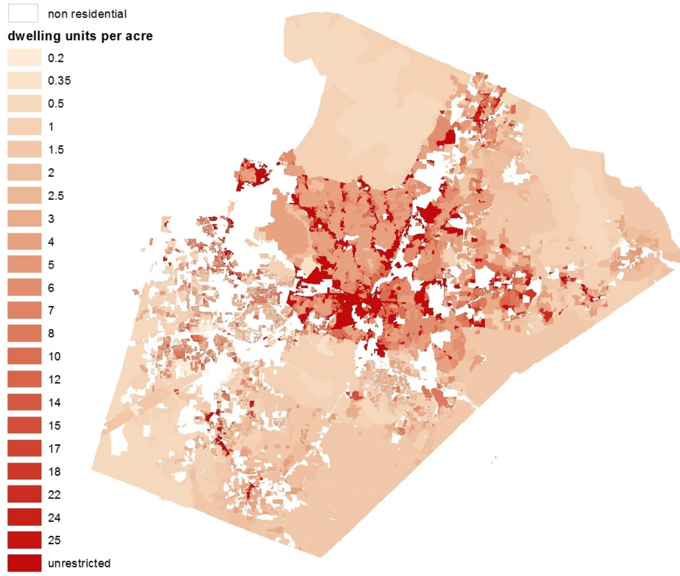
Table 1.2 shows summary statistics for the HMDA-matched Zillow sample. The table is split into an overall sample (columns 1 and 2) and then lots with a minimum lot size regulation of 1 acre (columns 3 and 4) and those with a minimum lot size regulation of one tenth of an acre (columns 5 and 6). Some of the sorting patterns that I study later are apparent in the raw data means. Overall, houses built in areas regulated for 1 acre lots are more expensive than those built in areas regulated for one tenth of an acre lots and the average sale price is around $261,000. Both the loan amount and household income are sorted in the same way, suggesting some of the sorting patterns that I find later. The average age of buildings in my sample is 24 years.

Minimum lot sizes are often measured in dwelling units per acre. The second row helps with this correspondence - if an area has a one acre minimum lot size it means that it allows one dwelling unit per acre of land, an area with a tenth of an acre minimum lot size allows ten dwelling units on one acre of land. The average minimum lot size is between a fifth and a quarter of an acre. Finally, I compute the ratio of the realized to the minimum lot size for each property in my sample. Table 1.2, row four shows the averages for this ratio. If the ratio is exactly one, the regulation can be considered binding and it is possible that given the possibility, households would prefer to build on smaller lots. On average the properties on one acre lots are very close to being binding (mean ratio of 1.07) while properties on a tenth of an acre lots are on average 1.5 times the minimum lot size (mean ratio of 1.49). Hence the regulation is more likely to be binding for larger minimum lot sizes. This does not have to be the case - it is possible that strictly regulated lots lie in undesirable areas with extremely cheap land where the regulation does not impose a constraint. These summary statistics provide suggestive evidence that the lot size regulation is more binding for stricter regulations.

Minimum lot size regulation data

Wake county provides geographic information on zoning regulations by lot. However, each municipality and unincorporated area in Wake County varies in their labeling of different regulations. After standardizing the lot size information for all the municipalities in Wake County as well as the unincorporated areas using the individual zoning codes, I am able to

Figure 1.1: Density regulations in Wake County, NC



Notes: Map of Wake County showing density regulations in dwelling units per acre. The corresponding minimum lot size is the inverse of the density, i.e. 2 dwelling units per acre correspond to a minimum lot size of half an acre. White areas are non-residential. The central city of Wake County is Raleigh.

merge information on the land use regulation with other information on lots to create a new data set with detailed geographic information on density regulations for the second most populous county in North Carolina which is also one of the fastest growing in the country over the last two decades.

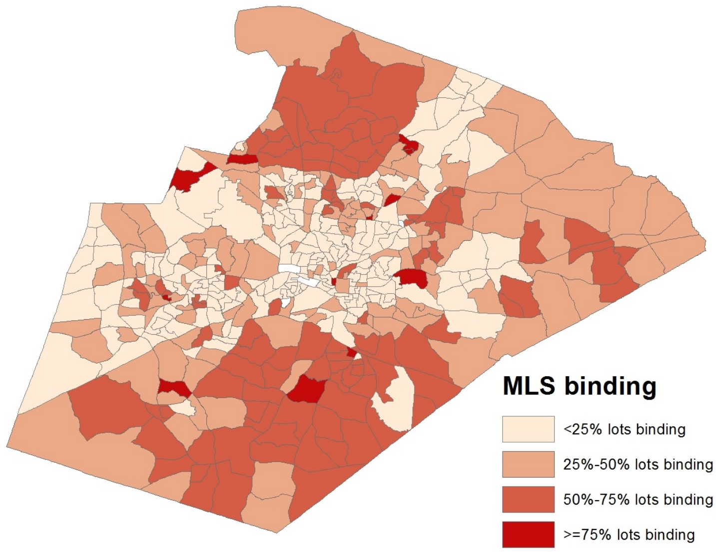
Figure 1.1 shows the distribution of density regulations across Wake County. Overall there is a fair amount of variation in the density regulation across Wake County. The darker the color, the more density is allowed and the smaller the lot sizes. White areas are non residential and hence not of interest here. The biggest block of density is the city of Raleigh but the surrounding areas also show pockets of more dense areas surrounded by less density. The variation is not just limited to the denser areas of Raleigh. There are centers of high allowed density even outside the main city of Raleigh in different suburban centers. Regulations range from 0.5 dwelling units per acre (du/a) to 25 dwelling units per acre. Outside of this range, there are some areas that are zoned with less than half a dwelling unit per acre. These are residential agricultural homes and I will drop these from my sample since our intuition for large lots being a restrictive practice does not apply in the same way to farmland. They make up a tiny proportion of my sample (about a tenth of a percent) and including these observations does not change any results. On the least regulated side, there are some areas in the inner city of Raleigh that are unrestricted in terms of their density. These will also not be included in my sample as these belong to the few areas in Wake County that allow apartment buildings. The transactions data does not include apartment buildings. Overall, less than 10% of residential land in Wake County allows apartment buildings. The average regulation in Wake County overall is four dwelling units per acre, or a minimum lot size of 0.25 acres.

To get a better understanding of what regulations can be considered strict, it is helpful to draw a comparison with the data that has been used in the past to study land use regulations. The Wharton Residential Land Use Regulation Index (WRLURI, Gyourko, Saiz and Summers, 2008) combines several aspects of land regulation into an index. The index is composed of the answers to questions such as the time to get certain building permits, the number of regulatory bodies whose approval is required for a project as well as the salience of minimum lot sizes. A municipality is coded as having restrictive minimum lot size regulations if there is any area within it that mandates a minimum lot size of 1 acre (amounting to a density regulation of 1 du/a). In my sample of lots a substantial 14% are regulated with either the most restrictive regulation of 1 dwelling unit per acre or an only slightly relaxed version of 1.5 dwelling units per acre.

Importantly, there is variation in density both within smaller geographic areas as well as across the county. If the strictness of the regulation was evenly spread throughout the entire county then the case for minimum lot sizes being a neighborhood sorting mechanism would be more tenuous. In reality however, there are large areas of Wake County that allow only very low density.

Figure 1.2 again shows a map of Wake County. The boundaries that are visible are

Figure 1.2: Bindingness of the regulation

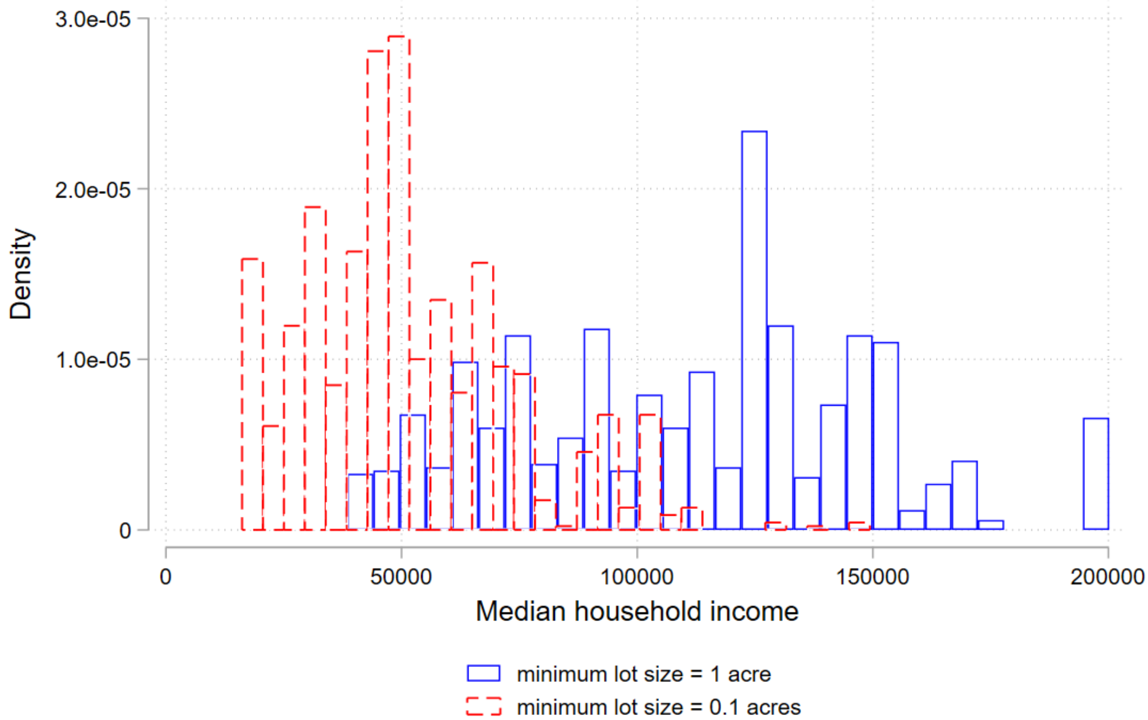


Notes: Map of Wake County showing the proportion of lots of exactly the minimum lot size (MLS) by Census block group.

Census block group boundaries. The map presents the proportion of lots in Wake County that are exactly at the minimum lot size, i.e. lots where the minimum lot size regulation binds. Since the regulation allows realized lots to be larger than the minimum lot size, the fraction of binding properties is a first indication of the degree to which households are constrained. There is variation in the fraction of lots at the minimum lot size. Overall, the lot size is more likely to be binding when the allowed density is lower and minimum lot sizes are larger (compare to Figure 1.1). One can see a high proportion of binding lots southwest of Raleigh, where the popular suburbs of Fuquay-Varina, Holly Springs and Cary lie. Another area with a large fraction of binding lots lies to the North of Raleigh where Wake Forest University is situated. Given that evidence on the bindingness of land use regulations is sparse, this information is useful. We can infer that the regulation binds quite widely, especially in areas that are in high demand.

Aggregating the minimum lot size regulation data at the Census block level and combining it with information on median household income from the Census illustrates

Figure 1.3: Income distribution by minimum lot size



Notes: Distribution of median household income in Census blocks regulated with 1 acre minimum lot sizes and Census blocks regulated with lot sizes of 0.1 acres. Median household income is taken from the 2013 ACS.

spatial inequality in income in differently regulated areas. Figure 1.3 shows the income distribution in Census blocks of Wake County that are regulated with a one acre minimum lot size versus a tenth of an acre minimum lot size. This does not control for amenities, so if stricter zoning is correlated with more desirable amenities and hence higher incomes then that is included in this histogram. Most strikingly, the overlap between both distributions is small. There is no Census block with median household income of less than $40,000 that is zoned with the most restrictive land use regulation and there is almost no block with median income above $110,000 that is zoned for a minimum lot size of 0.1 acres.

Location choice data

In the second part of the paper, I focus on the location choices of households with elementary school children to estimate a model of neighborhood choice that includes the minimum lot size as a constraint. The reasons for reducing the sample to this group are driven by current data availability as well as substantive reasons. In addition to considering other neighborhood amenities, parents face an important tradeoff between consumption and school quality in a given location. Particularly elementary schools with their small attendance areas can offer very little variation in housing options. Since schools are among the most important neighborhood amenities, parents of elementary school age children are likely to be quite sensitive to regulations on housing. They are therefore an interesting and key group to study in the context of sensitivity to minimum lot sizes.

To study the location decisions of households with elementary school age children, I use administrative school data from North Carolina (NCERDC). The advantage of using this data compared to the HMDA-matched Zillow data to estimate a model of neighborhood choice is two fold: First, the administrative school data allows me to study heterogeneity in preferences along more dimensions than the HMDA data which only contains the race of a mortgage applicant. The administrative school data contains, race, household size and free and reduced price lunch status. Second, this sample allows me to focus on a particularly interesting group of households. Since public education is one of the main reasons that neighborhoods are important for later life outcomes, understanding the trade off between consumption and neighborhood amenities that is imposed by minimum lot size regulations is especially important in the context of families that are choosing what school attendance area to locate in. I restrict the sample to elementary school children in grades 3-5 and their location choices in terms of Census block groups (this is the location definition available in the NCERDC) for the years 2010-2013. More recent data is available now that I am working on including in my sample.

Household income is imputed in the following way: I first impute income based on

Table 1.3: Sample of children from North Carolina administrative school data

	mean	sd	p25	p50	p75
Male	.51	.499			
Asian	.07	.26			
Black	.21	.41			
Hispanic	.14	.35			
White	.53	.5			
Number of kids	2.46	1.32	2	2	3
Free/reduced lunch	.32	.47			
Household income ($)	74,942	40,233	40,261	67,665	100,034
Observations	107,348				

Notes: This table shows summary statistics for the North Carolina Administrative School Data. The sample is composed of elementary school children (grades 3 to 5) in the years 2010 to 2013.

the free and reduced price lunch status, household size and the year. In a given year the free lunch status is determined in relation to the federal poverty line. If a child is eligible for either free or reduced price lunch, I impute the cutoff income for a given year for households of that particular size. For households that are not eligible for free or reduced price lunch, I impute income based on the Census block group of residence and the child's race. I use the American Community Survey that publishes median income by race for each block group by year. I drop households from the sample if the annualized cost of the cheapest available housing in the Census block group that they live in is larger than their annual imputed income. This affects less than 5% of the sample[9].

As a result of this imputation, I end up with types of households by free and reduced price lunch status, household size, race and home location. Table 1.3 shows summary statistics for the sample of children that I use. There is variation in race and free and reduced price lunch status as well as household size that I will exploit to study heterogeneity in preferences for neighborhood amenities. The imputed household income yields a median value of $67,665 which is lower than the median household income of Wake County (around $75,000). Given that this is a sample of households with young children as well as households that send their children to public school instead of private school, it is reasonable to expect a median

[9]The reason that I see such households is likely to be due to the fact that I only observe households for a short period of time. This may not adequately represent their financial situation at the time of purchasing the home. Alternatively, it is possible that household income misrepresents household wealth in these cases.

income that lies below the median of the county as a whole. The average income that I impute amounts to the median income in Wake County.

Other data sources

In addition to the above mentioned sources of data I also use data from four additional sources.

To exclude minimum lot size regulation boundaries that overlap with geographic and economic boundaries I obtain data on school attendance area boundaries over time from the Wake County Public School System. I use Wake County's open data portal to obtain spatial data on geographic boundaries such as highways, rivers and large streets.

I perform a version of the boundary discontinuity analysis using Census data. For this version I use Census block level data from the American Community Survey (2010-2013)[10] and Decennial Census 2010 to study changes across minimum lot size regulation boundaries. My level of observation for these analyses is either the Census block or the Census block group. To compute the distance to the boundary, I compute distance of the Census block or Census block group centroid to the nearest minimum lot size regulation boundary. The minimum lot size regulation is constant within all Census blocks and for more than 95% of Census block groups. For those Census block groups with multiple minimum lot sizes, I use the lowest minimum lot size to obtain a conservative estimate on sorting.

Finally, I obtain data on the price of land for Census Tracts in Wake County from 2012-2013 from Davis et al. (2019) .

[10]U.S. Census Bureau. (2013). 2010-2013 American Community Survey 3-year Public Use Microdata Samples. Retrieved from https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t

1.4 The effect of minimum lot sizes on neighborhood composition

What are the mechanisms through which the minimum lot size affects neighborhood composition? If the regulation is enforced and has an impact, then it should lead to larger lots in more tightly regulated areas. As a consequence, it might lead to larger constructed homes and higher property values, e.g. for a given price per square foot of land if the minimum lot size increases the overall land expenditure is higher. The higher property value in and of itself should lead to households of different incomes moving into differently regulated areas, and a marginal person of some level of income being priced out of the area with stricter regulation. In addition, if the regulation itself creates an amenity or households value having neighbors that live on regulated lots, there might be additional sorting due to this amenity.

To study the causal effect of minimum lot sizes on neighborhood composition, I conduct a boundary discontinuity design at the minimum lot size boundary. The goal is to isolate the effect of density regulations on different aspects of neighborhood composition, most importantly, neighborhood income but also ethnicity and poverty levels. The biggest concern with comparing neighborhood incomes in differently zoned areas is that both observed and unobserved amenities might differ systematically by density regulation and were taken into consideration when the regulation was originally set. For example, one might imagine that green, spacious neighborhoods with good air quality might also be more tightly regulated. In this case, simply running a regression of neighborhood income on regulation would not yield a causal estimate of the treatment coefficient, since amenities are at least partially driving the sorting on income.

To account for this endogeneity problem, it is crucial to compare lots that are differently regulated but have access to identical observed and unobserved neighborhood amenities. In the cleanest possible scenario, I would compare two houses on either side of a street, where the regulation boundary coincides with a neighborhood street, i.e. the house on one side of the street is regulated with a higher minimum lot size/ lower density and the house on the other side is regulated less strictly with a lower minimum lot size/ higher density. If the street separating the two houses is not a highway or a different type of divide that is hard to cross and likely to create very different neighborhoods on either side, then the two houses have the same access to all neighborhood amenities like parks, cafes, restaurants, etc. Similarly, they share the same neighbors. It also seems difficult to believe that the inhabitants of either house would be subject to different air quality or levels of pollution. It is unlikely that crime levels will differ on either side of the boundary.[11]

If the boundary also allocates the houses to different schools, then there is likely to be sorting based on the fact that houses lie in different school attendance areas. If one house lies within the attendance zone of a better school and higher income households have a higher willingness to pay for better education then part of the sorting on income at that boundary stems from school quality. I therefore exclude regulation boundaries that overlap with elementary school attendance zone boundaries[12]. Excluding school attendance area boundaries captures other boundaries that potentially confound the effect of the minimum lot size. Other local public goods such as police and fire stations or public libraries tend to have larger areas assigned to them. Therefore, school attendance boundaries run through areas that belong to the same public library or fire station. The majority of municipality boundaries are captured by a combination of highways and rivers as well as school attendance boundaries. I control for systematic differences by municipality at the remaining boundaries that coincide with municipal borders.

Lastly, idiosyncratic differences in unobserved amenities at particular boundaries are captured by boundary fixed effects. The only remaining difference between either side of the boundary then is the land use regulation, i.e. conditional on the running variable (distance

[11] I am currently searching for detailed data on crime and pollution that will allow me to test the continuity assumption. For now, I refer to the assumptions that have been made in the literature of boundary discontinuity designs which are similar to the assumptions I am making here.

[12] Since multiple elementary school boundaries feed into one middle or high school, focusing on elementary school attendance boundaries captures middle and high school attendance boundaries

to the boundary), observable amenities X_i as well as the boundary fixed effects it holds that: $\mathbb{E}(\epsilon_i|X_i = x)$ is continuous at the boundary $X_i = x^*$. Any remaining unobserved heterogeneity is either continuous at the boundary or a consequence of the difference in the minimum lot size regulation. There is no reason to believe that the preference for large houses should jump at the zoning boundary because larger houses can always be built in the less restricted area. If the preference for large houses is correlated with higher incomes, then it would be cheaper to build large houses in less restrictively zoned areas. However, if there is a preference for living near other households with larger lots, this is part of the selection effect. I will discuss this channel in more detail in Section 4.7.

While the two mechanisms outlined above both suggest a positive sorting effect at the minimum lot size boundary, with richer households sorting into neighborhoods with larger minimum lot sizes, one can also imagine cases in which there is no sorting or sorting is negative. If wealthier households prefer smaller houses, then this preference might be priced into the smaller houses (this might be true in the case of gentrifying inner cities that attract wealthy households to apartments and townhouses) and I might find a negative effect of stricter regulation on household income. This preference could jump at the boundary since smaller houses are not allowed on the more restrictive side. If there is no effect at the boundary this implies zoning does not put a binding constraint on households of any income and that all sorting effects are captured by variables other than land use regulation, for example observed and unobserved local amenities.

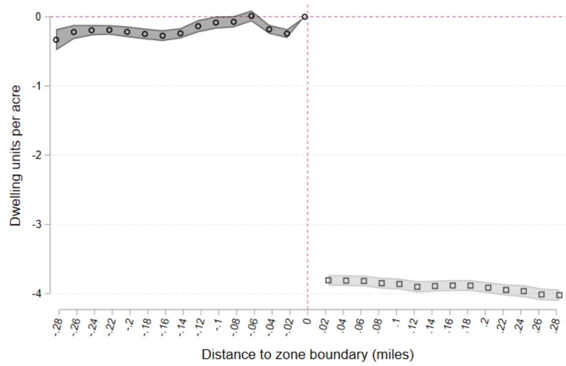
1.4.1 Graphical Analysis

I first follow Bayer, Ferreira and McMillan (2007) to construct graphs of the discontinuity at the minimum lot size boundary ((Bayer, Ferreira and McMillan, 2007) study school attendance area boundaries). For each lot I calculate the distance to the closest straight minimum lot size boundary segment that is not excluded. If the closest boundary is an excluded boundary, I drop that lot. I discuss the selection of the remaining boundaries in

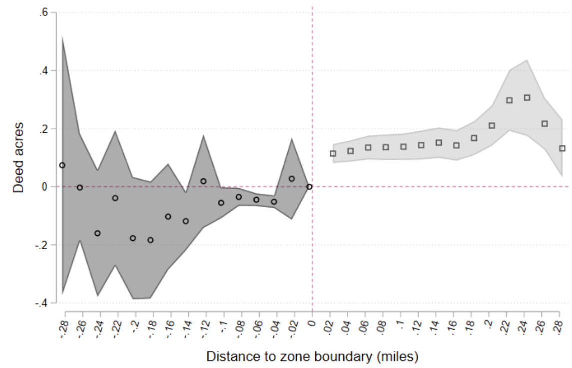
Appendix A. For each lot, I also define an indicator that is one if the lot is on the more strictly regulated side of its closest boundary compared to the other side, i.e. the indicator takes on the value of 1 if the lot is regulated to have a higher minimum lot size/ lower allowed density respectively relative to the other side of the boundary. Each graph is constructed from a regression of the outcome of interest on boundary fixed effects as well as bins of distances to the boundary (in either 0.01 or 0.02 mile steps). Negative values of the distance are for the less restrictive side of the boundary. Hence each point represents a conditional mean of the outcome variable of interest for a specific distance bin. The bin closest to the boundary on the less regulated side is normalized to 0 to avoid multicollinearity.

Figure 1.4 shows the plots of the conditional means for the most important variables of interest. The first thing to check is whether boundaries are being correctly assigned and whether the housing characteristics do in fact change at the minimum lot size boundary. If this is not the case then minimum lot sizes are not actually binding and there is no reason to believe sorting as a result of the regulation would occur.

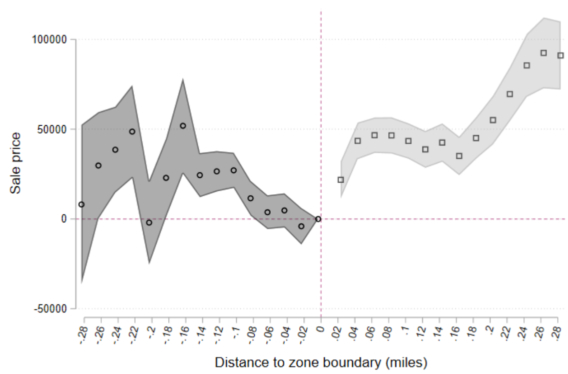
Figure 1.4a plots the conditional means for the regulation in dwelling units per acre. At the boundary there is a distinct jump downwards. This jump represents the average jump for the boundaries in the sample which is four dwelling units per acre. Figure 1.4b shows that there is a clear discontinuity in deed acres (the lot size) at the minimum lot size boundary. Lots on the more regulated side of the boundary are larger by about 45% at the mean (0.02 miles from the boundary on the less regulated side) indicating that minimum lot sizes are salient and change the characteristics of lots on either side of the boundary. Interestingly, the confidence intervals on the less regulated side are large which implies that there is a larger variety of lots on the less regulated side than the more regulated side of the boundary. This corresponds to the fact that the regulation specifically makes it impossible to build on small lots in regulated areas but does not restrict the building on large lots in less regulated areas. Figure 1.4c shows the sales price of the lot which is the sum of the value of the land and the value of the construction on that land. The jump amounts to $22,616, which is an increase of



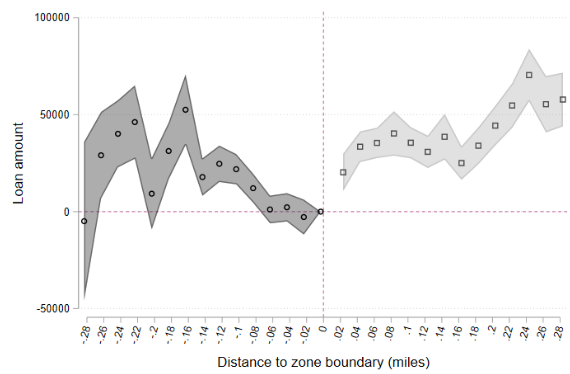
(a) Minimum lot size



(b) Deed acres



(c) Total sales price



(d) Loan amount

Figure 1.4: Graphical analysis - lot characteristics

Notes: Plots are created by regressing a dependent variable on a series of boundary fixed effects and bins of distance to the boundary (bins of 0.01 or 0.02 miles). Here I plot the the coefficients on the distance bins. Negative distances indicate the less regulated side of a boundary, positive distances indicate the more regulated side of the boundary. I normalize the bin closest to the boundary on the less regulated side (0 to -0.02 miles to the boundary) to zero. 95% confidence intervals are shown.

8.8% at the mean. This is one of the main channels through which minimum lot sizes seem to operate - the increase in the price of property then leads to households with different incomes sorting into differently zoned areas. Finally, Figure 1.4d shows a jump in the loan amount taken for properties on the more regulated side of the boundary of $20,820. At the mean, this jump amounts to 9.3% which corresponds reasonably well with the increase in sales price at the boundary. Similar to 1.4b both figures 1.4c and 1.4d demonstrate a higher variance in effects on the less regulated side of the boundary.

The graphs shown in Figure 1.5 are estimated using Census data at the census block level (Figures 1.5b) or HMDA matched Zillow data (Figures 1.5a and 1.5c).

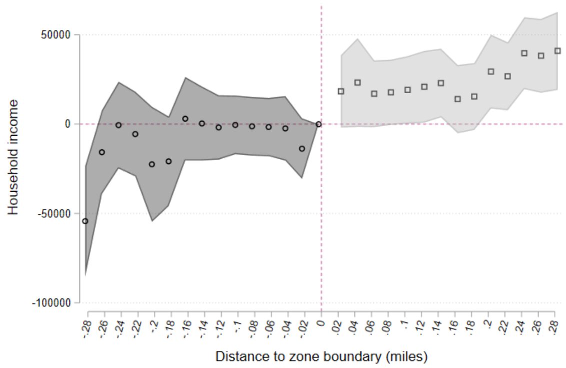
The first two figures consider aspects of income sorting along the minimum lot size boundary. Figure 1.5a depicts the conditional means of household income. There is a jump of $17,377 at the boundary, amounting to an increase of 15.4% relative to the mean. The confidence intervals suggest that this jump is not very precisely estimated. This could indicate that sorting on income occurs at some boundaries but not at others. I explore this possibility further in section 4.4. Another reason for this might be that the graphical analysis conflates boundaries with different types of regulations. Hence a regulation of 4 dwelling units per acre for example could be on the regulated side of one boundary and the unregulated side of another depending on the regulation on the other side. Figure 1.5b shows the effect on neighborhood poverty at the boundary. Data on the fraction of households with income below the poverty level by Census block group comes from the ACS. The variance on the unrestricted side is high while the variance of the coefficients on the restricted side is much lower. This is likely due to the fact that nothing restricts richer households from building large houses or living in more dense areas, but there is a restriction for less wealthy households in the more restricted areas. The restricted side does show coefficients that are almost always smaller than those on the less restricted side. The jump can be estimated to be a decrease in the fraction of households with income below the poverty level of 7.5%. This fact emphasizes that the jump is coming from the marginal person being priced out by the higher regulation rather than a complete income sorting occurring at this boundary.

Figure 1.5c considers the effect at the boundary on racial composition. The dependent variable here is an indicator for whether the main mortgage applicant is white. Again, similarly to the fraction of households with income below the poverty line there is a much bigger variance in race on the less regulated side of the boundary. However, the more regulated side of the boundary displays a 5.1% increase in the probability of the applicant being white. While there is clear sorting on race with households on the more regulated side of the boundary more likely to be white, the effect size is small relative to the other effects. One possible explanation is that sorting on race happens at a higher geographic level and that within the neighborhoods studied here, race is fairly homogeneous.

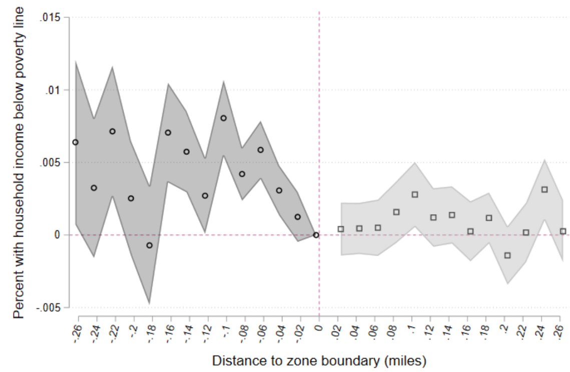
1.4.2 Estimation

The graphical analysis suggests that the difference in minimum lot size at the boundary generates a discontinuity in neighborhood composition by leading to different types of housing on either side. The effect of the regulation is to set a floor on the size of the lot - building on larger lots always remains possible. If amenities that are priced into housing values and neighbors are continuous at the regulation boundary, then the only discontinuity lies in the allowed building characteristics on either side which will influence housing prices and thereby have an effect on the incomes of households living there. If the regulation is not binding, rather than seeing housing characteristics change discontinuously at the boundary we would expect them to simply vary smoothly according to the preferences of households on either side.

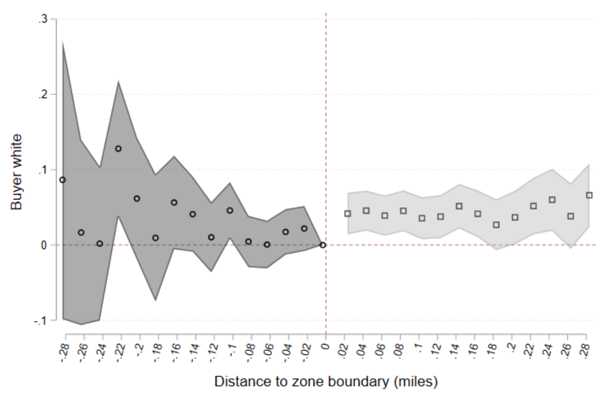
As mentioned above, the graphical analysis does not control for the difference in the regulation, conflating jumps of different sizes. To take this jump into account and estimate



(a) Household income (HMDA)



(b) % of households with income below poverty level (ACS)



(c) Buyer white (HMDA)

Figure 1.5: Graphical analysis - neighborhood composition

Notes: Plots are created by regressing a dependent variable on a series of boundary fixed effects and bins of distance to the boundary (bins of 0.01 or 0.02 miles). Here I plot the the coefficients on the distance bins. Negative distances indicate the less regulated side of a boundary, positive distances indicate the more regulated side of the boundary. I normalize the bin closest to the boundary on the less regulated side (0 to -0.02 miles to the boundary) to zero. 95% confidence intervals are shown.

the effect of a one unit change in the regulation, I use the following estimation strategy:

$$Y_h = \alpha_0 + \beta T_h + \alpha_1 dist_h + \alpha_2 dist_h \mathbb{1}\{\text{more regulated}_h\} + \alpha_3 dist_h^2 + \alpha_4 dist_h^2 \mathbb{1}\{\text{more regulated}_h\} + \gamma_{city(h)} + \theta_{bd(h)} + \epsilon_h \quad (1.1)$$

where h refers to one particular lot, Y_h is the outcome of interest at the lot level (household income, race, etc.), $dist_h$ is the distance in miles of lot h to nearest regulation boundary. The distance to the boundary is the running variable - I control for the running variable using a quadratic polynomial that I estimate separately on either side of the boundary (α_2 and α_4).

θ_{bd} is a set of boundary fixed effects. $T_h = \mathbb{1}\{\text{more regulated}_h\} * \Delta_{density(h)}$ is the treatment indicator, an interaction of the block being on the more regulated side of the nearest boundary, $\mathbb{1}\{\text{more regulated}_h\}$, and $\Delta_{density(h)}$ - the difference between the density regulation on either side of the boundary. The difference in the density is defined as: $\Delta_{density(h)} = |density_{\text{more regulated}} - density_{\text{less regulated}}|$.

γ_{city} is a set of city level fixed effects. There are 11 cities in Wake County apart from Raleigh and it is possible that lots in these cities differ systematically in their neighborhood characteristics (there could be richer cities, or more ethnically diverse cities for example) but not for reasons related to land use regulation. It is possible that some cities might have more job opportunities for example. There could be more reasons like this that will lead to different compositions of neighborhoods in some cities for reasons unrelated to zoning, hence I control for systematic differences by city. The fixed effect also captures systematic differences in the level of land use regulation by city.

The outcomes of interest Y_h that I examine are the lot level household income, race of the person taking out the mortgage, mortgage loan amount and property sales price. Appendix B additionally shows results at the Census block level that include neighborhood variables such as the % of households with income below the poverty level. The interpretation of β is the effect on Y_h of a decrease in one dwelling unit per acre of mandated density.

1.4.3 Threats to identification

The main assumption of the regression discontinuity approach is that:

$$\mathbb{E}(\epsilon_b | X_b = x) \text{ is continuous at } X_b = x^* \tag{1.2}$$

Here, this translates to ϵ_b being continuous at the regulation boundary. There are several possible threats to identification related to the boundary discontinuity estimation strategy.

One worry relates to the fact that these boundaries were drawn several decades ago, starting in the 1930s. One might either worry that they are no longer relevant today since they were set so long ago or that what is being picked up today is not the effect of the regulation but rather the equilibrium sorting that occurred over decades. In response to the first concern, there is literature about redlining boundaries that show that those boundaries, though drawn in the 1930s have persistent effects on neighborhoods today (Aaronson, Hartley and Mazumder (2017), Appel and Nickerson (2017)). Additionally, the regulation boundary of course still matters for actual regulation today while redlining practices are now banned.

The second issue - what is actually being picked up by this boundary - is a more serious concern. If the regulation does not bind, then it is unclear what effect is picked up at the boundary. I do robustness to consider boundaries where the regulation is salient and compare them to other boundaries where the regulation is not salient. In particular, I focus on neighborhoods that are new relative to the introduction of zoning and also neighborhoods where I observe a high proportion of binding regulations at the lot level.

The next concern pertains to the exogeneity of the zoning boundaries. Again, one may be worried both about the historical exogeneity but also about exogeneity today. In terms of historical exogeneity I can point to the fact that density regulations were officially imposed to regulate the growth of cities, i.e. which areas would be allowed to grow and which would not. If minimum lot sizes separate systematically different qualities of land then that would be a concern since it would affect housing prices and quality by itself but not because of the actual restriction on the lot size[13]. If, historically, zoning was imposed to separate already different neighborhoods further, then examining newer neighborhoods and the bindingness of zoning today would address some of those concerns. Close to the boundary, neighborhood composition should be affecting either side equally and therefore not be generating differences in incomes of households on either side because of differences in neighbors.

Regarding exogeneity today, the media may create the perception that battles against rezoning to higher densities are often fought and that home owner's associations are often successful in excluding less wealthy neighbors through zoning. Examining the rezoning history of Wake County, these types of concerns do not seem of great importance even if they might be occurring in other places around the country. On average, the entirety of Wake County receives about 30 rezoning requests per year, of which a little more than half are not relevant for neighborhood sorting because they affect commercial and not residential property. Among the remaining 12-15 rezoning requests, about half are either withdrawn or denied. There remain on average 5-6 residential rezoning cases that are successful and will affect usually around 20 acres of land at most. In contrast to popular perception, it does not seem that rezoning cases for higher density are routinely denied. Initiating a rezoning request is also costly, $1,000 per case, and might not be as easily affordable to all.

Since a difference in regulation will necessarily change the observable characteristics of the neighborhood, it is difficult to show continuity at the boundary for observable characteristics. For my approach, it is most important that factors influencing sorting based on income are continuous so that the only remaining factor that households sort on is the density regulation. The most important factors in this context are amenities. There is a large literature studying the willingness to pay for amenities and neighborhood sorting. The most important amenity that will be discontinuous at boundaries is public education. I exclude regulation boundaries

[13] I am in the process of compiling data on soil quality at the appropriate level to check this assumption.

that overlap with public elementary school attendance zone boundaries. This also excludes municipality boundaries that might offer different levels of public goods. In contrast to Turner, Haughwout and van der Klaauw (2014) I am examining much lower-level boundaries. I also exclude boundaries that are likely to separate neighborhoods, i.e. highways, major roads, streams and rivers. It stands to reason that the boundaries that remain after all these exclusions might systematically vary in some ways.

Idiosyncracies at a particular boundary with respect to unobservable amenities, are captured by the boundary fixed effect. If there remains anything that is systematically correlated with a particular level of regulation but has nothing to do with the regulation then this would bias the estimate of β. If the assumption of continuity holds, then I am estimating the effect of only the difference in allowed density on neighborhood sorting independently of other factors that we know households sort on.

1.4.4 Results and contextualization

Table 1.4 shows the results of estimating equation 1.1 where the dependent variable is the household income of the household living in a given lot. It is helpful to consider when a density regulation is relevant and when it is not: The regulation imposes a floor on lot size. If all lots are larger than the minimum allowed lot size, then one might argue that the constraint is not binding for this neighborhood since all lots exceed the minimum lot size anyway. Consequently, we would expect to see less sorting at this boundary, if the regulation is imposing no constraint to begin with.

For each lot, I calculate the ratio of the actual lot size to the minimum lot size: $\frac{\text{actual lot size}}{\text{minimum lot size}}$. When $\frac{\text{actual lot size}}{\text{minimum lot size}} = 1$, the regulation is binding, when $\frac{\text{actual lot size}}{\text{minimum lot size}} > 1$ it is not. For $\frac{\text{actual lot size}}{\text{minimum lot size}} < 1$ the regulation is not being enforced. I focus on boundaries with different proportions of lots with a binding lot size and compare the results to a sample of boundaries where a lower proportion of lots is binding. If it is the minimum lot size that leads to sorting and not other more permanent features of the neighborhood sorting equilibrium

Table 1.4: Household income

	(1) full sample	(2) full sample	(3) full sample	(4) 25% binding	(5) not binding	(6) 25% binding	(7) not binding
$\mathbb{1}\{\text{regulated}\} * \Delta_{density}$	1916.0*	1927.5	1493.5	3812.0*	1608.0	4131.1**	1081.4
	(1160.8)	(1172.2)	(1257.9)	(2039.5)	(1397.3)	(2075.4)	(1539.0)
distance to boundary	full	≤0.5mi	≤0.2mi	≤0.5mi	≤0.5mi	≤0.2mi	≤0.2mi
boundary fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
municipality fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	43031	42975	40760	11530	31445	11185	29575

Standard errors in parentheses clustered by minimum lot size polygon, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is household income in 2010 dollars from the Home Mortgage Disclosure Act merged with house transaction data from Zillow. All regressions include boundary fixed effects and municipality fixed effects. I vary the distance to the boundary. Columns 4 and 6 focus on boundaries where at least 25% of lots are binding compared to boundaries where less than 25% of lots are binding. $\mathbb{1}\{\text{regulated}\}$ is an indicator that is 1 if a lot is on the relatively more regulated side of its boundary. $\Delta_{density}$ is the absolute value of the difference between the density regulation on the more regulated side of the boundary and the density regulation on the less regulated side of the boundary. $\mathbb{1}\{\text{regulated}\} * \Delta_{density}$ can be interpreted as the change in household income for a decrease in one dwelling unit per acre of allowed density.

then sorting at binding minimum lot size boundaries should be stronger than at non-binding boundaries.

Table 1.4 uses individual lot-level data for Wake County to perform the boundary discontinuity. I exclude from the sample properties that are zoned for agricultural homes, to focus completely on the urban and suburban context. Table 1.4 column 1 estimates β for the overall sample without any restrictions. The treatment coefficient can be interpreted as estimating the difference in income for a decrease in one dwelling unit per acre of allowed density relative to the other side of the boundary. The overall effect suggests an increase in neighborhood income by nearly $2,000 for a decrease in one dwelling unit per acre on the more restricted side of the boundary. Columns 2 and 3 decrease the distance to the boundary and focus only on lots that are at most 0.5 or 0.2 miles away from the boundary. Overall, the smaller the bandwidth, our intuition predicts the more the effect of regulation should be picked up, because the observations on either side become more similar to each other as the distance to the boundary decreases. In this case however, it seems that decreasing the distance to the boundary slightly lowers the size of the effect. In the case of the sample of lots within 0.2 miles to the boundary the effect is not precisely estimated. These results reflect that there are many boundaries where the regulation exists but is not salient. At those boundaries, it is not clear that the effect being captured is a consequence of the minimum lot size regulation changing. However, part of the effect of the regulation is its intent to treat effect which encompasses boundaries at which there is no sorting.

To focus on boundaries where the regulation is binding, columns 4 and 6 vary the distance to the boundary for boundaries where at least 25% of lots are binding on the more regulated side. The effect size in column 4 doubles to an increase in income of nearly $4,000 for a decrease in the density regulation by one dwelling unit per acre. Now, decreasing the bandwidth to 0.2 miles around the boundary increases both the size and the precision of the estimated effect. My preferred estimate is presented in column 6 and shows an increase in neighborhood income of $4,131 for a decrease in one dwelling unit per acre.

Further support that these effects are indeed capturing the effect of the minimum lot size regulation is given by the estimates in columns 5 and 7. These specifications focus on boundaries where less than 25% of lots are binding. The standard errors are large and the overall size of the effect is less than $2,000 for a decrease in one dwelling unit per acre. This supports the intuition that the effect of sorting should be larger at boundaries where the minimum lot size is more binding. Compared to the first three columns it is interesting that decreasing the distance to the boundary only increases the size of the treatment effect at boundaries that have binding regulations and not in the overall sample. This again gives rise to confidence that when the regulation is salient, the correct sorting effect is being captured by the boundary discontinuity design. When the regulation is not salient, it is not clear how to interpret the coefficient.

Table 1.B-2 shows results from a similar analysis that uses Census block centroids as units of observation instead of individual lots. The results are qualitatively similar but overall, the effect size using individual lots is much larger than the effect size in the analysis at the census block level. The fact that this more detailed data results in larger effects indicates

that average Census block group incomes mask a lot of the variation at regulation boundaries and emphasizes the importance of the HMDA-matched lot level income information.

It is worthwhile to put these effects into context of existing results on sorting in the literature. One of the few existing results on income sorting into neighborhoods comes from Bayer, Ferreira and McMillan (2007). They find sorting on income 0.2 miles from the school attendance zone boundary of $2,800. Comparing this to the results that I find for the same distance (column 5 of Table 1.B-2), a decrease in one dwelling units per acre leads to a similar amount of sorting. Remembering that the density regulations go up to 24 du/a, this suggests that imposing minimum lot sizes can easily lead to similar and greater sorting effects than sorting that occurs due to different preferences of households for local public goods. The most common size of the jump at the boundary is either 3 dwelling units per acre. Extrapolating the coefficients linearly implies sorting on income of $12,300 for a decrease in three dwelling units per acre which is already larger than the sorting found by Bayer, Ferreira and McMillan (2007). This comparison is the most direct comparison I can make.[14]

I can summarize that at a boundary where one side allows one dwelling unit per acre less than the other, sorting on income is about $4,000. The size of this effect suggests that zoning alone could impose a large constraint on the neighborhood choice of households with at least median household income ($74,000 in Wake County). Certainly, this effect would not matter much for high-income households choosing where to live. Imagine now, a possible difference in Wake County, going from 1 du/a to 10 du/a. About 5% of my sample lie at boundaries with this size of a jump. The boundary predicts sorting on income at such a boundary of $37,179 (9*$4,131). In areas where the zoning is unrestricted, median household income is about $51,000. Median household income in Wake County overall is

[14] One can also somewhat compare my sorting results to Black (1999). She finds that the capitalization into house prices of 5% higher test scores is $3,950. This can be compared to a decrease in density by one dwelling unit per acre. Finally, one can compare my findings to Banzhaf and Mangum (2018) to some extent. They find that given the same amenities, ticket prices for neighborhoods (measured through property values) are 65 cents higher for a unit increase in the Wharton Residential Land Use Regulatory Index. It is difficult to compare these results to what I find but the WRLURI is usually standardized, so a one unit increase would mean a one standard deviation increase. In that sense the ticket price is quite small once amenities are taken into account.

Table 1.5: Other borrower/property characteristics

	(1) Loan amount	(2) Loan amount	(3) White	(4) White	(5) Sale price	(6) Sale price	(7) Deed acres	(8) Deed acres
$\mathbb{1}\{\text{regulated}\} * \Delta_{density}$	6504.8***	6373.9***	0.00813	0.00827	5781.8*	5955.5*	0.0478**	0.0478**
	(2269.7)	(2268.4)	(0.00810)	(0.00841)	(3186.9)	(3165.8)	(0.0237)	(0.0231)
Distance to boundary	full	≤0.2mi	full	≤0.2mi	full	≤0.2mi	full	≤0.2mi
Distance	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
Boundary fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12277	11905	9159	8853	12277	11905	12277	11905

Standard errors in parentheses clustered by minimum lot size polygon,* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are the loan amount in 2010 dollars taken by owners of each property (columns 1 and 2), the race of the main borrower (columns 3 and 4), the total sales price of a property in 2010 dollars (columns 5 and 6) and the actual size of the deed in acres (columns 7 and 8) from the Home Mortgage Disclosure Act merged with house transaction data from Zillow. All regressions include boundary fixed effects and municipality fixed effects. I vary the distance to the boundary. Columns 2,4,6 and 8 focus on boundaries where at least 25% of lots are binding compared to boundaries where less than 25% of lots are binding. $\mathbb{1}\{\text{regulated}\}$ is an indicator that is 1 if a lot is on the relatively more regulated side of its boundary. $\Delta_{density}$ is the absolute value of the difference between the density regulation on the more regulated side of the boundary and the density regulation on the less regulated side of the boundary. $\mathbb{1}\{\text{regulated}\} * \Delta_{density}$ can be interpreted as the change in dependent variable for a decrease in one dwelling unit per acre of allowed density.

around $74,000. If there were only two school districts and one of them was regulated by 1 du/a and the other by 10 du/a then simply the zoning would exclude the left tail of the income distribution from living in the strictly regulated area.

1.4.5 Other neighborhood characteristics

Table 1.5 additionally looks at other characteristics of the property and borrower. Here, I again restrict the sample to boundaries where at least 25% of properties are at the minimum lot size. Columns 1 and 2 show that the loan amount taken increase by about 3% relative to the mean for a difference in density regulations by one dwelling unit per acre at the boundary. This happens for an average increase in deeded acreage of 0.05 acres (columns 7 and 8) which is a 17% increase relative to the mean and an increase in sales price of 2.7% relative to the mean (columns 3 and 4). As in Table 1.B-3 there is no precisely estimated indication of a change in the probability of the borrower being white, even though the size of the effect amounts to a 1% increase relative to the mean. Overall, there is strong evidence that households are sorting on income and income-related characteristics at the minimum lot size boundary. On the other hand, there is no strong indication of land use regulations being a salient boundary for sorting by race. A possible confounder of the boundary discontinuity design is that land use regulation boundaries might overlap with former redlining boundaries[15]. If this were the case then land use regulations would be separating neighborhoods that were already very different from each other for reasons unrelated to land use regulations. The lack of large and precisely estimated effects of sorting on race at the minimum lot size boundary is reassuring evidence that the treatment effect here is not picking up the effects of a policy such as redlining which as been shown to generate persistently different neighborhoods till today.

Table 1.B-3 discusses the differences in Census block level characteristics across the land use regulation boundary. Table 1.C-4 considers a different dependent variable, namely test scores, and studies the composition of households in terms of the test scores of their children on either side of a minimum lot size boundary.

1.4.6 Age of property

Table 1.6 considers properties of different ages (constructed before 1990, before 2000 or after 2010). If the regulation was initially not exogenous or did not apply to older buildings then the previously shown effects would be biased. In particular, if the regulation were endogenous effect sizes in Table 1.4 would be overstating the true effect and if older and smaller properties were grandfathered into the regulation then the results in Table 1.4 would be biased downwards. Table 1.6 splits the sample by properties of different ages. Since most of the regulations were finalized in the 1970s and 1980s, focusing on houses built more recently will capture the effect on neighborhood sorting for houses that definitely were constructed after the minimum lot size regulations were put in place and hence took the regulation as

[15] The term redlining refers to racially motivated practices in the 1930s in the United States that denied applicants from specific, mostly African American neighborhoods, mortgages to purchase homes.

Table 1.6: Age of property

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		before 1990			before 2000			after 2010	
	all	25% binding	not binding	all	25% binding	not binding	all	25% binding	not binding
$\mathbb{1}\{\text{regulated}\} * \Delta_{density}$	1481.0	1423.9	1152.5	1343.1	3037.9	311.1	7298.1**	13142.3	9452.2**
	(1156.9)	(1909.6)	(1685.5)	(1009.0)	(2096.7)	(1342.8)	(3233.6)	(8452.3)	(4282.6)
distance to boundary	≤0.5mi	≤0.2mi	≤0.2mi	≤0.5mi	≤0.2mi	≤0.2mi	≤0.5mi	≤0.2mi	≤0.2mi
distance	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic
boundary fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
municipality fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21315	5147	15471	28730	7713	20007	4620	910	3117

Standard errors in parentheses clustered by minimum lot size polygon, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is household income in 2010 dollars from the Home Mortgage Disclosure Act merged with house transaction data from Zillow. All regressions include boundary fixed effects and municipality fixed effects. I vary the distance to the boundary. Columns 1-3 focus on houses built before 1990, columns 4-6 on houses built before 2000 and Columns 7-9 on houses built after 2010. Columns 2,5 and 8 focus on boundaries where at least 25% of lots are binding compared to boundaries where less than 25% of lots are binding. $\mathbb{1}\{\text{regulated}\}$ is an indicator that is 1 if a lot is on the relatively more regulated side of its boundary. $\Delta_{density}$ is the absolute value of the difference between the density regulation on the more regulated side of the boundary and the density regulation on the less regulated side of the boundary. $\mathbb{1}\{\text{regulated}\} * \Delta_{density}$ can be interpreted as the change in household income for a decrease in one dwelling unit per acre of allowed density.

given.

As before, I vary both the distance to the boundary as well as the degree to which lots are binding at the minimum lot size (at least 25% of properties binding or not). In the time period before 1990 sorting is always positive with effect sizes around \$1,500 but large standard errors. This might be an indication that minimum lot sizes were less binding for these older buildings. Including properties built between 1990 and 2000 (columns 4-6) yields similar effect sizes as before but the effects are estimated with increased precision. Binding boundaries show even higher effects, around \$3,000 of sorting but no statistical significance. Finally, sorting for properties built after 2010 (columns 7-9) is large (\$7,298) and precisely estimated. The pattern across these variations is as before, boundaries bordering binding boundaries show a higher degree of sorting and focusing on narrower bands around the boundary leads to larger effects. Focusing on boundaries with a newer housing stock seems to yield larger sorting effects, perhaps due to the fact that those newer buildings were built taking the regulation to be exogenous.

1.4.7 "Neighborhood effect" of minimum lot sizes

Next, I address the issue that households right at the minimum lot size border might be different from households on the interior. This is possible if households have a preference for the type of neighbors that live in highly regulated areas or if they have a preference for amenities created by the density regulation itself. In either case households on the interior are more exposed to this effect because all their neighbors are tightly regulated, whereas right at the boundary only half the neighbors are tightly regulated and the other half are less strictly regulated. I call this effect the "neighborhood effect" of minimum lot sizes. If households indeed have a preference for regulated neighbors, then the sorting effect at the boundary is an underestimate of the overall effect of the minimum lot size. The overall effect is the sum of the sorting effect at the boundary as well as the additional effect on the interior of the boundary.

I follow a similar strategy as Turner, Haughwout and van der Klaauw (2014) to estimate this effect. The effect that is estimated at the boundary is the effect of the difference in the regulation on sorting on income, however on either side of the boundary, households have both some strictly regulated and some less strictly regulated neighbors. Consequently, the neighborhood effect does not change at the boundary, only the primary effect of the regulation does. Comparing a boundary parcel to an interior parcel regulated with the same minimum lot size isolates the neighborhood effect - the difference between those two parcels is how their neighbors are regulated - while excluding the sorting effect into that particular type of regulation.

Let $\mathbb{1}\{\text{interior parcel}\}$ be an indicator for lot h being on the interior of a boundary. I vary the distance at which I define an interior parcel. McConnell and Walls (2005) find that the value of open space such as might be caused by a low density regulation decreases quickly with distance, ranging from a block to a mile away. I therefore define an boundary parcel as lying at most 0.2 miles away from the boundary and experiment with the defintion of an interior parcel. The treatment effect,α_1, is given by the interaction of the interior dummy with

Table 1.7: Neighborhood effect of the minimum lot size

	(1) Income	(2) Income	(3) Income	(4) Income	(5) Income	(6) Income
$\mathbb{1}\{\text{Interior parcel}\} * \Delta_{Density}$	602.4***	405.3**	363.7**	707.4***	444.3***	410.3***
	(133.8)	(133.6)	(135.1)	(119.0)	(120.7)	(121.1)
Boundary parcel (miles from boundary)	$0 < h \leq 0.2$	$0 < h \leq 0.2$	$0 < h \leq 0.2$	$0 < h \leq 0.2$	$0 < h \leq 0.2$	$0 < h \leq 0.2$
Interior parcel (miles from boundary)	$0.3 < h \leq 0.4$	$0.3 < h \leq 0.4$	$0.3 < h \leq 0.4$	$0.3 < h \leq 0.6$	$0.3 < h \leq 0.6$	$0.3 < h \leq 0.6$
Boundary f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Municipality f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Parcel controls	No	No	Yes	No	No	Yes
Interior parcel indicator	No	Yes	Yes	No	Yes	Yes
Observations	56641	56641	56641	57720	57720	57720

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The dependent variable is household income in 2010 dollars from the Home Mortgage Disclosure Act merged with house transaction data from Zillow. All regressions include boundary fixed effects and municipality fixed effects. I vary the distance to the boundary. Columns 4 and 6 focus on boundaries where at least 25% of lots are binding compared to boundaries where less than 25% of lots are binding. $\mathbb{1}\{\text{Interior parcel}\}$ is an indicator that is 1 if a lot is located a certain distance away from the boundary. Row three indicates how this distance is varied. $\Delta_{density}$ is the absolute value of the difference between the density regulation on the more regulated side of the boundary and the density regulation on the less regulated side of the boundary. $\mathbb{1}\{\text{Interior parcel}\} * \Delta_{density}$ can be interpreted as the difference in household income on the interior compared to the boundary for a decrease in one dwelling unit per acre of allowed density. Parcel controls include the age of the house and the deeded acreage.

the change in density at the boundary which can be interpreted as the change in the income of households living on interior parcels compared to those at the boundary for a difference in the density of one dwelling unit per acre ($\Delta_{density} = density_{\text{less regulated}} - density_{\text{more regulated}}$).

$$Y_h = \alpha_0 + \alpha_1 \mathbb{1}\{\text{interior parcel}\} * \Delta_{density} + \alpha_2 \mathbb{1}\{\text{interior parcel}\} + X_h'\gamma + \gamma_{city} + \theta_{bd} + \epsilon_h \tag{1.3}$$

θ_{bd} is a set of boundary fixed effects and X_h is a vector of lot characteristics (i.e. lot size, age of house). It is important to control for house characteristics here since those are not reflective of neighbors but rather of decisions that the home owners themselves make.

Table 1.7 shows the results of the estimation. All estimates lie between \$350 and \$700 for a decrease in density by one dwelling unit per acre. Columns 3 and 6 include the indicator for interior parcels separately along with parcel controls, hence these estimates are preferred. The definition of the interior parcel in miles from the boundary does not change the treatment coefficient much. The additional effect of the regulation from having regulated neighbors is about $400. This amounts to about one tenth of the overall effect estimated at the boundary. It therefore seems that for the most part, the effect of the minimum lot size is to sort households into differently regulated neighborhoods. In addition, there is an effect of about a tenth of the size of the primary effect that can be attributed to a preference for having other regulated neighbors. This finding is important for the modeling decisions that I make in the following sections.

1.5 A model of neighborhood choice with minimum lot sizes

1.5.1 Towards a rezoning counterfactual

The boundary discontinuity shows that minimum lot sizes are an important mechanism that cause households to sort into different neighborhoods and that the effects are large compared to other sorting effects in the literature. Current policy debates discuss different ways in which land use regulations could be relaxed in the future in either existing neighborhoods[16] or how vacant land that is about to be developed should be regulated [17]. The main reasons for relaxing zoning constraints are to create more and more affordable housing options. This includes the possibility of giving households the opportunity to move from a less desirable neighborhood that has relaxed regulations to a more desirable rezoned neighborhood. The trade-off between consumption (which is affected by the regulation) and amenities is important

[16] One such example is from Minneapolis. In December 2018, the City Council of Minneapolis passed the Minneapolis 2040 plan (https://minneapolis2040.com/) that includes a policy that seeks to end zoning of neighborhoods for single family houses exclusively. The city hopes to thereby increase the supply of housing and offer a wider variety of more affordable housing options.

[17] 20% of land in Wake County is vacant. The Raleigh Comprehensive Plan Update 2030 which considers future land use in the area seeks to zone vacant land in a way to "Accommodate growth in newly developing or redeveloping areas of the city through mixed-use neighborhoods with a variety of housing types."

to learn about which households are affected by this policy on the margin as well as the distribution of willingness to pay for different types of households.

The boundary discontinuity approach is crucial to precisely isolate the causal effect of the minimum lot sizes on neighborhood composition which is challenging to do with broad comparisons of locations. In order to estimate the causal effect it relies on boundaries at which amenities are continuous. As a result the estimates are limited to the regulation's effect right at the boundary and it is not possible to study the trade-off between consumption and amenities. Therefore, from the boundary discontinuity we learn about the extent to which location choices are distorted but not how spatial sorting is broadly affected across areas with different amenities and regulations. It is likely that relaxing the regulation in an area causes equilibrium price effects which depend on the desirability of the new neighborhood as well as the supply of land. The boundary discontinuity estimates sorting in partial equilibrium but cannot capture price effects in the counterfactual scenario. To take a first step towards answering these important policy questions, I develop a static equilibrium neighborhood choice model that focuses on the trade off between amenities and consumption. I also use this model to compare and to inform the difference between the model predictions and the boundary discontinuity estimates.

The main feature this model captures is how minimum lot sizes contribute to sorting of households into different neighborhoods. The model hones in on the first order mechanism of the minimum lot size regulation, namely distorting neighborhood choice by constraining the choice of housing. Households of different incomes sort into neighborhoods due to housing expenditures that are in part determined by the minimum lot size imposing a floor on land consumption. Households trade off consumption and available neighborhood amenities in different neighborhoods. I follow the structure of existing urban sorting models (Diamond and McQuade, 2019; Galiani, Murphy and Pantano, 2015; Bayer, Ferreira and McMillan, 2007; Diamond, 2016; Colas and Hutchinson, 2019) that model locations as separate goods characterized by their amenities as well as the consumption that households can enjoy. Most previous models of neighborhood choice incorporate the possibility that a household of any income may live in an expensive neighborhood by consuming a tiny amount of the housing good. By incorporating minimum lot sizes into this sorting model, this is no longer an option. Further, this model nests previously chosen specifications of preferences over consumption and housing. However, by allowing for preferences that can be non-homothetic, sorting into neighborhoods based on income can arise naturally in this model through income effects that the estimated curvature in consumption generates. To study equilibrium effects resulting from policy changes, prices are endogenized in the counterfactual and are an important determinant of welfare consequences.

The model is designed to focus on the primary channel through which the minimum lot size operates - mandating a minimum quantity of land that must be purchased. As such it does not include certain aspects of the minimum lot size. First, it does not endogenize amenities created by the minimum lot size. Second, it takes the minimum lot size regulation as given. Rationalizing the existence of zoning regulations has been the subject of past work, however the distortionary effects remain mostly unstudied. In the future it may be interesting to bring together both the benefits and costs to consider an optimal level of land use regulation. Finally, the model follows large parts of the neighborhood choice literature and is static. This implies that a house does not play a role as an asset and households can instead be considered renters of their dwellings. The model captures the behavior of households that are constrained in their location choice by the minimum lot size. These households are the population targeted in this paper and in policy debates. Inspite of the choices I make, the model fits sorting patterns by income and preferences for locations.

1.5.2 Setup

The model is a static neighborhood choice model. Households choose a neighborhood and at the same time choose the size of their lot (land consumption) and non-land consumption within a neighborhood. They face two constraints: a budget constraint and the fact that the lot size they pick has to be at least the minimum lot size. The fact that the regulation affects the size of the lot and not the size of the house motivates this formulation of the utility function in which households choose lot size and consumption. I discuss the consequences of this choice below. Each neighborhood is characterized by a level of neighborhood amenities, a minimum lot size and a set of available lot sizes.

Households, indexed by i, living in neighborhood n have preferences over non-land consumption C_{in}, the size of their lot L_{in} (in square feet), and neighborhood amenities, δ_n. The amenities are non rival – each person receives the same amount independent of how large a lot they purchase. Each households belongs to an observable type, τ. τ is defined by the observable characteristics in the data, in particular number of children and ethnicity. While δ_n is the valuation of neighborhood amenities that is common to all households, ψ^z represents heterogeneity in preferences for a vector of neighborhood characteristics Z_n by types. Each household also receives an unobserved taste shock, ϵ_{in}, in each neighborhood which is assumed to be distributed i.i.d Type I extreme value. The variance of ϵ_{in} is fixed which normalizes the scale of utility. Households' utility function is given by:

$$u_{in} = \alpha \frac{C_{in}^{1-\theta_c} - 1}{1-\theta_c} + \beta \frac{L_{in}^{1-\theta_l} - 1}{1-\theta_l} + Z_n' \psi^z \tau_i + \delta_n + \epsilon_{in} \tag{1.4}$$

Households have CRRA preferences over land and non-land consumption. α governs how households trade off overall consumption and neighborhood amenities, β scales land consumption relative to non-land consumption. θ_c and θ_h are the parameters of risk aversion for non-land consumption and land respectively. Z_n is a vector containing characteristics of the neighborhood n.[18] τ_i is a vector of demographic characteristics of household i, i.e. number of children and ethnicity. ψ^z allows preferences for neighborhood amenities to differ by characteristics of households. Since minimum lot sizes enforce a minimum land consumption per neighborhood, the fact that consumption would be very low or negative for some households

[18]In practice, I choose the average end of grade test score in the local public school, the poverty rate in the neighborhood and average travel time to work.

in strictly regulated neighborhoods drives households of different incomes to sort into different neighborhoods. CRRA preferences are not homothetic, so the optimal quantities are not just determined by the price of land relative to the price of consumption.

Households simultaneously pick a neighborhood, n, to live in along with housing consumption in square feet, L_n, and non-housing consumption, C_{in}. Since this is a static model the choice of neighborhood is assumed to be a once and for all choice. While this behavior does not correspond to the reality it approximates the behavior of a household with children picking a neighborhood to live in for the duration of their children's school education. Households maximize utility subject to two constraints: The first is a budget constraint where Y_i is household i's household income (which does not vary across neighborhoods) and r_n is the price of a square foot of land in neighborhood n.

$$Y_i = C_{in} + r_n L_{in} \tag{1.5}$$

The second constraint introduces minimum lot sizes into the model. All neighborhoods are regulated by a minimum lot size, $\underline{L}_n$, i.e. every house in n must be built on a lot of at least $\underline{L}_n$. The role that the minimum lot size plays is twofold: firstly it can force higher than optimal consumption of land and secondly it can change the set of feasible locations.Rather than lot size being a continuous choices, within each neighborhood, households pick from a discrete set of l available lot sizes. This is in accordance with the fact that in reality, unless there is vacant land available, households cannot freely pick any lot size but must pick from the available lots in a neighborhood.

$$L_{in} \in L_n = [L_{1n}, L_{2n}, ..., L_{ln}] \text{ and } L_{in} \geq \underline{L}_n \tag{1.6}$$

The CRRA preferences over housing and other consumption with different curvature parameters do not allow for a closed-form representation of optimal quantities of C_{in} and L_{in}. If the optimal quantity of housing is smaller than the minimum lot size, the minimum lot size has to be chosen. Consumption is the residual income after expenditure on land.

$$L^*_{in} = \max\left\{\underline{\mathrm{L}}_n, L_{in}(Y_i, r_n)\right\}$$

$$C^*_{in} = Y_i - r_n L^*_{in}$$

Households optimize by picking the location that maximizes their indirect utility while simultaneously picking the optimal lot size and non-land consumption in that neighborhood. The choice to focus on lot size as opposed to housing services more generally is explained in the context of the land use regulation. Minimum lot sizes directly affect the size of the lot but they do not impose a restriction on the size of the construction on that lot[19] This paper focuses on the minimum lot size, so I isolate land consumption from housing consumption. Underlying this modeling strategy is an assumption that markets are thick in the characteristics of actual properties so that households can find their preferred combination of lot size and construction. This assumption is supported by the data from Wake County which shows a high degree of variability of the house by lot size.

The indirect utility associated with each location can be written in the following way:

$$v_{in} = \alpha \frac{C^{*\ 1-\theta_c}_{in}}{1-\theta_c} + \beta \frac{L^{*\ 1-\theta_l}_{in}}{1-\theta_l} + Z'_n \psi^z \tau_i + \delta_n + \epsilon_{in} \tag{1.7}$$

Under the distributional assumption about the unobserved idiosyncratic preferences, the probability that household i chooses location n can be written as follows:

$$\pi_{in} = \frac{exp\Big(\alpha \frac{C^{*\ 1-\theta_c}_{in}-1}{1-\theta_c} + \beta \frac{L^{*\ 1-\theta_l}_{in}-1}{1-\theta_l} + Z'_n \psi^z \tau_i + \delta_n\Big)}{\sum_{k\in\Omega_i}\Big(exp\Big(\alpha \frac{C^{*\ 1-\theta_c}_{ik}-1}{1-\theta_c} + \beta \frac{L^{*\ 1-\theta_l}_{ik}-1}{1-\theta_l} + Z'_k \psi^z \tau_i + \delta_k\Big)\Big)} \tag{1.8}$$

Ω_i represents the set of feasible locations for households i. I define Ω_i based on i's household income and the price of land such that a household cannot pick to live in a location

[19]While there exist regulations that can directly affect for example the ratio of the property area to the lot size (floor to area ratio) this type of regulation is not present in Wake County.

where the price of every lot is higher than i's household income.

1.5.3 Identification

This section discusses how the model parameters are identified. While the identification of the mean neighborhood utilities (δ_n) and characteristic-specific preferences (ψ^z) follows standard arguments, identification of the remaining parameters α, β, θ_c and θ_l is more subtle. The aim of the model is to study the primary way in which the minimum lot size affects neighborhood choice and perform welfare analysis for counterfactuals in which neighborhoods with desirable amenities receive relaxed zoning regulations. If households do not value neighborhood amenities highly relative to consumption (high value of α), then relaxing minimum lot sizes in tightly regulated areas may have little effect because higher aggregate consumption can be achieved more cheaply in areas with worse amenities. If households do value neighborhood amenities highly then we expect poorer households to choose neighborhoods with good public schools or other neighborhood amenities when minimum lot sizes are relaxed. Observing locations with different degrees of bindingness of the minimum lot size, aids with identification of the parameters related to the marginal utility of consumption, α and θ_c. These parameters are typically difficult to identify in the context of neighborhood choice models.

Finding separate variation in neighborhoods in lot size (or more broadly in housing choice) and consumption can be quite difficult. Consequently, disentangling preferences for lot size from the curvature of the utility in non-land consumption is challenging. The minimum lot sizes offers this variation - neighborhoods with high bindingness of the regulation. In these neighborhoods there is essentially only variation in consumption but no variation in housing, thus helping to identify parameters related to the marginal utility of consumption (θ_c, α).

θ_c governs the curvature of consumption and α determines the trade off between consumption and neighborhood amenities. The larger the value of θ_c, the more curved utility is and the less sensitive households are with respect to consumption. In neighborhoods where the minimum lot size binds for all lots, i.e. there is only one type of lot size, then the variation in the rate at which households of different incomes choose the same neighborhood identifies the curvature parameter θ_c. Intuitively, if households of different incomes choose the same neighborhood at similar rates, given that there is a fixed expenditure per neighborhood, this will mean that the curvature of consumption is high (θ_c large).

After θ_c has been pinned down, α is identified by the importance of consumption relative to the error term and mean neighborhood utilities. A high value of α indicates that consumption is important compared to neighborhood amenities and the error term. In this case we expect to see sorting of households by income into neighborhoods, with richer households choosing more expensive neighborhoods. The degree of bunching at the minimum lot size, while not strictly needed, helps identify α. In this model, bunching at the minimum lot size indicates that there are households that prefer a smaller lot size but still choose this neighborhood because they enjoy high utility from neighborhood amenities or a high idiosyncratic shock. Hence, it carries information about how much households care about neighborhood amenities compared to aggregate consumption. When there is variation in lot size within neighborhood, α and θ_c are identified holding fixed the size of the lot while comparing how households of different incomes pick the same neighborhood.

The curvature parameter on the lot size θ_l is identified from variation in how households of different incomes sort into different lot sizes for different prices of land. Within a neighborhood, one can compare two households with different incomes and study how a pure income effect changes the chosen lot size. If the relative choice probabilities net of consumption differences are small, then we can infer that households are not very sensitive to lot size, i.e. θ_l is large.[20] The other source of variation that identifies θ_l stems from variation in lot size chosen

[20]When there is no variation in lot size within a neighborhood, it is difficult to identify β and θ_h. Since there is no variation in lot size within neighborhood, the only way to identify preference for lot size is by making comparisons across neighborhoods. One can think of varying income and looking at the difference in lot size while holding fixed the price of a square foot of housing. However it is not possible to distinguish between a preference for desirable neighborhood amenities and a preference for larger houses if nicer neighborhoods tend to mandate larger houses. Hence identifying preferences for housing requires variation

for households of the same income that choose to locate in different neighborhoods with different prices of land. The stronger substitution is towards housing when the price is lower, the smaller θ_l is going to be and the stronger preferences households have for larger lots. Once θ_l has been pinned down, β is identified from the relative importance of lot size to aggregate consumption and the error term.

Finally, identification of the mean neighborhood utilities (δ_n) and characteristic-specific preferences for neighborhood amenities (ψ^z) parameters follows standard arguments as in Berry (1994), Berry, Levinsohn and Pakes (2004), Bayer, Ferreira and McMillan (2007) and Bayer, Keohane and Timmins (2009). The δ_n parameters capture the mean utility by location which is common to all households. They are identified from the shares of households that choose each neighborhood such that the predicted location shares of the model match the actual shares.

1.5.4 Estimation

The estimation of the model follows standard methods in the estimation of discrete location choice models as in for example Diamond (2016), Bayer, Ferreira and McMillan (2007) or Galiani, Murphy and Pantano (2015). These papers follow methods from the empirical industrial organization literature described in Berry, Levinsohn and Pakes (2004) and Berry, Levinsohn and Pakes (1995). The approach is often used to estimate preferences for aspects of products that are endogenous to characteristics unobserved by the researcher. The focus of this paper is on understanding how important overall neighborhood amenities are compared to consumption in the presence of a regulation that limits consumption in attractive neighborhoods. Since none of the counterfactuals I perform require knowing separate preferences for different amenities I shall not separately estimate preference parameters for specific local amenities.

I follow the neighborhood choice literature to estimate the vector neighborhood fixed effects $\boldsymbol{\delta}$ as well as the parameter vector $\boldsymbol{\theta} \equiv (\alpha, \theta_c, \beta, \theta_l, \boldsymbol{\psi}^{\boldsymbol{z}})$ using maximum likelihood.

within neighborhoods in how households of different incomes choose the size of their lot.

The common utility from one neighborhood is normalized to 0. Given the assumptions on the error term and the form of the choice probabilities the likelihood function is given by:

$$L(\boldsymbol{\theta}, \boldsymbol{\delta}) = \sum_{i=1}^{N} \sum_{n=1}^{J} c_{in} ln(\pi_{in}(\boldsymbol{\theta}, \boldsymbol{\delta})) \tag{1.9}$$

where c_{in} is an indicator that is one if household i chooses location n and π_{in} is the probability that household i chooses neighborhood n. In my context the vector $\boldsymbol{\delta}$ is 427 long (the number of locations, i.e. number of Census block groups in Wake County). Computationally, searching over 427 + dim(θ) parameters is challenging. Berry, Levinsohn and Pakes (1995) showed that one can simplify computation by recognizing that for each vector of $\boldsymbol{\theta}$there is a unique vector of $\boldsymbol{\delta}$ such that the model's predicted location shares match the observed shares. This vector $\boldsymbol{\delta}$ is the unique fixed point of the following contraction mapping:

$$\delta_n^{l+1} = \delta_n^l + [log(\pi_{in}) - log(\pi_{in}(\boldsymbol{\theta}, \boldsymbol{\delta}))] \tag{1.10}$$

Lot size is picked from the discrete set of available lot sizes in each location. To predict the choice of housing, for each guess of the vector of parameters $\boldsymbol{\theta}$ I first determine the optimal house choice of each individual given these parameters in all neighborhoods. These choices are then fed into the maximum likelihood procedure. A future draft of this paper will consider a version model where there is a discrete choice both over the neighborhood and the lot size. In a typical neighborhood there are around 3-6 different lot sizes to pick from. Given 427 neighborhoods, this significantly expands the space of choices and makes computation much more costly. I therefore begin with this simpler version of the model.

The price of a neighborhood is endogenous to unobserved neighborhood amenities. This could potentially bias the parameters determining the marginal utility of consumption, α and θ_c. Both unobserved and observed aspects of the neighborhood that are common to all households are captured within δ_n. In addition to that, preferences over certain neighborhood characteristics vary with observable characteristics of households (ψ^z). Under the assumption that this heterogeneity captures preferences for neighborhood characteristics, the parameters estimates of α and θ_c are unbiased. If there are unobserved differences in neighborhood preferences by type of person, then these may lead to an over- or underestimate of sensitivity to consumption. For example, if after taking into account common preferences and heterogeneous preferences by number of children for some neighborhood amenities, there remain systematic differences in preferences for other neighborhood aspects between households with more and less children, then sensitivity to consumption would be downward biased if those aspects are also correlated with high land prices. This issue can be partly improved upon by estimating common neighborhood utilities by observable types (i.e. $\delta_n^{\text{one child}}$, $\delta_n^{\text{two children}}$) which will be part of a future version of this paper.

1.5.5 Results

Table 1.8 shows the parameter estimates for the model with exception of the mean neighborhood utilities. The first four parameters are the utility weights on land and non-land consumption as well as the curvature on both types of consumption. Compared to the curvature in non-land consumption, the curvature in land consumption (in square feet) is higher. For a household with consumption net of land expenditure of \$70,000, these parameters imply that the marginal utility of consumption is lower than the marginal utility of land if that household is living on a lot that is one tenth of an acre large. In terms of housing, an interesting aspect is that due to the discreteness of housing options, it is possible for households to choose a lot size where the marginal utility of land is above the marginal utility of consumption if the next available lot size is much larger. The fact that the curvature on land is larger than the curvature on consumption implies that land expenditure as a fraction of household income decreases with income, i.e. lower income households spend a larger fraction of their income on land than higher income households. The parameter estimates imply that as income doubles from \$25,000 to \$50,000 households will optimally choose a lot size that is about 30% larger and as income doubles from \$50,000 to \$100,000 households would optimally pick a lot size that is 32% larger if lot size choice were continuous.

The lower half of the table shows heterogeneity in neighborhood amenities for different types of households. The neighborhood fixed effect δ_n represents the common utility for white households with only one child (reference group). I pick three important characteristics of the neighborhood by which I allow preferences for different types to vary: school quality, i.e. the average end of grade math test score for 3rd to 5th grade exams of the school in a neighborhood, the poverty rate in that neighborhood (Census block group) and average travel time to work - which is taken from the ACS and captures a similar effect as distance to the Central Business District. Compared to the reference group, a higher poverty rate is uniformly less preferred by all other types of households, but there is variation in how much different types value school quality as well as travel time relative to the baseline. This is a very parsimonious specification of heterogeneity in preferences and it can be extended further in the future.

Since showing the 427 fixed effects in a list is not feasible, I summarize my findings graphically. Figure 1.6 illustrates the estimates of the common neighborhood utilities for some different types of households. Locations that remain white are not in the sample because they are entirely non-residential areas. The gradient is created such that neighborhoods that are colored in the same way are in the same decile of the mean neighborhood utility for this demographic group. Figure 1.6a shows the mean neighborhood utilities ($\boldsymbol{\delta}$) for households with one child (averaged across different preferences for race). This graph lends credibility to the estimated parameters: Overall, the map tells the story of households preferring the amenities in suburbs around the inner city of Raleigh. Suburbs to the West and Southwest of Raleigh are particularly popular, which corresponds to the reality in Wake County, with the municipalities of Fuquay-Varina, Cary and Holly Springs being especially demanded areas to live. On the other hand, suburbs in the East and Northeast of Raleigh tend to be poorer and more rural and are also estimated to be less preferred in terms of their amenities.[21]

[21] To understand the correlation between the estimated neighborhood fixed effect δ and neighborhood characteristics, I run a linear regression. The results are shown in Appendix Table 1.F-5. A higher

Table 1.8: Parameter estimates

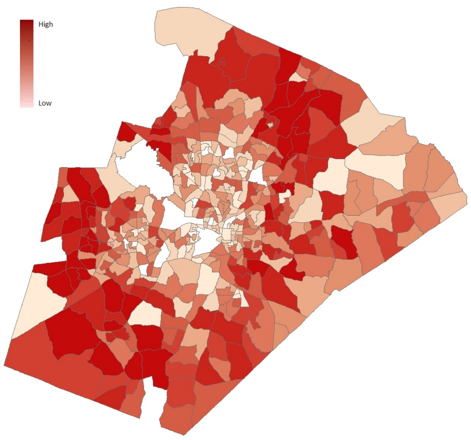
Consumption	
Weight on consumption utility (x10^{-2}), α	1.277
(x10^{-2})	(0.041)
Coefficient of relative risk aversion, θ_c	7.143
	(1.91)
Weight on lot size utility (x10^{-4}), β	7.155
(x10^{-12})	(5.41)
Curvature in lot size, θ_l	18.094
(x10^{-6})	(4.00)
Type specific heterogeneity, ψ^z	
$\mathbb{1}$[two children] x school quality	0.002
	(22.606)
$\mathbb{1}$[two children] x poverty rate	-0.05
	(0.054)
$\mathbb{1}$[two children] x travel time	0.004
	(2.986)
$\mathbb{1}$[$\geq$ three children] x school quality	0.001
	(18.648)
$\mathbb{1}$[$\geq$ three children] x poverty rate	-0.116
	(0.069)
$\mathbb{1}$[$\geq$ three children] x travel time	0.018
	(2.620)
$\mathbb{1}$[asian] x school quality	0.017
	(9.590)
$\mathbb{1}$[asian] x poverty rate	-0.022
	(0.016)
$\mathbb{1}$[asian] x travel time	-0.049
	(0.902)
$\mathbb{1}$[black] x school quality	-0.006
	(12.610)
$\mathbb{1}$[black] x poverty rate	-0.041
	(0.057)
$\mathbb{1}$[black] x travel time	0.009
	(1.818)
$\mathbb{1}$[hispanic] x school quality	-0.005
	(10.247)
$\mathbb{1}$[hispanic] x poverty rate	-0.002
	(0.050)
$\mathbb{1}$[hispanic] x travel time	0.052
	(1.612)
Observations	107,348

Notes: This table shows model parameters estimated using maximum likelihood estimation. Standard errors in parentheses. The estimation uses data on public elementary school children (grades 3-5) from North Carolina Administrative School Records (2010-2013) combined with information on available lot sizes by neighborhood in Wake County. School quality is measured as the average end of grade math score for grades 3-5, poverty rate is the fraction of households with income below the poverty line by Census block group from the American Community Survey and travel time is average travel time to work in minutes by Census block group from the American Community Survey.

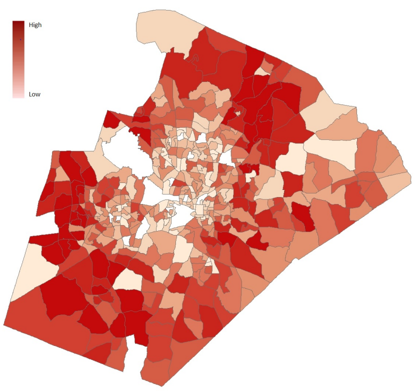
Figures 1.6b and 1.6c show the same map with the valuations for households with two children and three or more children respectively. The biggest differences for households with more children arise in the valuation for the periphery as well as the inner cities - it seems that in comparison to households with one child, households with two or more children have a higher valuation for neighborhoods that lie either on the outskirts of Wake County or further in the inner city.

Figure 1.6d and 1.6e show the map holding fixed the amenity value and varying the income for a household with one child. Figure 1.6d plots deciles of the utility for households with an income of \$20,000 and Figure 1.6e plots the utility for households with an income of \$75,000 (median income in Wake County) respectively. For households with an income of \$20,000 the most preferred locations lie in the East of Wake County. Those locations tend to be places with low cost of land and housing as well as low neighborhood amenities (compare to Figure 1.6a). It therefore seems that for households with low income, consumption is important enough to outweigh the preference for attractive neighborhood amenities. For households with median income (Figure 1.6e) preferred locations shift to the South and Southwest of Wake County reflecting the fact that the floor on land consumption imposes less of a constraint on this population and they prefer living in suburbs with desirable amenities. The fact that inner city Raleigh locations are not highly preferred could be for two possible reasons: The price per square foot of land in the inner city is higher than in the suburbs. So purchasing larger lots is quite expensive and can be done more cheaply outside the inner city. At the same time, the amenities of inner city areas are not highly valued to households so overall these locations are not very attractive. The second reason might be that the undesirability of inner city neighborhoods is particularly true for the sample that I use to estimate the model - households with elementary school age children. If schools are valued highly as a local amenity, this makes the neighborhoods in the center of Raleigh even

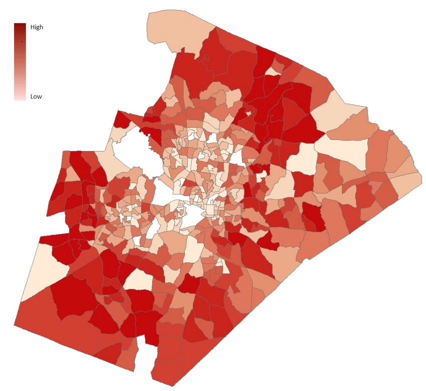
neighborhood fixed effect is consistently associated with higher performing schools and a lower fraction of households with incomes below the poverty level and a lower price of land conditional on the other factors. In addition, there is a slight indication that travel time to work is longer in neighborhoods with a higher amenity value, suggesting that these neighborhoods often lie in the suburbs.



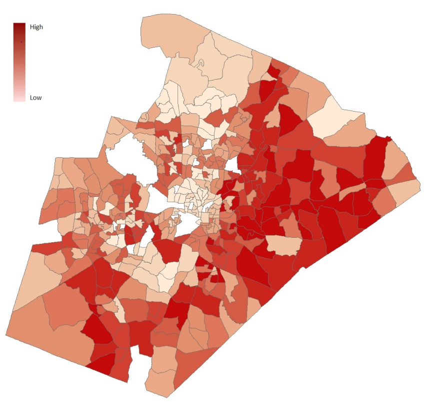
(a) Households with one child (only neighborhood valuation, $\boldsymbol{\delta}$)



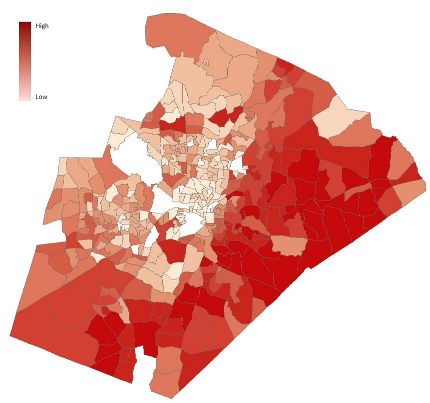
(b) Households with two children (only neighborhood valuation, $\boldsymbol{\delta}$)



(c) Households with three or more children (only neighborhood valuation, $\boldsymbol{\delta}$)



(d) Households with one child and household income of $20,000



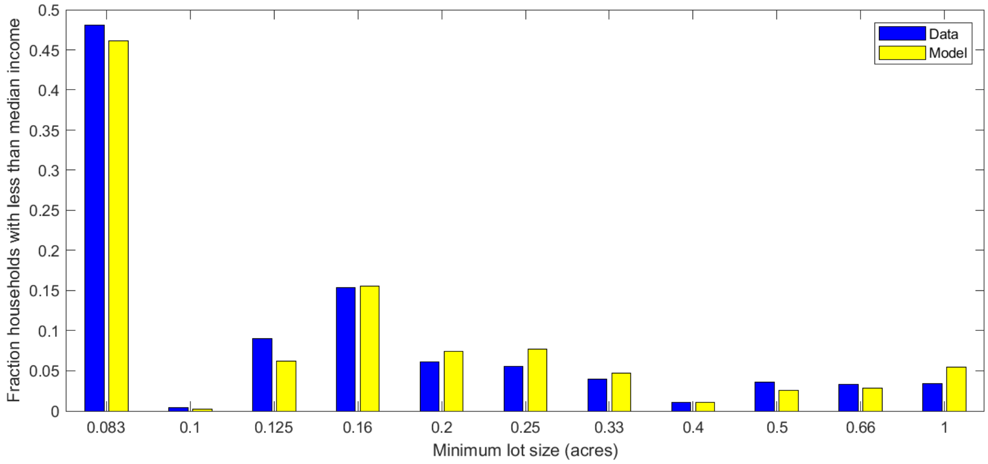
(e) Households with one child and household income of $75,000

Figure 1.6: Deciles of utility

Maps a), b) and c) plot only the estimated neighborhood fixed effect in deciles for households of different sizes. Maps d) and e) consider households with one child and plot deciles of the utility values for different incomes.

less desirable.

Figure 1.7: Fraction of households with less than median income that choose to locate in neighborhoods zoned with different minimum lot sizes



Notes: Fractions of households income below median income in Wake County ($75,000) that choose to locate in differently regulated neighborhoods. Figure compares the model's predictions with the raw data fractions from the administrative school data.

Another important aspect of the model is to predict the shares of households that live in differently regulated neighborhoods, particularly for income constrained households. Figure 1.7 shows the models predictions and the shares in the data for households with income below the median income. For households with less than median income, the model predicts the shares in differently regulated areas very well.

1.6 Relaxing minimum lot size regulations

Having estimated the parameters of the model, I can conduct counterfactuals in which I change the minimum lot size and study the resulting changes in neighborhood sorting, the price of land as well as implications for welfare. In particular, I conduct two counterfactuals: The first counterfactual considers a relaxation of the minimum lot size in one particular neighborhood with amenities at the 75th percentile of the estimated neighborhood fixed effect in Wake County. The second counterfactual considers the zoning regulations for vacant land. Imagine that there is a new neighborhood which is attached to either a high- or a low-amenity neighborhood and which will either be zoned with strict or relaxed regulations. I study prices and average neighborhood income in this counterfactual neighborhood.

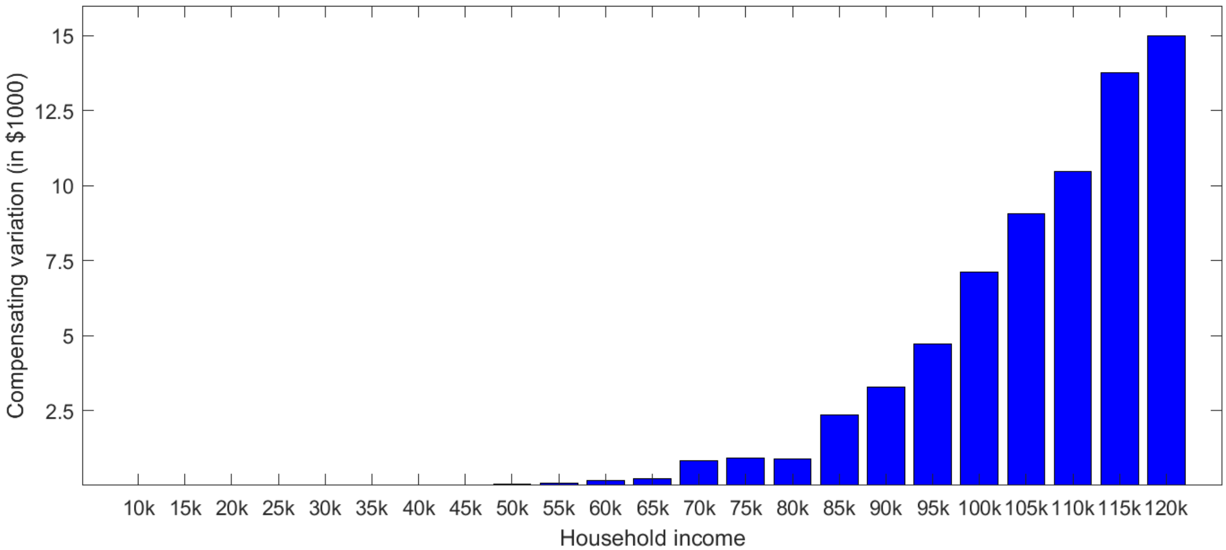
To close the model, I use the fact that land supply of a given neighborhood is fixed. Using this fact, in each counterfactual I re-compute the land price per square foot in each neighborhood such that demand for land as estimated by the neighborhood choice model equals the available supply of land in each neighborhood. Given that my choice data are the choices of households with children in elementary school this can be interpreted as rezoning the land that is occupied by such households and resolving for prices on these pieces of land.

1.6.1 Lowering the minimum lot size in one neighborhood

The first counterfactual deals with a change in the existing neighborhood regulation. Changing the minimum lot size regulation for the entire county is very likely to lead to sizable equilibrium effects as households might resort across the entire county thereby significantly changing who lives where. In addition, considering changing a current neighborhood raises the question of how existing buildings will be treated. A dynamic treatment of this question seems inherently more important in that context. Since I have specified a static neighborhood demand model without endogenous amenities I instead reduce the minimum lot size in one particular neighborhood. Rather than thinking of this as a change of the existing neighborhood one can imagine a counterfactual county that is identical to Wake County in all respects except that there is this slight change in the zoning regulation in one neighborhood. The counterfactual is informative about sorting in the alternative county.

In particular, I change the minimum lot size ($\underline{\mathrm{L}}_n$) from half an acre to a tenth of an acre in an existing neighborhood whose value of the neighborhood fixed effect is at the 75th percentile. I compute land prices in all neighborhoods such that the demand for land equals the supply of land. Since allowing a smaller subdivision of land without changing the number of inhabitants in the county is comparable to an increase in possible housing supply, it is

Figure 1.8: Compensating variation by income (high-amenity neighborhoods)



Notes: Figure shows the compensating variation by household income for a counterfactual policy in which the minimum lot size is held at its current level everywhere except for one neighborhood. The minimum lot size is reduced from 0.5 acres to 0.1 acres in one neighborhood with neighborhood amenities at the 75th percentile. I calculate the compensating variation required to make households indifferent between the counterfactual and actual scenarios.

not surprising that the price per square foot of land (r_n) falls in the counterfactual for the neighborhood in which the regulation is being changed. I compute the compensating variation for households of different incomes and characteristics by finding the reduction in income that would make them indifferent between the current scenario and the counterfactual in which one neighborhood has relaxed zoning regulations. The compensating variation, CV_i, is given by[22]:

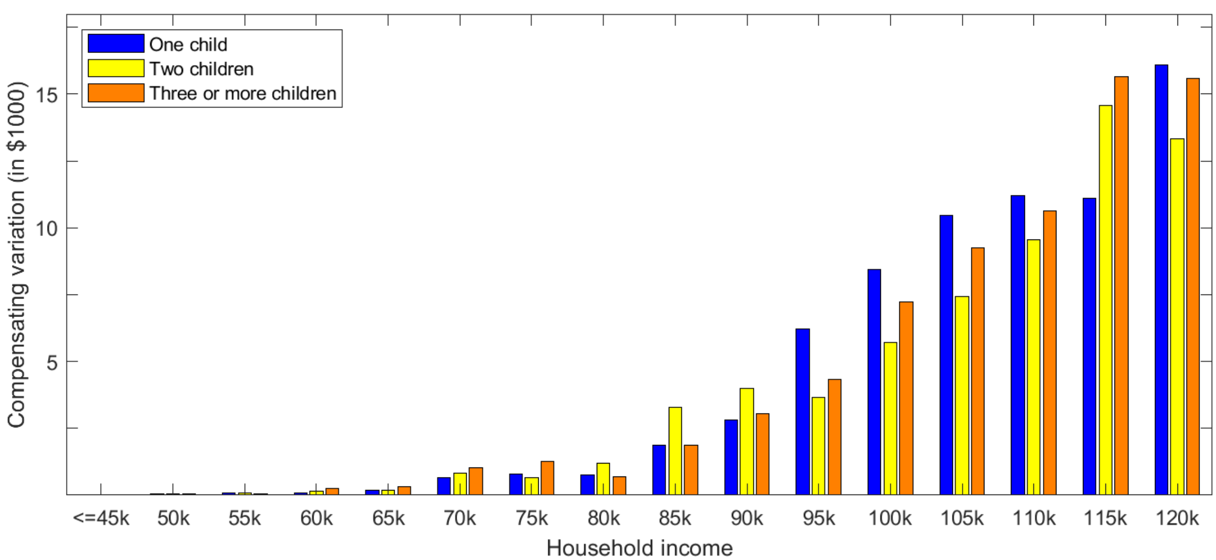
$$\begin{aligned}\mathbb{E}\Big[\max_n\{v_{in}(Y_i, r_n, \underline{\mathrm{L}}_n)\}\Big] = \\ \mathbb{E}\Big[\max_n\{v_{in}^{counterfactual}(Y_i - CV_i, r_{counterfactual}, \underline{\mathrm{L}}_{counterfactual})\}\Big]\end{aligned} \tag{1.11}$$

Figure 1.8 shows the compensating variation averaged over observable characteristics

[22] Here I take the expectation over the values of the idiosyncratic shock, so the counterfactuals give the compensating variation before the location-specific shocks are drawn.

for households with different levels of household income. There are no households that are worse off in the counterfactual which follows from the fact that overall prices are similar or slightly lower and that amenity levels do not change. Households with incomes below $45,000 do not value the counterfactual scenario since they can still find land that is cheaper and allows even smaller lots in other neighborhoods and neighborhood amenities are valued less than consumption. Households in the middle income ranges benefit from having the counterfactual neighborhood available. For households with the Wake County median income of $75,000, the compensating variation is $905. For households with $85,000 it is $2,352. These numbers can be interpreted as annual willingness to pay to have the counterfactual neighborhood available. The compensating variation increases with income since the cost of housing falls for everyone in that neighborhood and the house does not play a role as an asset in this model.

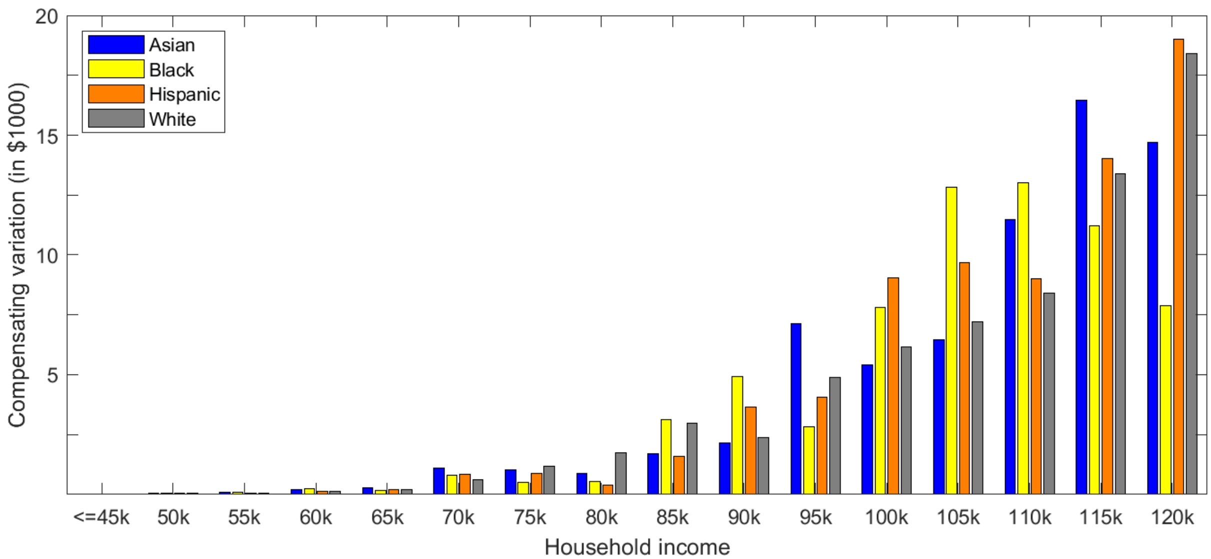
Figure 1.9: Compensating variation by number of children and income



Notes: Figure shows the compensating variation by household income and number of children for a counterfactual policy in which the minimum lot size is held at its current level everywhere except for one neighborhood. The minimum lot size is reduced from 0.5 acres to 0.1 acres in one neighborhood with neighborhood amenities at the 75th percentile. I calculate the compensating variation required to make households indifferent between the counterfactual and actual scenarios.

Figure 1.9 shows the compensating variation by number of children in the households.

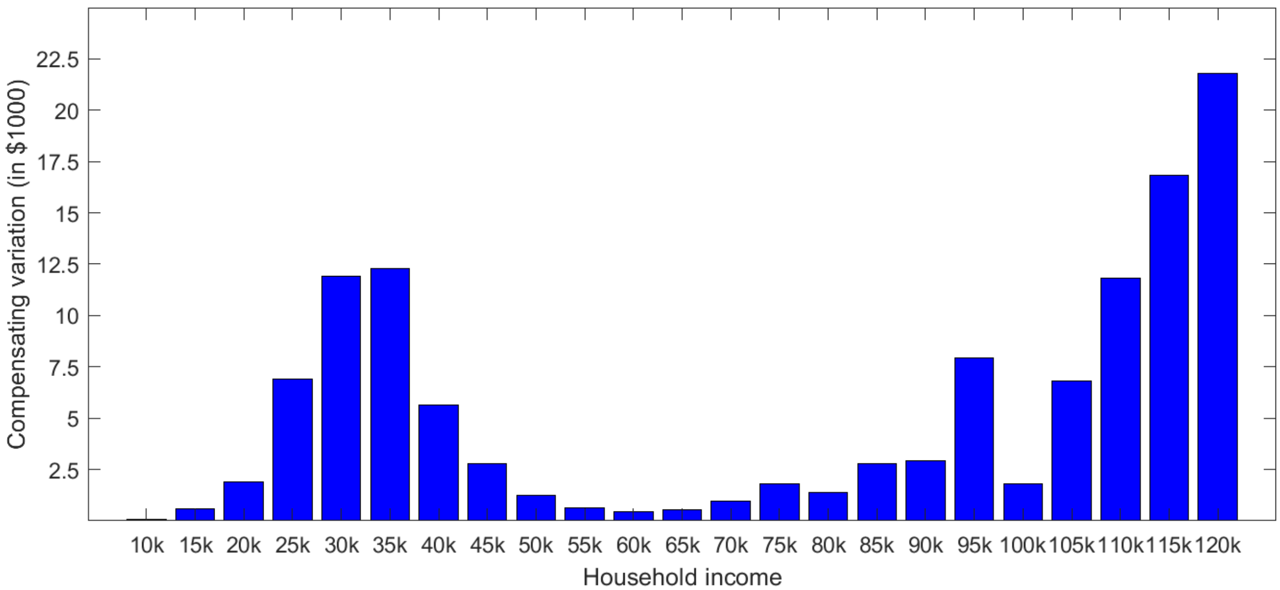
Figure 1.10: Compensating variation by race and income



Notes: Figure shows the compensating variation by household income and race for a counterfactual policy in which the minimum lot size is held at its current level everywhere except for one neighborhood. The minimum lot size is reduced from 0.5 acres to 0.1 acres in one neighborhood with neighborhood amenities at the 75th percentile. I calculate the compensating variation required to make households indifferent between the counterfactual and actual scenarios.

For households of around median income, the highest compensating variation is for households with two children. As household income increases, the compensating variation is largest for households with one child. This variation can be explained by differences in the valuation of neighborhood amenities for different types and by the difference in marginal neighborhoods for households of different incomes. For example, households with two children value school quality highly. That households with two children and incomes between $80,000 and $90,000 benefit the most is due to the fact that in that income range, the new neighborhood offers the best education at the lowest price compared to other neighborhoods that are chosen at high frequencies. Figure 1.10 shows the compensating variation by race of the child. The compensating variation is increasing with income for households of all races. For Black households the compensating variation is decreasing for household incomes greater than $110,000 and for Asian households it is decreasing after household incomes of $115,000. Given the type specific heterogeneity, the marginal attractiveness of the neighborhood that has a changed regulation differs by race. Since Asian households are estimated to have a high preference for the quality of the local public school, this might drive their compensating variation for households with incomes between $90,000 and $115,000 as the changed and now cheaper neighborhood contains a high-performing school.

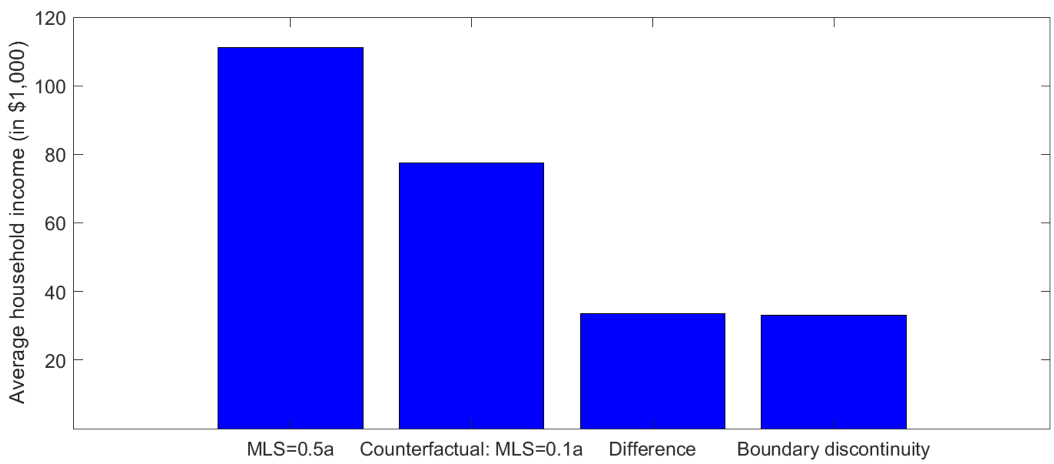
Figure 1.11: Compensating variation by income (low-amenity neighborhood)



Notes: Figure shows the compensating variation by household income for a counterfactual policy in which the minimum lot size is held at its current level everywhere except for one neighborhood. The minimum lot size is reduced from 0.5 acres to 0.1 acres in one neighborhood with neighborhood amenities at the 25th percentile. I calculate the compensating variation required to make households indifferent between the counterfactual and actual scenarios.

Figure 1.11 shows willingness to pay averaged across observable characteristics for households with different incomes. In this case, the neighborhood experiencing a decrease in the minimum lot size from 0.5 acres to 0.1 acres is a neighborhood with amenities at the 25th percentile of the neighborhood fixed effect. Figure 1.11 clearly shows that the households on the margin of choosing this neighborhood have household incomes between $25,000 and $40,000. I calculate an annual compensating variation of between $7,000 and $11,000 for this income range which is substantial. Intuitively, reducing the minimum lot size in this neighborhood is very attractive for households with low incomes who previously would have preferred to live in cheaper neighborhoods with lower amenities. Comparing Figure 1.11 with Figure 1.8 indicates that the marginal households that benefit from a reduction in the minimum lot size are different in high-amenity and low-amenity neighborhoods. The comparison also reinforces our intuition that households with low income are the most constrained by the minimum lot size. This is evidenced by the substantial amount of income that households with incomes between \$25,000 and \$40,000 are willing to give up when the minimum lot size is relaxed compared to a world with stricter minimum lot sizes.

Figure 1.12: Differences in neighborhood income



Notes: This figure refers to the neighborhood whose minimum lot size (MLS) is lowered in the counterfactual. The first bar shows the average income in the data. The second bar shows average income in the counterfactual when minimum lot sizes are lowered from 0.5 acres to 0.1 acres. The third bar plots the difference between average income in the data and in the counterfactual as predicted by the model. The fourth bar plots the difference in average income that is predicted by the boundary discontinuity design.

Figure 1.12 displays average household income for the neighborhood in the current scenario and the counterfactual. The first bar indicates that in the current scenario average neighborhood income is quite high, namely \$111,099. The second bar shows predicted average neighborhood income in the counterfactual with relaxed minimum lot sizes. When the minimum lot size decreases by 80% from 0.5 acres to 0.1 acres and the decrease in price per square foot of land, the average neighborhood income as is now \$77,581 which is only slightly higher than median income in Wake County and represents a fall of roughly 30% in neighborhood income. This is a substantial drop in neighborhood income and suggests that reducing the minimum lot size does indeed have an impact on the affordability of desirable neighborhoods. This counterfactual is computed holding the number of inhabitants in the county fixed. However, one might expect households to move in to the county if such a change were to actually occur. The additional demand might increase the price per square foot and average neighborhood income relative to the counterfactual income that I find now. On the other hand, if minimum lot sizes themselves generate an amenity that households value then by reducing the minimum lot size this amenity should fall, i.e. if low density is something that households inherently prefer then increasing the allowed density makes the neighborhood less attractive from that perspective. Demand for the neighborhood should fall and average income might fall further than is currently estimated.

One way to think about different mechanisms through which the minimum lot size works is to return to the original estimates from the boundary discontinuity design. How can the estimates of the model be compared to the estimates of the boundary discontinuity design? Bar 3 and 4 in Figure 1.12 speak to this difference. The boundary discontinuity design requires that amenities are continuous across the boundary while the minimum lot size regulation changes. In the model and counterfactual the level of the amenity is being held fixed while only the minimum lot size regulation changes. The closest comparison the model can make to the boundary discontinuity design is comparing household income in the neighborhood with identical amenities and a relaxed minimum lot sizes to the neighborhood with strict minimum lot sizes.

There are three crucial ways in which the model and the boundary discontinuity design differ from each other. First In the model there is no notion of distance to the boundary, instead its predictions about neighborhood composition in terms of income can be thought of as the average for the entire neighborhood. If there are differences between the interior of the neighborhood and the boundary, then the model and the regression discontinuity are going to differ. In particular, if we think that those households that particularly value having regulated neighbors sort on the interior of the neighborhood, then comparing the neighborhood average on the interior should be larger than comparing the average right at the boundary. In the boundary discontinuity results earlier, I showed that the difference between the interior and the boundary is small but positive. So if the model captures interior averages rather than the boundary, then the difference in incomes in either neighborhood should be larger than in the boundary discontinuity.

Second, the model counterfactual takes into account general equilibrium effects on price whereas the boundary discontinuity is estimated on the current equilibrium. If there are large equilibrium effects, one might expect the differences in the model to be larger than the differences found by the boundary discontinuity.

Finally, the model and the boundary discontinuity are estimated on different samples. While the boundary discontinuity uses data from all property transactions, the model uses only a sample of households with elementary school children. The difference between the model prediction and the prediction of the boundary discontinuity is also driven by the different samples. For example, if households with children in elementary school have a systematically higher willingness to pay for desirable neighborhood amenities, then the model may predict smaller differences in neighborhood income for this sample than for the overall sample.

Given the robustness check conducted in Section 1.4.7 that compares the interior of the neighborhood to the boundary, we may expect the first type of difference to be small. Considering general equilibrium effects, in this case the change in the existing equilibrium is relatively small - only one neighborhood receives a different minimum lot size, the price in this neighborhood falls but most other prices remain the same or extremely similar to the original equilibrium prices. The lower minimum lot size essentially leading to more land being available in that neighborhood now. The price per square foot drops until households prefer to purchase the larger lot sizes but there are no big changes in the other neighborhoods. Consequently, household's margins are not changing much, except for those who have a strong preference for the neighborhood with changed zoning. Comparing the change in neighborhood income predicted by the boundary discontinuity to the change predicted by the model in this case mostly is informative about the importance of neighborhood amenities created by the regulation apart from the restriction it imposes on the affordability of housing.

Bar 3 in Figure 1.12 shows the difference in neighborhood income for the sample in that neighborhood originally versus the predicted neighborhood income in the counterfactual. This difference in the regulation amounts to a difference of 8 dwelling units per acre (change from 2 to 10 dwelling units per acre) and $33,518. Bar 4 shows the difference in neighborhood income that the boundary discontinuity predicts for this size of jump. The difference between bar 3 and 4 is very small, the model predicts a difference that is $500 larger than what the boundary discontinuity predicts. The fact that this difference is low confirms that the main impact of the minimum lot size stems from the change in affordability at the boundary and suggests that the boundary discontinuity does a relatively good job of measuring the impact when general equilibrium effects are small.

1.6.2 Zoning vacant land

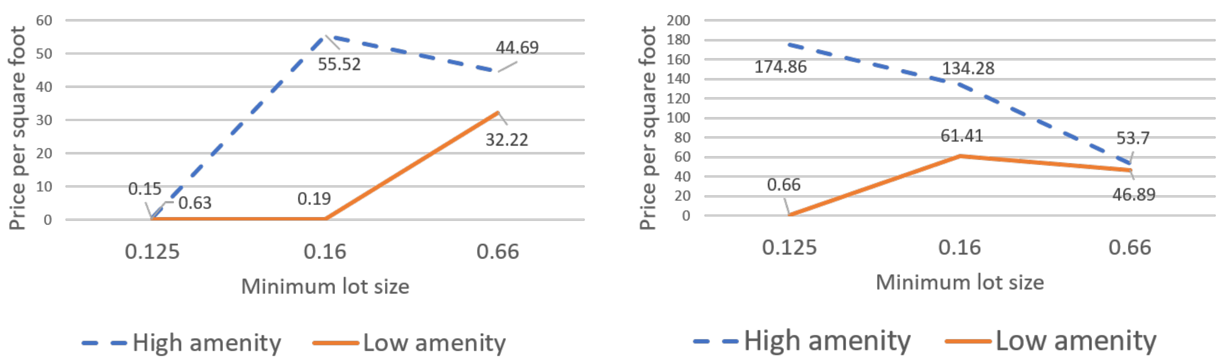
I now turn to the second counterfactual in which I explore the zoning of vacant land when it is developed. In Wake County, 20% of land is vacant. The zoning board of Wake County faces the consideration of how to zone this land for development as the population of the county has been growing rapidly in the last decade. Raleigh's Comprehensive Plan Update for 2030, essentially a proposal for urban planning going forward, explicitly contains the following objective: "Accommodate growth in newly developing or redeveloping areas of the city through mixed-use neighborhoods with a variety of housing types."

To mimic this policy, I now consider a vacant neighborhood and how it could be regulated. In this counterfactual one can imagine that the vacant land is adjacent to an existing neighborhood that has either high or low neighborhood amenities and that the vacant land receives the same amenities as its already established neighbor. I examine differences in the price of land per square foot as well as average neighborhood income for vacant neighborhoods that vary by: a) the minimum lot size, b) size of the neighborhood, c) desirability of neighborhood amenities. Since I am now adding new land, if the population remains fixed, then land prices should mechanistically fall throughout the county. To avoid this, I randomly draw additional households from the distribution of observable characteristics. The number of additional household is chosen as the average number of households in existing neighborhoods that have the same size as the vacant land.

Figure 1.13: Land prices per square foot in counterfactual neighborhoods of different sizes and with different level of amenities

(a) Land prices - median land

(b) Land prices - 25th percentile land



Notes: This figure plots the price of land per square foot for vacant neighborhoods with different combinations of the minimum lot size, neighborhood amenities and size. The left panel plots the price of land for vacant neighborhoods of median size when the neighborhood amenity is either at the 25th (low) or 75th (high) percentile for different minimum lot size regulations. The right panel plots the same values for vacant neighborhoods that are at the 25th percentile of neighborhood size in terms of land area.

Figure 1.13 shows how the price per square foot of land evolves for different values of the amenity, minimum lot size and size of the vacant neighborhood. There are two variations for the amenity - I either endow the vacant neighborhood with amenities at the 25th or the 75th percentile of the estimated neighborhood fixed effect. For the heterogeneity in preferences I assign average school test score, poverty rate and average travel time based on a neighborhood at the 25th or 75th percentile. In terms of the minimum lot size, I study three different values. The average minimum lot size in a neighborhood with school quality at the 75th percentile is 0.2 acres. I study vacant neighborhoods with minimum lot sizes of

0.125 acres, 0.16 acres and 0.66 acres. Finally, I vary the size of the vacant neighborhood to either be the median neighborhood size or a neighborhood of the 25th percentile size.

Two opposing effects on demand determine the price of land in the vacant neighborhood. Supply of land in square feet is fixed in the vacant neighborhood. When the minimum lot size is lowered there are two effects on the demand for square feet of land in the vacant neighborhood that oppose each other. First, demand for square feet of land in the neighborhood increases when the minimum lot size falls as the option value of land increases. Some households that previously did not demand land now do. Second, the demand for square feet of land falls if households that were hitting up against the minimum lot size constraint previously now switch to a smaller lot size. This puts downward pressure on the price. The overall effect on the price of land is the sum of these two effects. Appendix E gives an additional simple intuition for the ambiguity of land price by considering the market for small and large lots before and after the minimum lot size is relaxed.

Figure 1.13a shows the prices in different counterfactuals for a vacant neighborhood of median size with either high or low amenities. I first focus on the case where the vacant neighborhood has high amenities. Compared to the case with intermediate minimum lot size (0.16 acres) the price in the scenario with a lower lot size of 0.125 acres is much lower, namely 63 cents compared to $55 in the case with 0.16 acres. In this case many households seem to switch from large lots to small lots putting a downward pressure on the price that is not outweighed by the fact that demand for square feet of land increases from households that previously did not pick this neighborhood. Overall the price falls. On the other hand, when the minimum lot size falls to 0.16 acres from 0.66 acres, the price per square foot increases from $45 to $55 (it is important to keep in mind that the total minimum expenditure on land is still much higher in the case of a minimum lot size of 0.66 acres compared to 0.16 acres). Going in this direction it seems that overall the demand for square feet in the neighborhood increases, i.e. households now demanding square feet in the neighborhood outweigh the loss in demand from those switching to a smaller lot size. Given that the jump from 0.66 acres to

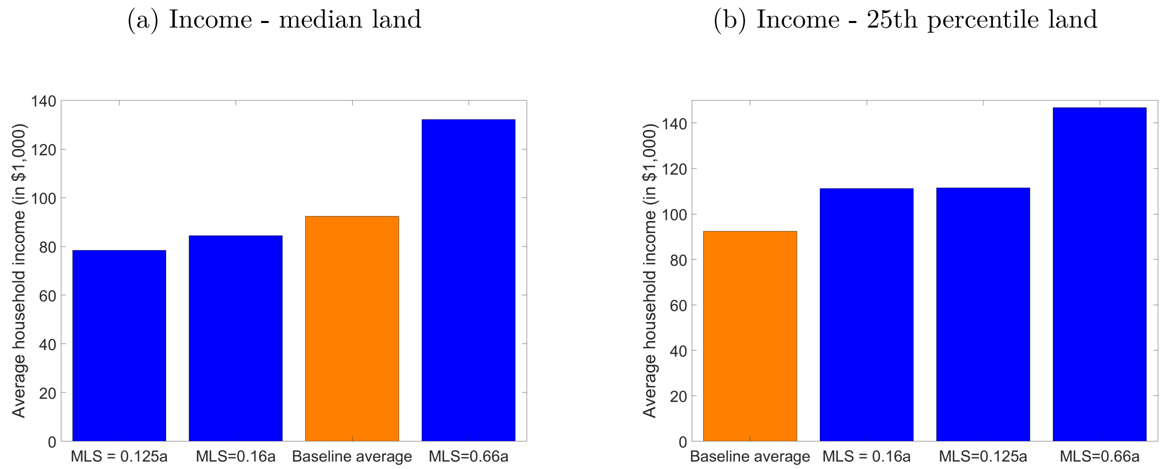
0.16 acres is quite large, this is not surprising. Overall expenditure in each neighborhood is still ordered by minimum lot size with the lowest minimum lot size resulting in the cheapest minimum land expenditure.

Comparing this to the prices for the high amenity neighborhood when the vacant neighborhood has the size of a neighborhood at the 25th percentile in Figure 1.13b, shows that the available quantity of land can be crucial. All prices are larger in the case where the new neighborhood is smaller. In Figure 1.13b, the highest price per square foot of land is in the counterfactual neighborhood with the lowest minimum lot size. Compared to the intermediate regulation, strong new demand for the neighborhood implies a higher price. In the scenario with a neighborhood at the 25th percentile in terms of size, the overall minimum expenditure on land (square foot price multiplied by the minimum lot size) is highest in the neighborhood with the lowest minimum lot size. This illustrates that if the vacant neighborhood is small, it is important to consider that what seems like an objectively small minimum lot size may still lead to a very unaffordable neighborhood if there is a tight land constraint.

The results for overall land expenditure are mirrored by the average neighborhood income in each of these different scenarios. Figure 1.14 shows the average income for neighborhoods with amenities at the 75th percentile from the data (bar 3) and the average income in the counterfactual scenarios with three different lot sizes and amenities at the 75th percentile. Average income falls by 15% relative to the baseline average when the minimum lot size is 0.125 acres. It falls by 8.5% in the case with the intermediate lot size of 0.16 acres and then increases by 30% in the case with strict minimum lot sizes of 0.66 acres. Panel 1.14b shows the same statistics when the vacant neighborhood is smaller. In this case all counterfactual neighborhoods are more expensive, which is a consequence of the fact that overall the amount of land being added is small and desirable so that the upward pressure on land price dominates.

For low amenities in the median neighborhood, Figure 1.13a shows the price per

Figure 1.14: Average neighborhood income in counterfactual neighborhoods of different sizes and with different level of amenities compared to average in actual neighborhoods with these amenities



Notes: This figure plots average income for different sizes of the vacant neighborhood (in terms of land area) and different counterfactual minimum lot size (MLS) regulations. Both panels show vacant neighborhoods endowed with neighborhood amenities at the 75th percentile. Baseline average refers to the average income in neighborhoods at the 75th percentile in the data. The other bars plot the average income predicted by the model for different levels of the regulation.

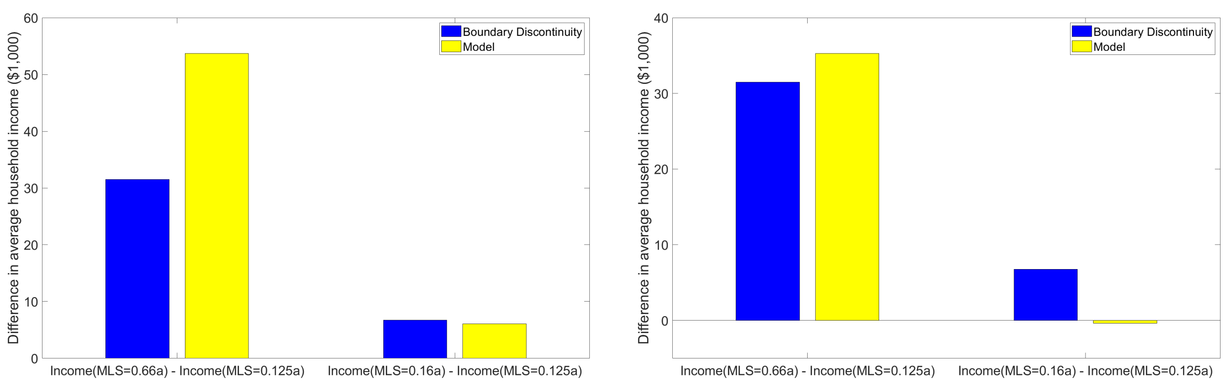
square foot is ordered by minimum lot size from lowest to highest. This can be explained by the fact the households that prefer a low amenity neighborhood will all shift to the smallest lot available causing prices to drop unless the lot size is large enough (0.66 acres) that overall demand for the square feet of land in the neighborhood increases and prices rise essentially because the supply of land is constrained. When the neighborhood is smaller (Figure 1.13b), upward pressure on the price outweighs the downward pressure from households shifting to lower lot sizes for the intermediate lot size but not for the highest lot size, thereby leading the price per square foot to be highest in the case of the intermediate lot size. A discussion of the average income for the low amenity case can be found in Appendix D.

Figure 1.15 again compares the predictions about sorting on income from the boundary discontinuity design with the predictions of the model. Figure 1.15a compares the smallest counterfactual regulation, a minimum lot size of 0.125 acres to the largest (0.66 acres) and the intermediate (0.16 acres) counterfactual regulations. The model predicts a much larger change in average neighborhood income when going from 0.66 acres to 0.125 acre than the boundary discontinuity does. In contrast, the model and the boundary discontinuity are quite similar comparing 0.125 acres of minimum lot size to 0.16 acres of minimum lot size. Going from 0.66 acres to 0.125 acres of minimum lot size creates large equilibrium price effects since supply of land is constrained when the minimum lot size is 0.66 acres. This causes the price difference between the two neighborhoods to be larger than in the boundary discontinuity. The equilibrium price effects of going from 0.16 to 0.125 acres of minimum lot size are smaller. Overall minimum cost of land does not change very much in those cases and the supply constraint is not as binding. As a consequence, the boundary discontinuity and the model are fairly close.

Figure 1.15: Comparison between difference in neighborhood income for regulated vs. less regulated areas predicted by model and boundary discontinuity

(a) Median land

(b) 25th percentile land



Notes: This figure plots the difference in average income predicted by the model and the boundary discontinuity estimates for different levels of the regulation (MLS). Both panels show vacant neighborhoods endowed with neighborhood amenities at the 75th percentile. The left panel shows results for a vacant neighborhood of median size in terms of land area and the right panel shows results for a vacant neighborhood with land area at the 25th percentile.

In Figure 1.15b there is an interesting reversal in the neighborhood income - income is actually lower in the slightly more regulated case (0.16 acres) than when the minimum lot size is 0.125 acres. This is another equilibrium effect that cannot be captured by the boundary discontinuity. The case for a low amenity counterfactual neighborhood is discussed in Appendix D.

Overall, the zoning of vacant land must consider the minimum lot size compared to the available amount of vacant land. If the available land is large relative to the minimum lot size, land prices will be low and the newly developed neighborhood can be affordable when zoned with a small minimum lot size. If minimum lot size is large relative to the available amount of land, then even for small minimum lot sizes the price per square foot can rise enough to ensure that the new neighborhood is even more expensive than an average existing neighborhood with similar amenities. Additionally, the boundary discontinuity and the model agree in their predictions about income sorting as long as the equilibrium price effects in the counterfactual land use regulation schemes are not large. For large equilibrium effects, the boundary discontinuity predictions will be wrong in size and may even have the wrong sign.

1.7 Discussion and Conclusion

This paper considers how land use regulations such as minimum lot sizes affect the neighborhood choices of households by imposing a minimum quantity of housing that needs to be purchased. While previous literature has explored the effect of land use regulations on housing prices, we have little knowledge of whether they are an important determinant of sorting on income into different neighborhoods. Land use regulations are also absent from the study of neighborhood choice more broadly. I begin to fill this gap in the literature by studying the effects of minimum lot sizes.

Exploiting lot level variation within areas with constant amenities and public goods in the minimum lot size regulation and in income, I conduct a boundary discontinuity design at the minimum lot size regulation boundary to study the effect of minimum lot sizes on neighborhood income. I show that in places where the regulation binds it has a considerable effect on income sorting, increasing average income by over $4,000 for a decrease in one dwelling unit per acre. This effect seems to be driven by the fact that lots are larger and housing generally more expensive when land is regulated more strictly. A second order effect is that households have a small willingness to pay for neighbors that live on regulated lots.

To study the welfare implications of changes in the regulation and to explore the primary mechanism through which the minimum lot size operates further, I then develop a model of neighborhood choice that incorporates minimum lot sizes as a constraint on households deciding where to live. The model naturally leads to sorting on income into different neighborhoods as the minimum lot size affects the minimum level of non-land consumption. One caveat of the model is that it does not take into account the possibility that relaxing the minimum lot size might change the amenity level of a neighborhood and make it less desirable. My analysis in this paper suggests that this effect might be secondary compared to the effect of the regulation on construction. Nevertheless, for large-scale changes in land use regulations, peer effects could play an important role and a deeper analysis is left for future research. Another limitation is given by the fact that land has no value as an asset and hence all households benefit from a reduction in the price of land. These are both important extensions that are left for future research. Importantly, the model particularly captures the preferences and behavior of households whose neighborhood choices are distorted by the minimum lot size. When the minimum lot size is lowered, some of these households would prefer to live in a desirable neighborhood. I use the parameters of the model to conduct counterfactuals that mimic current policy proposals involving minimum lot sizes.

The results of the model are informative for current policy debates surrounding the relaxation of land use regulations to end the housing shortage and provide a larger variety of housing options. I find that small, localized reductions in the minimum lot size can lead to desirable neighborhoods being more affordable to lower income households that value amenities highly. In the counterfactuals I simulate, households with household income below $45,000 to not benefit from this policy, but households with median income ($75,000) value it at $900 annually and households with household income of $85,000 value it at $2,500 annually. The results of the model imply that the affordability of vacant land will depend on the minimum lot size relative to the available amount of land that is to be zoned. Keeping this in mind, my results imply that vacant land with desirable neighborhood amenities can be zoned with minimum lot size regulations that make it more broadly affordable.

1.A Selection of remaining boundaries

Table 1.A-1 shows a comparison between the characteristics of lots whose closest boundary is excluded in the boundary discontinuity estimation and lots that are closest to admissible boundaries and are therefore included in the boundary discontinuity estimation. Boundaries that are excluded are minimum lot size regulation boundaries that overlap with elementary school attendance area boundaries, highways, rivers or streams. This selection of boundaries also excludes municipality boundaries within Wake County and the boundaries of the county itself.

Table 1.A-1: Comparison of lots near excluded and non-excluded boundaries

	Nearest boundary not excluded		Nearest boundary excluded		Difference	
	Mean	SD	Mean	SD	Difference	Standard error
Household income	98750.92	127043.75	96365.62	84398.16	2385.30*	(1131.18)
Sale price	263171.45	189875.28	252856.85	154737.81	10314.59***	(1688.94)
Loan amount	205860.80	132292.60	206567.14	120408.39	-706.34	(1200.24)
Minimum lot size	0.23	0.14	0.23	0.59	-0.00	(0.00)
Deed acres	0.30	0.75	0.26	0.47	0.04***	(0.01)
House size	0.05	0.02	0.05	0.02	-0.00***	(0.00)
Age of house	39.89	164.42	42.10	225.42	-2.21	(1.67)
Borrower black	0.11	0.32	0.10	0.31	0.01**	(0.00)
Borrower white	0.77	0.42	0.76	0.43	0.01	(0.00)
Observations	51467		15102		66569	

Notes: This table compares lots in my sample of merged Zillow transactions data with data from the Home Mortgage Disclosure Act whose closest boundary is excluded to lots whose closest boundary is not excluded. I exclude minimum lot size regulation boundaries that overlap with highways, rivers, streams or elementary school attendance areas. Columns 5 and 6 test for a difference in the means of both groups. My analysis uses the set of lots that are near non-excluded boundaries.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Columns 1 and 2 show the characteristics of lots that are retained in the sample, columns 3 and 4 show characteristics of lots whose closest boundary is excluded. Columns 5 and 6 show the difference and perform a means comparison test. Household income of lots near non-excluded boundaries is about $2,300 higher than that of lots near excluded boundaries. The sales price of lots near non-excluded boundaries is higher and the realized deeded acreage 0.04 acres larger. Given that many of the lots near excluded boundaries are lots near highways which tend to be less desirable locations, these trends are reasonable. However, the difference in household income is not very high. Reassuringly, the minimum lot size, house size and age of house are not statistically significantly different from each other for lots near excluded and non-excluded boundaries. In terms of race, lots near non-excluded boundaries are one percentage point more likely to have a black mortgage applicant.

Overall, it seems that the selection of the remaining boundaries is towards more valuable properties but in terms of the characteristics of the inhabitants and the regulation, the excluded properties do not strongly differ from the non-excluded properties.

1.B Results using Census data

In the following I discuss results of the sorting regression in Section 4 (equation 1.1) using data that is not at the lot level but instead at the Census block or Census block group level. There are 14,442 Census blocks in Wake County according to the 2010 Census geography definition. The minimum lot size regulation never varies within Census blocks. Distance to the boundary is now defined as distance of the Census block centroid to the closest regulation boundary. The set of excluded and non-excluded boundaries is the same as in the main analysis. These results should be viewed as additional evidence of sorting at the boundary even when the level of aggregation is increased to hide much of the variation right at the boundary.

Table 1.B-2: Median household income (ACS)

	(1) 30% binding	(2) not binding	(3) 50% binding	(4) not binding	(5) 30% binding	(6) not binding	(7) 50% binding	(8) not binding
$\mathbb{1}\{\text{regulated}\} * \Delta_{density}$	248.6**	-40.14	249.4*	151.2	503.9***	-366.0	593.9**	21.17
	(80.39)	(177.5)	(109.7)	(102.7)	(126.4)	(265.2)	(186.5)	(158.3)
Distance to boundary	≤0.5mi	≤0.5mi	≤0.5mi	≤0.5mi	≤0.2mi	≤0.2mi	≤0.2mi	≤0.2mi
Distance	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic
Boundary f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4450	2042	2417	4075	2328	1202	1243	2287

Notes: The dependent variable is median household income by Census block in 2013. All regressions include boundary fixed effects and municipality fixed effects. I vary the distance to the boundary. Columns 3 and 5 focus on Census blocks where at least 30% of lots are binding compared to blocks where less than 30% of lots are binding. Similarly for Columns 3 and 7. $\mathbb{1}\{\text{regulated}\}$ is an indicator that is 1 if a lot is on the relatively more regulated side of its boundary. $\Delta_{density}$ is the absolute value of the difference between the density regulation on the more regulated side of the boundary and the density regulation on the less regulated side of the boundary. $\mathbb{1}\{\text{regulated}\} * \Delta_{density}$ can be interpreted as the change in median household income for a decrease in one dwelling unit per acre of allowed density.
Standard errors in parentheses,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.B-2 column 1 estimates the main treatment coefficient β (from equation 1.1) for blocks where at least 30% of lots are exactly at the minimum lot size. The treatment coefficient can be interpreted as meaning that a block with a regulation allowing one less dwelling unit per acre than its neighbor has a median income of 248$ more than its neighbor. Column 2 looks at the same outcome variable but for the sample of blocks where less than 30% of lots are binding. The coefficient is much smaller and switches sign but is also very imprecisely estimated. This supports the intuition that the effect of sorting should be larger

at binding boundaries where the minimum lot size is more salient.

Columns 3 and 4 repeat the same exercise but this time consider boundaries that either have at least 50% of binding lots (column 3) or less than 50% of binding lots (column 4). The same pattern as before emerges, with a statistically insignificant and smaller effect for the not binding sample and a statistically significant effect of $249 for the binding sample. The expectation of a larger effect for more binding boundaries (50% binding vs 30% binding) is not fulfilled here, both effects seem to be the same.

Table 1.B-3: Other neighborhood characteristics (Census and ACS)

	(1)	(2)	(3)	(4)	(5)	(6)
	% poverty		% renting		% non-white	
$\mathbb{1}\{\text{Regulated}\} * \Delta_{Density}$	-0.00143**	-0.00143*	-0.0142***	-0.0183***	0.00242	-0.0000912
	(0.000453)	(0.000564)	(0.00336)	(0.00418)	(0.00245)	(0.00292)
Distance to boundary	full	≤0.5mi	full	≤0.5mi	full	≤0.5mi
Distance	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic
Boundary f.e.	Yes	Yes	Yes	Yes	Yes	Yes
City f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1296	1186	708	652	817	752

Notes: The dependent variables come from the American Community Survey and are the fraction of households with household income below the poverty level, the fraction of households renting instead of owning their dwelling and the fraction of households that are non-white by Census block in 2013. All regressions include boundary fixed effects and municipality fixed effects. I vary the distance to the boundary. All specifications only include Census blocks where at least 30% of lots are binding. $\mathbb{1}\{\text{regulated}\}$ is an indicator that is 1 if a lot is on the relatively more regulated side of its boundary. $\Delta_{density}$ is the absolute value of the difference between the density regulation on the more regulated side of the boundary and the density regulation on the less regulated side of the boundary. $\mathbb{1}\{\text{regulated}\} * \Delta_{density}$ can be interpreted as the change in median household income for a decrease in one dwelling unit per acre of allowed density.
Standard errors in parentheses,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Columns 5-8 do the same analysis as columns 1-4 but on a sample of blocks within 0.2 miles from the boundary. Columns 1-4 in contrast looked at a sample of blocks 0.5 miles from the boundary. Overall, the smaller the bandwidth, the more the effect of regulation should be picked up, because the observations on either side become more similar to each other the smaller the bandwidth. The effect size doubles in column 5, suggesting sorting on income of $504 for a decrease in one dwelling unit per acre. Compared to column 6, the effect is slightly larger for the more binding boundaries. Comparing with the non-binding boundaries the pattern is the same as before: effects of sorting at the boundaries where minimum lot size does not seem to be a constraint are imprecise and small. The fact that this more aggregate data results in smaller effects indicates that average Census block group income averages mask a lot of the variation at regulation boundaries.

Table 1.B-3 examines some other variables of neighborhood composition. I focus only on those boundaries where at least 30% of the lots are binding. I find no evidence of sorting by race at the minimum lot size boundary within school district for racial composition by Census block. I find evidence of a reduction in households renting, about 1.4% at the mean per decrease in 1 dwelling unit per acre and 2.6% decrease in poverty by decrease in 1 dwelling unit per acre.

1.C Sorting on test scores

Since public education is one of the main locally provided public goods and the North Carolina administrative school records contain a full set of end of grade test scores for all children in my sample, I can examine the difference in test scores on either side of a minimum lot size regulation boundary. These results speak to the degree of sorting that occurs at the minimum lot size boundary even within school attendance areas but should not be interpreted as the causal effects of the minimum lot size on school outcomes. Instead these results can be interpreted as suggestive evidence of neighborhood peer effects when the neighborhood is defined by the minimum lot size regulation. Since the administrative school data only contains location information at the Census block group level, I here calculate the distance of Census block group centroids to the land use regulation boundary. Since this variation is much more coarse, I cannot reduce the distance to the boundary to less than 0.4 miles.

Table 1.C-4 shows the results of those regressions. The dependent variable is the standardized end of grade math test score for children in grades 3-5. I restrict attention to those Census block groups where at least 30% of the lots are binding. At a distance of less than 1 mile from the boundary, for block groups where at least 30% of the lots are binding, a difference of one dwelling unit per acre less increases the standardized math test score by 0.00427 standard deviations after controlling for school fixed effects (column 4). At the average jump of 3 dwelling units per acre, this implies an increase in the math test score for one tenth of a standard deviation which is a considerable effect when compared to results of the school peer effect literature (for example Hanushek, Kain and Rivkin (2009)). When reducing the distance to the boundary even further and controlling for school quality through school fixed effects (columns 4 and 6) the effects are less precise.

One can speculate what these effects imply. One interpretation is that they speak to the importance of considering a definition of neighborhoods that includes land use regulation. The density regulation may be working to shape the neighborhood in ways that are not captured by commonly used definitions of neighborhoods like Census tracts for example.

Table 1.C-4: Boundary discontinuity with test scores

	(1) Std. math score	(2) Std. math score	(3) Std. math score	(4) Std. math score	(5) Std. math score	(6) Std. math score
$\mathbb{1}\{\text{regulated}\} * \Delta_{density}$	0.00361*	0.00427*	0.00333	-0.00291	0.0115*	0.00133
	(0.00184)	(0.00216)	(0.00324)	(0.00390)	(0.00516)	(0.00645)
Distance to boundary	$\leq$ 1 mile	$\leq$ 1 mile	$\leq$ 0.5 miles	$\leq$ 0.5 miles	$\leq$ 0.4 miles	$\leq$ 0.4 miles
Distance	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic
Boundary f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Year f.e.	Yes	Yes	Yes	Yes	Yes	Yes
School f.e.	No	Yes	No	Yes	No	Yes
Observations	39820	39820	21591	21591	16399	16399

Notes: The dependent variable is the standardized end of grade math test score for children in grades 3-5 in the years 2010-2013. All regressions include boundary fixed effects and municipality fixed effects. I vary the distance to the boundary. All specifications include controls for disability of the student and the student's sex. $\mathbb{1}\{\text{regulated}\}$ is an indicator that is 1 if a lot is on the relatively more regulated side of its boundary. $\Delta_{density}$ is the absolute value of the difference between the density regulation on the more regulated side of the boundary and the density regulation on the less regulated side of the boundary. $\mathbb{1}\{\text{regulated}\} * \Delta_{density}$ can be interpreted as the change in standardized math test score for a decrease in one dwelling unit per acre of allowed density.
Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

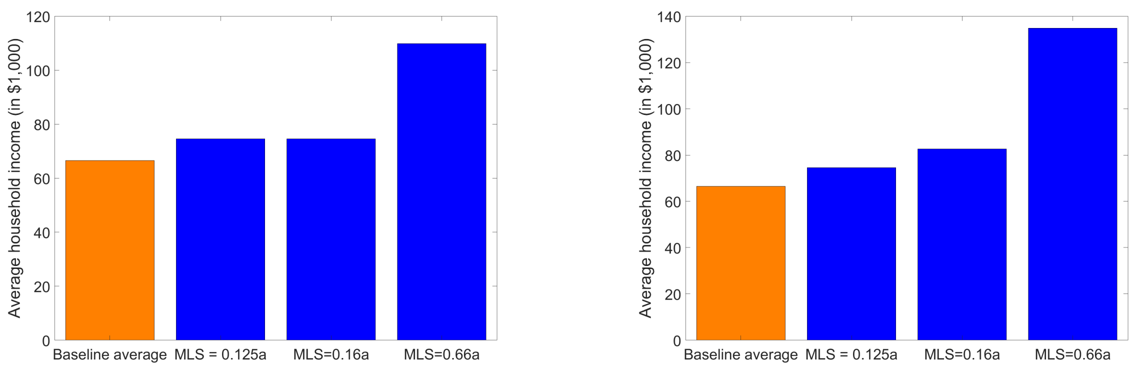
These findings indicate that there is a positive correlation between tighter land use regulation and school performance even within children attending the same school. This might be purely the result of sorting. However it could also mean that neighborhood effects are different in differently regulated areas. This explanation is one area for more exploration in further research.

1.D Vacant land with low amenities

Figure 1.D-1: Average neighborhood income in counterfactual neighborhoods of different sizes and with different level of amenities compared to average in actual neighborhoods with these amenities

(a) Income - median land

(b) Income - 25th percentile land

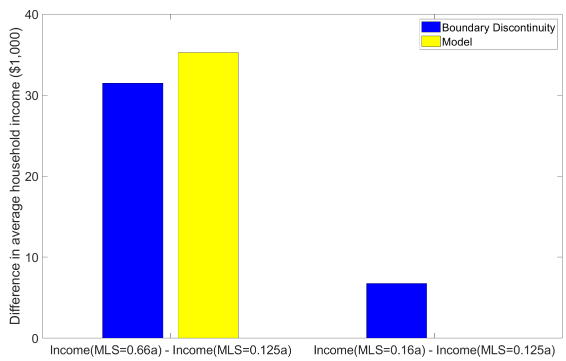


Notes: This figure plots average income for different sizes of the vacant neighborhood (in terms of land area) and different counterfactual minimum lot size regulations. Both panels show vacant neighborhoods endowed with neighborhood amenities at the 25th percentile. Baseline average refers to the average income in neighborhoods at the 25th percentile in the data. The other bars plot the average income predicted by the model for different levels of the regulation.

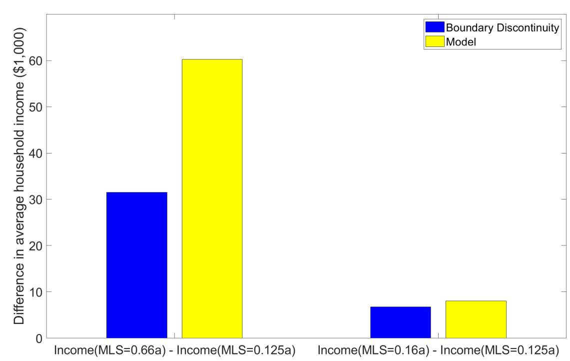
This appendix discusses average income and differences between the model and the boundary discontinuity for vacant neighborhoods with low amenities (amenities at the 25th percentile of the neighborhood fixed effect). Figure 1.D-1 shows average income for vacant neighborhoods with low amenities after they are regulated with different minimum lot sizes. Figure 1.D-1a shows neighborhood income when the vacant neighborhood is of median size. Compared to the baseline average of $66,525, neighborhood income increases for all types of the regulation (low,intermediate or high). This can be explained by recognizing that the average minimum lot size in neighborhoods with amenities at the 25th percentile tends to be lower than 0.125 acres. Hence all regulations in this case represent an increase relative to the baseline average and overall minimum housing expenditure is higher than in comparable existing neighborhoods. Therefore it is not surprising that average income rises despite the fact that the neighborhood is less desirable. Figure 1.D-1b shows average income for a vacant neighborhood at the 25th percentile of neighborhood size. Qualitatively the pattern is very similar to Figure 1.D-1a however the increase in neighborhood income for minimum lot sizes of 0.16 and 0.66 acres is stronger. This is a reflection of the strong price increases in Figure 1.13b which are a result of strong constraints on land supply.

Figure 1.D-2: Comparison between difference in neighborhood income for regulated vs. less regulated areas predicted by model and boundary discontinuity

(a) Median land



(b) 25th percentile land



Notes: This figure plots the difference in average income predicted by the model and the boundary discontinuity estimates for different levels of the regulation. Both panels show vacant neighborhoods endowed with neighborhood amenities at the 25th percentile. The left panel shows results for a vacant neighborhood of median size in terms of land area and the right panel shows results for a vacant neighborhood with land area at the 25th percentile.

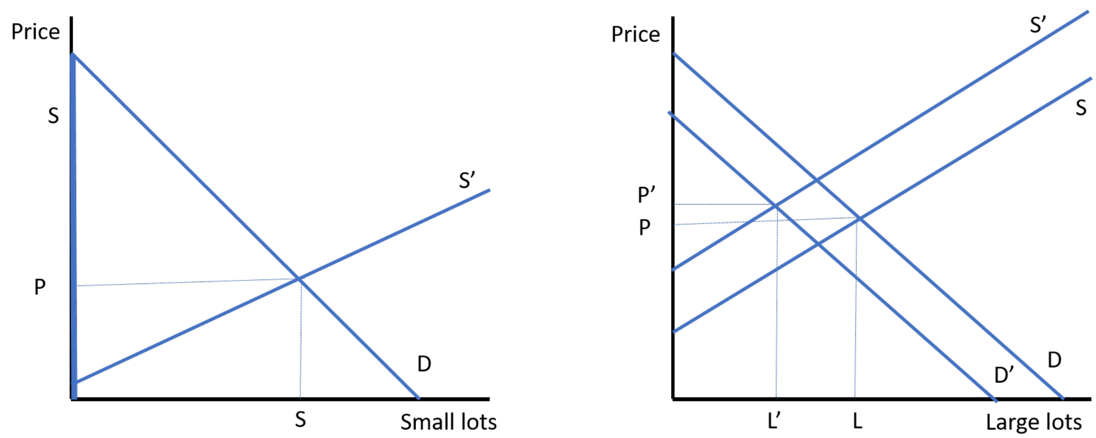
Figure 1.D-2 compares the income predictions of the model and the boundary discontinuity design. Figure 1.D-2a shows that the predictions of the boundary discontinuity and model differ in both cases but even more for the differences in income between a minimum lot size of 0.16 and 0.125 acres. A potential explanation is the following: The predictions from the boundary discontinuity design are from a sample of boundaries at which the regulation is binding which tend to be desirable neighborhoods. Figure 1.13a shows that for the low amenity neighborhood, prices hardly change for a minimum lot size of 0.125 or 0.16 acres which implies that demand overall does not change much going from one minimum lot size to another. Either new demand for the neighborhood is exactly offset by households shifting to smaller lot sizes and demanding less land or the regulation is not binding and households

do not switch to smaller lot sizes. Consequently, the neighborhood income does not change since the regulation was not sufficiently binding. Figure 1.D-2b shows a very similar pattern to Figure 1.15a. The model predicts a large change in the neighborhood income for a large change in the minimum lot size since equilibrium effects on the price are large. The boundary discontinuity fails to capture this. When the change in the lot size is small, the boundary discontinuity and the model predictions are very similar.

1.E Land price of vacant land

Figure 1.E-3 shows how the price of land is affected when the minimum lot size is relaxed. To illustrate this situation it helps to focus on two markets - the market for large lots and the market for small lots. Before the minimum lot size is relaxed, the market for small lots (given by the left panel of Figure 1.E-3) shows a demand of D and a supply of S, i.e. the supply of small lots is zero due to the minimum lot size regulation. In the market for large lots (right panel of Figure 1.E-3) the original equilibrium is given by the quantity L and price P, i.e. a positive quantity of large lots is sold in the original equilibrium.

Figure 1.E-3: Lowering minimum lot size and land price



Notes: This figure shows a simple illustration of the effects of a decrease in the minimum lot size on land prices. The panel on the left shows the the markets for small lots, the panel on the right shows the market for large lots. Initially, the equilibrium in the market for large lots is given by L and P. The market for small lots is regulated to have a supply of 0. After the minimum lot size is relaxed, the equilibrium in the market for small lots is given by S and P and the equilibrium in the market for large lots is given by L' and P'.

When the minimum lot size is relaxed, the supply of large lots decreases to S' as some developers now provide small lots instead. This results in an increase in the land price. At the same time, some households that previously were constrained by the regulation to purchase large lots, switch to buying smaller lots, so demand for large lots will also fall. The ensuing effect of the land price is now ambiguous. If demand for large lots falls by a lot relative to supply, i.e. if many households were constrained by the minimum lot size previously, then the price of large lots may fall lower than the original price. If demand for large lots does not fall strongly compared to the reduction in supply of large lots, then the price of land may rise compared to the initial state.

In the market for small lots the equilibrium is determined by the supply of and demand for small lots. The demand for small lots is a combination of demand from two sources: On the one hand there is demand stemming from households that switch from a large to a small lot. On the other hand there is new demand from households that previously did not demand land in this neighborhood due to the increased option value of land when the minimum lot size falls.

1.F Common neighborhood utility components

Table 1.F-5: Correlation of neighborhood fixed effects with neighborhood characteristics

	(1)	(2)	(3)	(4)	(5)
	δ_n	δ_n	δ_n	δ_n	δ_n
Average math score	0.0117***	0.0117***	0.0120***	0.0111***	0.0102***
	(0.00141)	(0.00138)	(0.00164)	(0.00164)	(0.00161)
Fraction college degree	-0.318		-0.0141	0.791*	0.597
	(0.223)		(0.261)	(0.321)	(0.363)
Fraction below poverty level	-1.771***		-1.585***	-1.002*	-0.877
	(0.342)		(0.348)	(0.499)	(0.495)
Fraction asian households		3.568***			3.425**
		(0.984)			(1.092)
Fraction black households		2.591**			1.990*
		(0.848)			(0.876)
Fraction white households		2.530**			1.688
		(0.794)			(0.890)
Average travel time to work			0.0342**	0.00926	0.0178
			(0.0118)	(0.0119)	(0.0121)
Block group population			-0.0000925	-0.0000858	-0.0000963
			(0.0000510)	(0.0000485)	(0.0000493)
Fraction unemployed				1.657*	1.015
				(0.742)	(0.760)
Median age				-0.000143	0.000155
				(0.00771)	(0.00809)
Price of land per acre				-0.00000150***	-0.00000144***
				(0.000000382)	(0.000000377)
Observations	427	427	427	427	427
R^2	0.142	0.134	0.166	0.228	0.249

Notes: The dependent variable is the estimated common neighborhood utility δ_n. The regressions are run at the location/Census block group level. The fraction of households with college degree, with income below poverty level, the fraction of households by block group that are black, white and asian, the overall block group population, the median age by block group and the fraction of households that are unemployed come from the American Community Survey by Census block group in 2013. I vary the distance to the boundary. Average math score is the average end of grade mathematics test score in each location's elementary school. Standard errors in parentheses,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Chapter 2

Rural Physician Shortages and Policy Intervention (with Dennis B. McWeeny)

Chapter Summary

Compared to populations in non-shortage areas, populations in areas with physician shortages have measurably worse health outcomes. We analyze the effects of government incentive programs intended to eliminate physician shortages. Using a differences-in-differences approach, we estimate that student loan forgiveness programs lead to an increase of three physicians on average per rural county. We then estimate a structural model of physician specialty and location decisions to simulate counterfactual incentive policies. Physicians are relatively unresponsive to differences in salaries across locations and prefer to practice medicine in their home state. Consequently, current incentive payments are too small to eliminate shortages.

2.1 Introduction

Achieving equitable access to healthcare for all citizens of the United States has long been a goal of policymakers. Healthcare provision is labor-intensive, requiring a large and highly trained workforce of nurses and physicians. However, there are persistent shortages of primary care physicians (PCPs) in many rural areas of the United States.[1] Although fourteen percent of the U.S. population lives in rural areas, only ten percent of PCPs practice medicine there. Areas with fewer PCPs per capita have measurably worse health outcomes, including higher rates of preventable hospitalizations, higher mortalities from preventable chronic diseases like diabetes and heart disease, and higher incidence of mental illness. All of these health outcomes can be improved through the preventative care typically provided by PCPs.

Rural areas lack many amenities that well-educated people like physicians tend to desire, making these areas less attractive places to live. In the classic Rosen-Roback model of location choice in labor economics (Rosen (1979); Roback (1982)), locations are characterized by wages, rents, and amenities. If the amenities are non-productive, then this framework predicts that wages will be higher and rents will be lower in areas with few amenities. While this model predicts that workers will be compensated for living in less desirable locations, government intervention in the healthcare marketplace prevents wages from fully adjusting for physicians. Medicare, the primary insurer for people 65 and over, dominates the landscape of healthcare spending and accounts for 20 percent of all healthcare expenditures in the United States.[2] Medicare adjusts its reimbursement rates for the costs of living and running a medical practice in a particular geographic location.[3] As a result, wages for physicians in attractive urban areas are actually adjusted upwards relative to what the Rosen-Roback framework would predict, which generates a disparity in wages for physicians across locations

[1]Primary care physicians perform the essential tasks of diagnosing medical conditions and prescribing drugs or other forms of treatment, sometimes referring patients to a more specialized physician.

[2]https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nhe-fact-sheet.html

[3]Recent evidence suggests that Medicare has a powerful influence on reimbursement rates for private insurers; Clemens and Gottlieb (2017) find that a $1.00 increase in Medicare's fees increases the reimbursement rates of private insurers by $1.16.

that contributes to shortages (see Figures 2.1 and 2.2).[4]

In order to combat PCP shortages, most states have introduced policies to incentivize physicians to practice in rural areas. As of 2015, the Distict of Columbia and every state except Florida had introduced a loan forgiveness or scholarship program for primary care physicians; on average, these programs grant $98,447 in debt relief per physician.[5] Despite these programs, shortages have persisted for decades and have even intensified in certain areas.

In this paper, we seek to determine if the incentive programs designed to attract physicians to rural areas are effective at reducing shortages. To do so, we provide both reduced-form and model-based evidence on the effects of incentive payment programs on the number of physicians practicing in rural areas.

First, we use a differences-in-differences identification strategy to estimate the average treatment effect of student loan forgiveness programs on the number of physicians practicing in rural counties. We exploit variation across states in the year of implementation of these policies to identify causal effects. We also test for pre-trends as well as differences in the short- and long-run effects using an event study design.

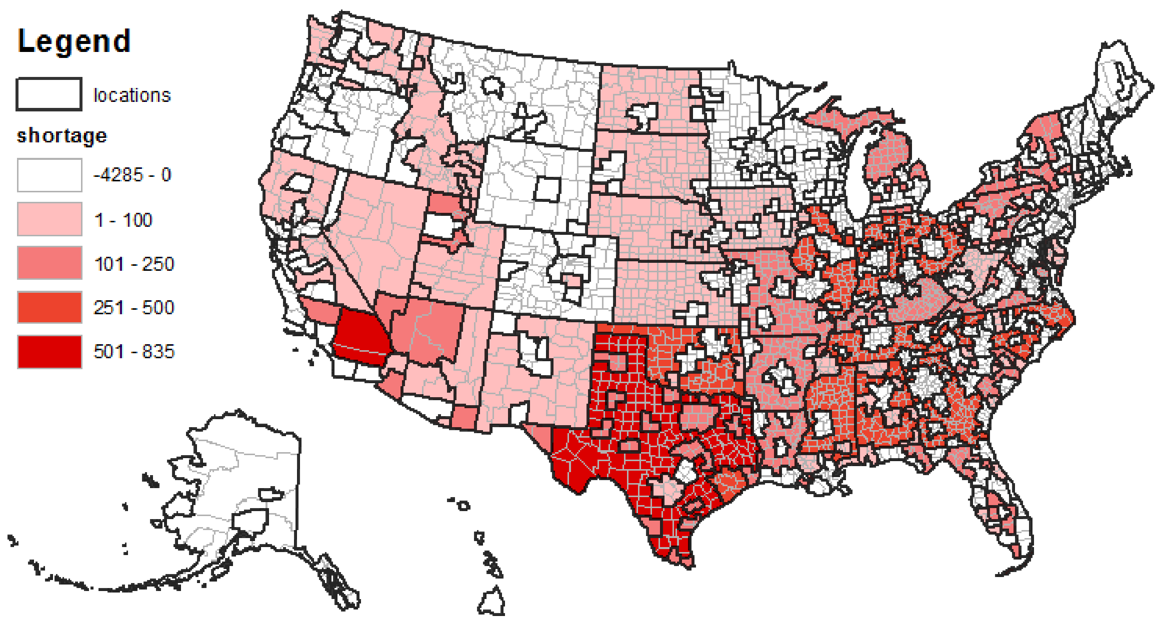
Our differences-in-differences estimates indicate that loan forgiveness programs cause a long-run increase of about three physicians per rural county, on average. This increase is driven almost entirely by MDs (Doctors of [Allopathic] Medicine) who form a much larger fraction of physicians than do DOs (Doctors of Osteopathic Medicine).[6] These results are robust to the inclusion of additional fixed effects as well as a placebo test of randomized implementation years.

Building on these results, we then estimate a static discrete choice model of physician

[4]We discuss our definition of shortages and locations in more detail below. Our shortage calculations correspond reasonably well with shortage calculations done by the Health Resources and Services Administration (HRSA). The HRSA defines some additional shortage areas in Montana, Wyoming and Colorado, probably because they take into account time to nearest source of care along with physician work capacity. Overall, we take a very conservative stance on what constitutes a PCP shortage.

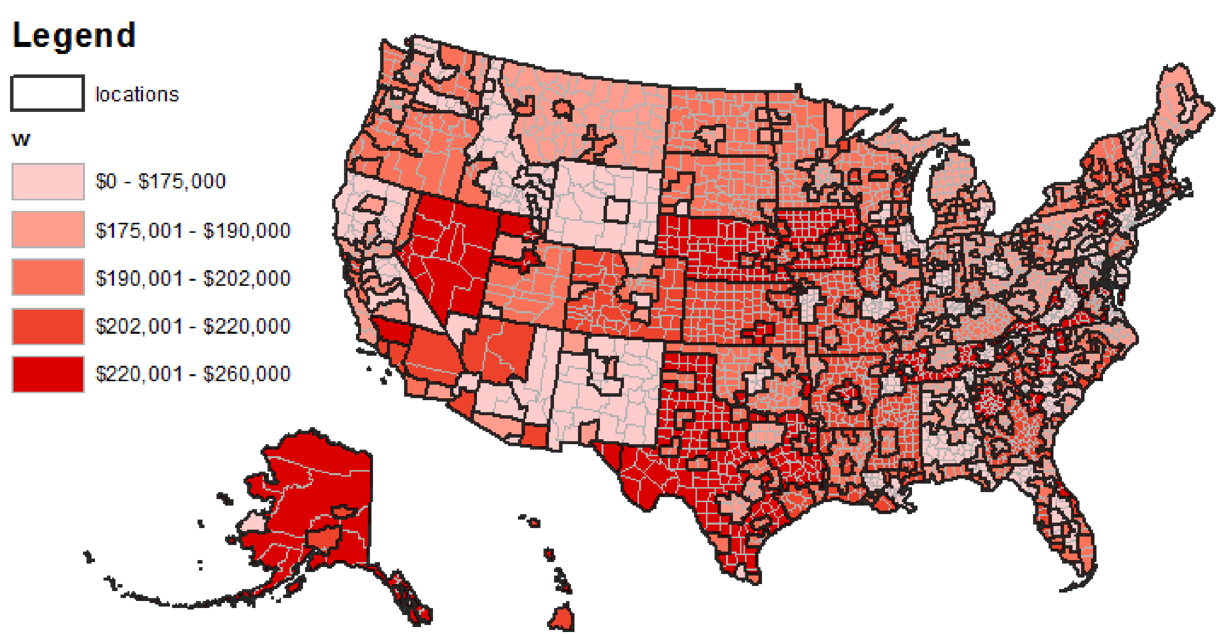
[5]Table 2.6 in Appendix 2.A presents a summary of these programs.

[6]The training for DOs focuses on a more holistic approach to heathcare than the training for MDs, including a greater focus on preventative care.



Source: Authors' calculations from HRSA and Census Bureau data, using 1,500:1 population-to-PCP ratio

Figure 2.1: Shortages of primary care physicians across the United States, 2014



Source: BLS Occupational Employment Statistics (OES)

Figure 2.2: Annual salaries of primary care physicians across the United States, 2014

specialty and location decisions. This model is similar to those commonly used in the migration and empirical industrial organization literatures. Our model accounts for salaries, rents, and other location-specific amenities that may be important factors for physicians. We address the endogeneity of salaries by using an instrument that exploits exogenous variation in the number of births across time within a given location. We also incorporate physicians' preferences for living close to their home state and allow preferences to vary by gender, medical degree type, and country of origin.

The results from the structural model indicate that physicians are relatively unresponsive to small differences in salaries across locations and strongly prefer to live in their home state. The income elasticity for the average location is 1.91, meaning a one percent increase in annual salaries generates just a 1.91 percent increase in the fraction of PCPs who choose to practice medicine in that location. We also find that American physicians would need to be paid an additional $9,338 per year on average in order to forgo living in their home state.

We then use the model to simulate several counterfactual incentive payment policies. We find that the current loan forgiveness programs reduce shortages by less than one percent per year — far smaller than necessary to eliminate shortages. However, we also simulate several revenue-neutral policies that tax PCPs who practice in urban areas and subsidize those who practice in rural areas. These policies are much more effective at reducing shortages; a ten percent tax rate and corresponding 35.41 percent revenue-neutral subsidy rate would eliminate between five and ten percent of the shortage per year in rural areas. Given these results, we suggest that policymakers either implement a nationwide revenue-neutral policy or focus their efforts on recruiting more physicians who were raised in rural areas to stay and practice medicine there.

The rest of this paper proceeds as follows. In the next section, we discuss the related literature on physician shortages, physician location decisions, and static migration models more generally. In Section 2.3, we discuss our data sources, define shortage areas, and discuss current incentive payment programs. We present the differences-in-differences methodology and results as well as several robustness checks in Section 2.4. In Section 2.5, we introduce the structural model, discuss identification and estimation, and conduct simulations of alternative incentive payment schemes. We then discuss limitations of our analysis and opportunities for future work before concluding in Section 2.6.

2.2 Related Literature

The existence of physician shortages has been well-documented among policymakers and the healthcare community (see, for example, IHS Inc. (2016); Petterson et al. (2012); Goodman et al. (1996); Schwartz (2011)). Petterson, Rayburn and Liaw (2016) estimate that the United States will require 51,880 additional PCPs by 2025. Nevertheless, physician shortages have received little attention in the economics literature due to a lack of publicly available data and the difficulty of pinning down the frictions that cause shortages (Newhouse et al., 1982). We take a step towards filling this gap in the literature.

Physician shortages are related to two decisions that physicians make: the choice of specialty and the choice of location. Since PCPs tend to earn less than more specialized physicians, medical school graduates increasingly decide against becoming generalists (Schwartz et al., 2005). However, the role of income in the location decision has not been studied extensively. There is some evidence suggesting that medical school and residency location are important determinants of practice location (Burfield, Hough and Marder, 1986). Additionally, certain types of medical school graduates such as DOs are more likely to practice in rural areas (Chen et al., 2010).[7] We account for these findings in our model of physician specialty and location choice.

One might argue that physician shortages are only a problem if they lead to adverse health outcomes in shortage areas; the evidence on health outcomes by metropolitan status is mixed. However, Reschovsky and Staiti (2005) and Chan, Hart and Goodman (2005) find

[7] The training for DOs focuses on a more holistic approach to healthcare than the training for MDs, including a greater focus on preventative care.

that travel distances are a significant barrier to PCP access in rural areas, and that access to mental health care is considerably worse in rural areas. Laditka, Laditka and Probst (2009) find that rates of preventable hospitalizations[8] are 90 percent higher in rural areas than in urban areas for patients aged 18–64, and 45 percent higher for patients over 64. In addition, Liu et al. (2008) show that rural residents are more likely to bypass[9] local primary care.

This paper is most closely related to Hurley (1991), Bolduc, Fortin and Fournier (1996), Zhou (2017), and Falcettoni (2018).[10] Hurley (1991) uses a nested logit framework to model physicians' choices of specialty, location, and type of practice. The author finds that shrinking differences in income across alternatives would cause more doctors to choose primary care but would not significantly affect location decisions. While still considering specialty choice, our model also allows physicians to choose among 200 distinct locations across the United States rather than the seven simplified "community sizes" that proxy for locations in Hurley (1991). Furthermore, our methodology allows us to address the endogeneity of salaries. Bolduc, Fortin and Fournier (1996) is somewhat closer to our paper; it focuses on the initial location decisions of general practitioners in Québec, Canada. Specifically, the authors partition Québec into 18 regions and estimate a multinomial probit model of location decisions. Both of these papers find that physicians are unresponsive to geographic differences in compensation. Hurley (1991) and Bolduc, Fortin and Fournier (1996) report income elasticities of 1.05 and 1.11 respectively; our estimate of this elasticity is 1.91. Zhou (2017) uses a dynamic discrete choice model to examine the location and specialty choices of physicians in North Carolina and simulates the effects of incentive policies. Though our model is static rather than dynamic, we model the location and specialty choice of all physicians in the country, not just those in North Carolina. Finally, Falcettoni (2018) uses a spatial equilibrium model to analyze physician location choice. Falcettoni (2018)

[8]These are also referred to as "ambulatory care sensitive conditions" in the medical literature. The diseases captured in this definition include pneumonia, urinary tract infections, and dehydration, among others.

[9]That is, use medical facilities that are not the closest to the patient's home due to the scarcity of required services (Bronstein, Johnson and Jr. (1997)).

[10]Our paper and Falcettoni (2018) were developed contemporaneously and independently of one another. Our first draft was completed in August 2017; we became aware of Falcettoni (2018) in November 2018.

focuses on differences in preferences between primary care physicians and specialists but does not endogenize the choice of specialty; in contrast, our model allows physicians to choose their specialization as well as their practice location. Specialty choice may be an important margin for policies that target PCP shortages; our counterfactual results indicate that some physicians switch from another specialty to primary care in response to the introduction of loan forgiveness programs. While Zhou (2017), Falcettoni (2018), and our paper simulate counterfactual loan repayment schemes, we also explore the effects of hypothetical revenue-neutral incentive payment policies that may be more effective at reducing shortages. Further, we use a simple differences-in-differences identification strategy to estimate the average treatment effect of current student loan repayment programs on the number of physicians in rural counties; our paper is the first to provide reduced-form evidence of the effects of these policies.[11]

Our work is also related to several papers that estimate static models of individual migration decisions. For example, Diamond and McQuade (2019) uses a spatial equilibrium model to explain increased inequality between low- and high-skilled workers over the past few decades, Piyapromdee (2019) uses a similar model to determine how native workers are affected by immigration, and Colas and Hutchinson (2018) apply the framework to study how federal income taxes affect the sorting of workers across locations. All three of these papers incorporate a static discrete choice model à la Berry, Levinsohn and Pakes (2004) to explain workers' location choices; we closely follow this modeling strategy. In addition, these papers also find that high-skilled workers are relatively unresponsive to differences in salaries across locations. For example, Diamond and McQuade (2019) reports an income elasticity for college graduates of 2.12, which is very similar to our estimate for primary care physicians.

[11] Pathman et al. (2004) survey participants in all rural physician recruitment and retention programs as of 1996, but their focus is mostly on job satisfaction and retention rates, not the effect on physician shortages.

2.3 Data

To estimate the structural model in this study, we assemble data from several different sources. The assembled data can be thought of as two distinct datasets. The first dataset, which we refer to hereafter as the physician dataset, contains demographic information as well as specialty and location decisions for individual physicians who bill to Medicare. The second dataset, which we call the location dataset, contains annual PCP salaries, observed amenities, and other characteristics of each location. The construction of both datasets is described below. First, we discuss data collection and sample selection for our differences-in-differences analysis.

2.3.1 Data for differences-in-differences analysis of student loan policies

In order to provide reduced-form evidence of the effect of incentive programs on the number of physicians practicing in rural counties, we manually collected data on the earliest loan forgiveness or repayment program[12] for primary care physicians enacted by the legislature in each state. Although some states subsequently introduced additional loan repayment programs, we use the date of the first program to evaluate the effects of these policies. Most of these data were collected from the Rural Health Information Hub, which provides a list of funding opportunities for rural communities in each state.[13] However, this list does not include the year each program was first introduced; in order to find this information, we searched for the appropriate statute in each state's legal code and legislative bill history.[14]

Table 2.6 in Appendix 2.A lists these programs. As of 2015, the District of Columbia and every state except Florida had introduced a loan forgiveness or scholarship program for PCPs practicing in shortage areas. The average amount of loan repayment offered by the

[12]We use the terms "forgiveness" and "repayment" interchangeably in this paper.
[13]This list is available at https://www.ruralhealthinfo.org/funding/states
[14]A spreadsheet with detailed source information is available upon request. We thank the librarians at the University of Wisconsin-Madison Law School for their assistance.

Table 2.1: County summary statistics

	Overall		Program 2005-2015			
			Before		After	
	Mean	(Std. Dev.)	Mean	(Std. Dev.)	Mean	(Std. Dev.)
MDs only	25.90	(48.11)	29.18	(44.69)	30.86	(47.11)
DOs only	3.63	(7.57)	2.93	(5.26)	3.49	(6.24)
All physicians	29.53	(52.62)	32.11	(48.49)	34.35	(51.84)
Doctors per 1,000 people	0.94	(0.81)	1.03	(0.82)	1.00	(0.82)
Unemployment rate	6.67%	(2.98%)	6.22%	(2.75%)	7.72%	(3.15%)
Per capita income (in $1,000)	$37.50	($10.61)	$34.66	($10.47)	$37.19	($10.53)
Log(population)	9.59	(1.07)	9.53	(1.12)	9.71	(1.03)
Population per square mile	36.43	(39.14)	33.86	(40.60)	41.45	(39.15)
Observations	18,675		1,687		3,808	

states is $98,447; typically, this amount is paid in smaller annual installments over a period of two to four years, provided that the physician maintains a practice in the designated shortage area. In nearly all cases, participation in the program is restricted to U.S. citizens. Though some states introduced programs as early as the 1970s and many did so prior to 2000, 16 states introduced polices between 2005 and 2015. As our data on the number of physicians in each county is only available for 2005 onward, these recent policy changes provide the bulk of the variation that allows us to identify average treatment effects. Table 2.1 shows summary statistics by rural counties overall (columns 1 and 2) and for counties that are treated between 2005 and 2015 both before introduction (columns 3 and 4) as well as after (columns 5 and 6). Comparing columns 3 and 5 shows an increase in the raw average number of all physicians after the program introduction relative to before. On average there are roughly 30 physicians in a rural county. The counties in treated states between 2005 to 2015 look fairly similar in terms of unemployment rate, population and per capita income to the overall average which includes untreated counties as well as counties that were treated previously.

There are two other types of programs that are targeted at the rural physician shortage: tax credits for physicians practicing in shortage areas (implemented by Alabama, Georgia, Louisiana, New Mexico, North Carolina, Oklahoma, Oregon and Maine) and a federal program that facilitates J-1 work visas for foreign doctors willing to move to underserved areas. The tax credit amounts are typically much smaller than loan repayment amounts — around \$5,000 per year for a limited number of years. Since only one state (New Mexico) introduced such a policy between 2005 and 2015, we do not have enough variation in our data to analyze tax credit policies in a differences-in-differences framework. Furthermore, since the tax credits are an order of magnitude smaller than the loan forgiveness programs, we expect that any effect of these programs would be extremely difficult to detect. Finally, since this paper is mainly concerned with the responsiveness of physicians to monetary incentives, we do not analyze the visa program here. Thus, the remainder of the paper focuses on loan forgiveness programs.

We combine our collected information on introduction years and student loan forgiveness policies by state with data from the HRSA Area Health Resources File (AHRF).[15] In particular, we use the number of active DOs and MDs in each county reported by the AHRF as outcomes of interest. These data are available for the years 2005-2006, 2008, and 2010-2016.[16] The AHRF also contains county-level information such as basic demographic characteristics and employment and labor market related information which we use. As our focus is on rural physician shortages, we drop all non-rural counties. This leaves us with 1,957 rural counties and a total of 19,555 county-year observations.

2.3.2 Sources

Since 2012, the Centers for Medicare & Medicaid Services (CMS) has published the Physician Utilization and Payment Public Use File, which contains data on services provided and payments received by all health professionals who bill Medicare for services. This file contains the street address, gender, primary medical specialty, dollar amounts billed to Medicare, and characteristics of the professional's Medicare patient pool such as the average age of patients and the fraction of patients diagnosed with cancer. Starting in 2014, CMS was also required

[15] Data published by: HRSA, Bureau of Health Workforce (BHW), National Center for Health Workforce Analysis (NCHWA)

[16] These data were obtained by the HRSA from the AMA Physician Masterfile. The HRSA never obtained the Physician Masterfile data for 2007 or 2009, which is why these years do not appear in our data.

by a provision in the Affordable Care Act to publish the Physician Compare database, which allows patients to view additional demographics and performance measures for all physicians who accept Medicare. We merge these two sources to create a repeated cross-section of physicians who graduated from medical school ten years prior to the period 2012–2016 (i.e., physicians who graduated between 2002 and 2006).[17] We define any physician who lists their primary specialty as family practice, general practice, geriatric medicine, internal medicine, obstetrics & gynecology, or pediatric medicine as a primary care physician.

Though this data gives us quite detailed information on individual physicians, it has some limitations. First, Medicare patients are typically only a fraction of all the patients a physician treats. A physician may also serve patients with private insurance, patients with insurance through another government entity (e.g., Medicaid or the Department of Veterans Affairs), and patients who are uninsured. Because of this, we cannot use the Medicare data to determine physician salaries, as doing so would underestimate a physician's total compensation. Second, not all physicians accept Medicare patients,[18] meaning we cannot use the Medicare data to determine the overall number of physicians in each location.

To address the first issue, we use data from the Occupational Employment Statistics (OES) program of the Bureau of Labor Statistics (BLS) to determine annual salaries for PCPs in each location.[19] These data contain average salaries of workers in narrowly defined occupations in each metropolitan statistical area (MSA), each state, and the nonmetropolitan

[17] We use a ten year lag for a few reasons. First, both primary care physicians and specialists typically spend three or four years in a residency program prior to practicing on their own, and most specialists also complete a three or four year fellowship program that provides additional training after their residency. As a consequence of this additional fellowship training for specialists, the ratio of PCPs to specialists in the Medicare data is non-stationary until roughly ten years after medical school graduation. Since residencies and fellowships are determined by a nationwide matching process, a physician's location just a few years after finishing their degree does not solely reflect their own locational preferences. Second, our model assumes that location decisions are once-and-for-all; physicians tend to be fairly mobile in their late twenties and early thirties but tend to settle down in their late thirties and early forties, roughly 10 years after leaving medical school (see Figure 2.A-1 in Appendix 2.A).

[18] According to the Kaiser Family Foundation, the proportion who do accept Medicare patients is approximately 93 percent. See `http://www.kff.org/medicare/issue-brief/primary-care-physicians-accepting-medicare-a-snapshot/` .

[19] Again, we cannot use the ACS or CPS for this purpose because these surveys report occupational status on too broad a level. The finest level of detail on PCP salaries we would be able to obtain would be salaries for all physicians and surgeons, which would vastly overstate the true salary of a PCP.

areas of each state. We use the average annual salary for occupation code 29-1062: Family and General Practitioners.[20] Unfortunately, the OES program does not report salaries by age or gender.

To overcome the second issue with the Medicare data and obtain a more conservative estimate of the number of active PCPs in each location, we use the Area Health Resources File from the Health Resources & Services Administration (HRSA). This file contains data taken from the American Medical Association (AMA) Physician Masterfile, a comprehensive record of all physicians in the United States.[21] In particular, we use the number of physicians in direct primary patient care excluding hospital residents and physicians over the age of 75 as our measure of the number of PCPs.[22] The Area Health Resources File also contains many other statistics on health outcomes in each county by year as well other amenities such as home prices and average rents. We merge this with data from the Census Bureau, the Bureau of Labor Statistics, the Federal Bureau of Investigation, the Federal Emergency Management Agency, and the Dartmouth Atlas of Healthcare, which results in a large dataset of county characteristics.[23]

Most of our data are published at the county level. For the differences-in-differences analysis in Section 2.4, we use the county-level data directly, as this gives us a large number of observations. However, a structural model in which physicians choose to locate in individual counties would be intractable. For the structural model in Section 2.5, we therefore partition the 3,142 counties of the United States into 200 geographic units according to metropolitan status and population. First, metropolitan statistical areas (MSAs)[24] with 500,000 or more

[20] For some MSAs, there is not enough data for the BLS to report an accurate salary estimate. In these cases, we use the statewide average salary.

[21] We wish we could use the AMA Physician Masterfile itself, as it contains even more detailed information on individual physicians than the Medicare data. Unfortunately, the Masterfile is proprietary and difficult to access.

[22] This should yield a conservative estimate of the true number of PCPs since the Physician Masterfile underestimates the number of physicians who are retired. See Petterson, Rayburn and Liaw (2016) for a thorough explanation.

[23] The amenities used as control variables in both the reduced-form analysis and the structural model are listed in footnotes of the appropriate tables of estimation results.

[24] A MSA is defined by the Office of Management and Budget (OMB) as one or more adjacent counties that contain a city of at least 50,000 people.

people in 2010 are given their own distinct location. Second, all other counties that are part of a MSA within a state are aggregated into a "small metro" location for that state. Finally, all counties within a state that are not part of a MSA are grouped together as the rural areas of that state. This process results in 104 large metro areas, 49 small metro areas, and 47 rural areas.[25]

2.3.3 Summary Statistics

Table 2.2 displays summary statistics for our physicians dataset. This contains 75,675 physicians, 26,839 of whom are primary care physicians (35 percent). Overall, 35 percent of all physicians are female. However, female physicians disproportionately choose primary care as their specialty; 52 percent of PCPs are female. Overall, 28 percent of PCPs and 30 percent of all physicians practice medicine in the same state as their medical school. These patterns are slightly stronger in rural areas, as 33 percent of rural physicians practice in their home state. Foreign physicians (i.e., those who attended medical school outside the United States)[26] are disproportionately likely to choose primary care; 43 percent of PCPs are foreigners compared to 31 percent of all physicians. In addition, DOs are a significant source of rural primary care physicians; they account for 22 percent of rural PCPs but just 12 percent of the total physician population.

Table 2.3 summarizes the characteristics of rural, small metro, and large metro locations. On average, the annual salaries of PCPs are $8,000 higher in rural areas than in urban areas. This suggests that physicians in rural areas are being paid a small compensating differential for the lack of amenities. Despite this difference in pay, PCPs disproportionately practice

[25] Some states on the East Coast do not have any counties in the small metro or rural categories. We also note that several large MSAs overlap multiple states (e.g., New York-Newark-Jersey City, NY-NJ-PA MSA and Washington-Arlington-Alexandria, DC-VA-MD-WV MSA).

[26] In some cases, the medical school is listed in the Medicare data as "Other"; in these cases, we assume the physician attended a foreign medical school. The AAMC publishes the fraction of foreign medical school graduates by specialty (`https://www.aamc.org/data/workforce/reports/458506/1-7-chart.html`). According to this table, the overall fraction of foreign PCPs is 24 percent. Table 2.2 shows that 31 percent of all physicians in our data are classified as foreign. While we recognize that this somewhat overstates the fraction of international doctors, we maintain this definition to leverage the variation between foreign and domestic physicians.

Table 2.2: Physician summary statistics, 2012–2016

		Mean	(Std. Dev.)
PCPs	Rural areas (47 locations):		
	Fraction female	0.436	(0.496)
	Fraction foreign	0.366	(0.482)
	Fraction DOs	0.223	(0.416)
	Fraction living in med school state	0.326	(0.469)
	Small metro areas (49 locations):		
	Fraction female	0.481	(0.500)
	Fraction foreign	0.444	(0.497)
	Fraction DOs	0.168	(0.374)
	Fraction living in med school state	0.269	(0.444)
	Large metro areas (104 locations):		
	Fraction female	0.541	(0.498)
	Fraction foreign	0.429	(0.495)
	Fraction DOs	0.140	(0.347)
	Fraction living in med school state	0.284	(0.451)
	Overall (200 locations):		
	Fraction female	0.518	(0.500)
	Fraction foreign	0.425	(0.494)
	Fraction DOs	0.154	(0.361)
	Fraction living in med school state	0.286	(0.452)
All physicians	Overall (200 locations):		
	Fraction PCPs	0.355	(0.478)
	Fraction female	0.397	(0.489)
	Fraction international	0.316	(0.465)
	Fraction DOs	0.121	(0.326)
	Fraction living in med school state	0.299	(0.458)

medicine in urban areas; the average population-to-PCP ratio is 1,794 to one in rural areas but 1,324 to one in large metro areas.

2.3.4 Shortage Areas

In a population that is distributed according to the national average of gender and age groups, the medical literature is in agreement that one primary care physician can handle the healthcare demand of about 1,500 people (Goodman et al., 1996; Ricketts III et al., 2007). We therefore define a shortage area as a location with a population-to-PCP ratio greater than 1,500 to one.[27] Under this definition, the median rural area in our sample has

[27]The Health Resources & Services Administration (HRSA) defines a list of shortage areas assigns a score according to the severity of the shortage. Their shortage calculations are based on a very simple premise that takes into account population, health care access, and demographics; they define a shortage area as an area with an adjusted population-to-PCP ratio of more than 3,000 to one. Since our definition is based on

Table 2.3: Location summary statistics, 2012–2016

Variable	Mean	(Std. Dev.)	5%	Median	95%
Rural areas (47 locations):					
Annual salary of PCPs	$199,581	($22,742)	$164,054	$198,100	$238,764
Median home value	$145,777	($92,163)	$78,938	$113,500	$263,600
Persons per PCP	1,794	(412)	1,056	1,791	2,513
PCP shortage	121.73	(184.10)	-115	70	367
Percent w/ college degree	20.31%	(5.69%)	13.92%	19.04%	32.63%
Percent in poverty	16%	4%	10%	15%	24%
Population density	40.76	(36.74)	4.56	37.39	100.63
Small metro areas (49 locations):					
Annual salary of PCPs	$194,513	($26,688)	$145,724	$196,600	$234,350
Median home value	$171,645	($73,523)	$101,892	$151,110	$283,450
Persons per PCP	1,354	(259)	906	1,360	1,813
PCP shortage	-82.73	(128.71)	-265	-72	59
Percent w/ college degree	26.72%	5.2%	19.79%	25.87%	36.77%
Percent in poverty	15%	3%	9%	15%	20%
Population density	152.72	(113.66)	22.18	131.86	411.62
Large metro areas (104 locations):					
Annual salary of PCPs	$192,123	($27,824)	$144,952	$192,759	$237,099
Median home value	$198,510	($104,472)	$111,629	$163,263	$443,500
Persons per PCP	1,324	(287)	974	1,270	2040
PCP shortage	-281.79	(595.51)	-1405.5	-127.5	159.5
Percent w/ college degree	30.47%	(6.6%)	19.36%	29.91%	44.06%
Percent in poverty	14%	(4%)	9%	14%	20%
Population density	546.45	(469.52)	107.79	383.86	1426.86

a PCP shortage of 70 physicians. Figure 2.1 shows a heat map of PCP shortages across the United States; comparing this to the heat map of annual salaries in Figure 2.2 shows that salaries in some rural shortage areas (e.g., Alabama, Indiana, and Ohio) are actually lower than the national average, which was $186,320 in 2014.[28]

Table 2.A-1 in Appendix 2.A shows the differences in health outcomes between shortage and non-shortage areas. Rural areas (which constitute over half of all shortage areas) have the greatest disparities in health outcomes. The means for preventable hospitalizations, percent of patients readmitted to a hospital within 30 days of release, and cardiovascular and diabetes deaths are significantly higher in shortage areas. Overall, this evidence suggests that shortage areas have worse health outcomes than non-shortage areas.

the consensus of the medical literature, we believe that a ratio of 1,500 to one is sufficiently conservative. Despite this difference, there is a very strong correlation between the localities defined as having a shortage under our definition and shortage areas based on the HRSA definition.

[28] The rural areas in these three states also happen to have relatively high housing expenditures compared to other rural areas; see Figure 2.A-2 in Appendix 2.A.

In addition, evidence that PCP salaries do not reflect shortages can be found in Figure 2.A-3 in Appendix 2.A, which shows that there is very little correlation between the size of the shortage and salaries. While rural areas have much larger shortages, PCP salaries in rural areas are very similar to those in other areas. Unsurprisingly, the largest metropolitan areas have surpluses of PCPs based on our measure, while the rural areas of Texas and Tennessee stand out as having particularly severe shortages.

2.4 Reduced-Form Evidence

In this section, we use a differences-in-differences approach to estimate the average treatment effect of state-level loan forgiveness programs on the number of physicians practicing in rural counties. We use a differences-in-differences design to control for possibly confounding variation at the state level. For example, states that introduce programs or introduce programs sooner might be states that have more rural areas, have particularly vulnerable rural populations or have the largest PCP shortages. States that introduce loan forgiveness programs sooner may also differ in unobservable ways from states that introduce these programs later. If states with rural areas that are least attractive for highly skilled individuals introduce programs, then running an OLS regression of the number of physicians on whether or not a program exists would lead to a downward biased treatment coefficient.

Our baseline empirical specification is given by

$$\text{Physicians}_{ct} = \alpha \text{Loan}_{ct} + \boldsymbol{x}'_{ct}\boldsymbol{\beta} + \mu_s + \tau_t + \epsilon_{ct}$$

where Physicians_{ct} is the total number of physicians practicing in rural county c in year t, Loan_{ct} is an indicator for whether county c is in a state with an active loan forgiveness program in year t, $\boldsymbol{x}_{ct}$ is a vector of time-varying county characteristics (e.g., population, income), μ_s are state fixed effects, and τ_t are year fixed effects. The treatment group consists of rural counties in states with a loan forgiveness or repayment program for primary care

Table 2.4: Effect of loan forgiveness programs on the number of physicians in rural counties

	(1) All physicians	(2) MDs only	(3) DOs only	(4) All physicians	(5) MDs only	(6) DOs only
Loan program	1.457*	1.235**	0.221			
	(0.738)	(0.604)	(0.197)			
Loan program (years 0–6)				1.587**	1.343**	0.244
				(0.716)	(0.594)	(0.189)
Loan program (years 7+)				3.374***	2.820***	0.555**
				(1.013)	(0.871)	(0.271)
Observations	19,555	19,555	19,555	19,555	19,555	19,555
R^2	0.648	0.630	0.435	0.648	0.630	0.435
Equality of effects F-stat				9.824	9.332	2.499
Equality of effects p-value				0.003	0.004	0.121
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are shown in parentheses and are clustered at the state level. One, two, and three stars indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Control variables include the population, labor force, per capita income, and unemployment rate at the county-year level.

physicians in year t, and the control group consists of rural counties in states that do not have an active loan forgiveness or repayment program in year t. The key assumption for identification of α, the average treatment effect, is the parallel trends assumption: in the absence of the loan program, counties in the treatment and control groups would have had parallel trends in the number of physicians over time. We will test this assumption later in our event study framework.

Table 2.4 displays the results from this differences-in-differences regression. We first estimate the average treatment effect for all physicians and then separately for MDs and DOs. Column (1) shows that the implementation of a loan forgiveness program results in approximately 1.46 more physicians practicing in each rural county, on average. Since the average number of physicians in rural counties is 29, this coefficient implies an increase of around five percent relative to the mean. Columns (2) and (3) show that MDs account for the majority of this effect. This is mostly due to the fact that DOs are a relatively small proportion of all physicians (just 13.8 percent of all rural physicians in 2016).

Several institutional factors might cause the short-run effects of loan forgiveness programs to differ from the long-run effects. First, these programs only benefit physicians with outstanding medical student loans. If many recent graduates have already paid off a substantial portion of their loans when the policy is introduced, they might not consider enrolling in the program. In this case, the effects of the policy will be small until new medical school graduates with outstanding debt apply. Second, many of the loan forgiveness programs allow physicians currently located in shortage areas to apply for loan forgiveness. If these physicians constitute a substantial proportion of program participants, then we would not see a change in the number of physicians in rural areas. Third, if medical students decide to switch intended specialties or practice locations in response to the introduction of the policy, we might see much larger effects after these students complete medical school and residency (about seven years after the introduction of the policy). Finally, if loan forgiveness programs are successful, physicians will stay in rural areas even after their two- or four-year service commitment ends. If this is the case, new physicians entering the program in later years will augment the additional physicians from previous cohorts, leading to much larger long-run effects.

To test for different effects in the short and long run, we estimate the following regression:

$$\text{Physicians}_{ct} = \alpha^{SR}\text{Loan}_{ct}^{SR} + \alpha^{LR}\text{Loan}_{ct}^{LR} + \boldsymbol{x}'_{ct}\boldsymbol{\beta} + \mu_s + \tau_t + \epsilon_{ct}$$

where Loan_{ct}^{SR} is an indicator for counties in states within six years after implementation of a loan forgiveness program, and Loan_{ct}^{LR} is an indicator for counties in states in the seventh or greater year after implementation. We then test the null hypothesis that $\alpha^{SR} = \alpha^{LR}$.

The results from this specification are shown in columns (4) through (6) of Table 2.4. In the first six years of the program, we find that the number of physicians in each county increases by 1.59 on average. In the seventh and later years, this effect more than doubles to 3.37 physicians per county. We can reject the hypothesis that $\alpha^{SR} = \alpha^{LR}$ at the one percent significance level, which indicates that there are substantial differences in the effects of loan forgiveness programs in the short and long run.

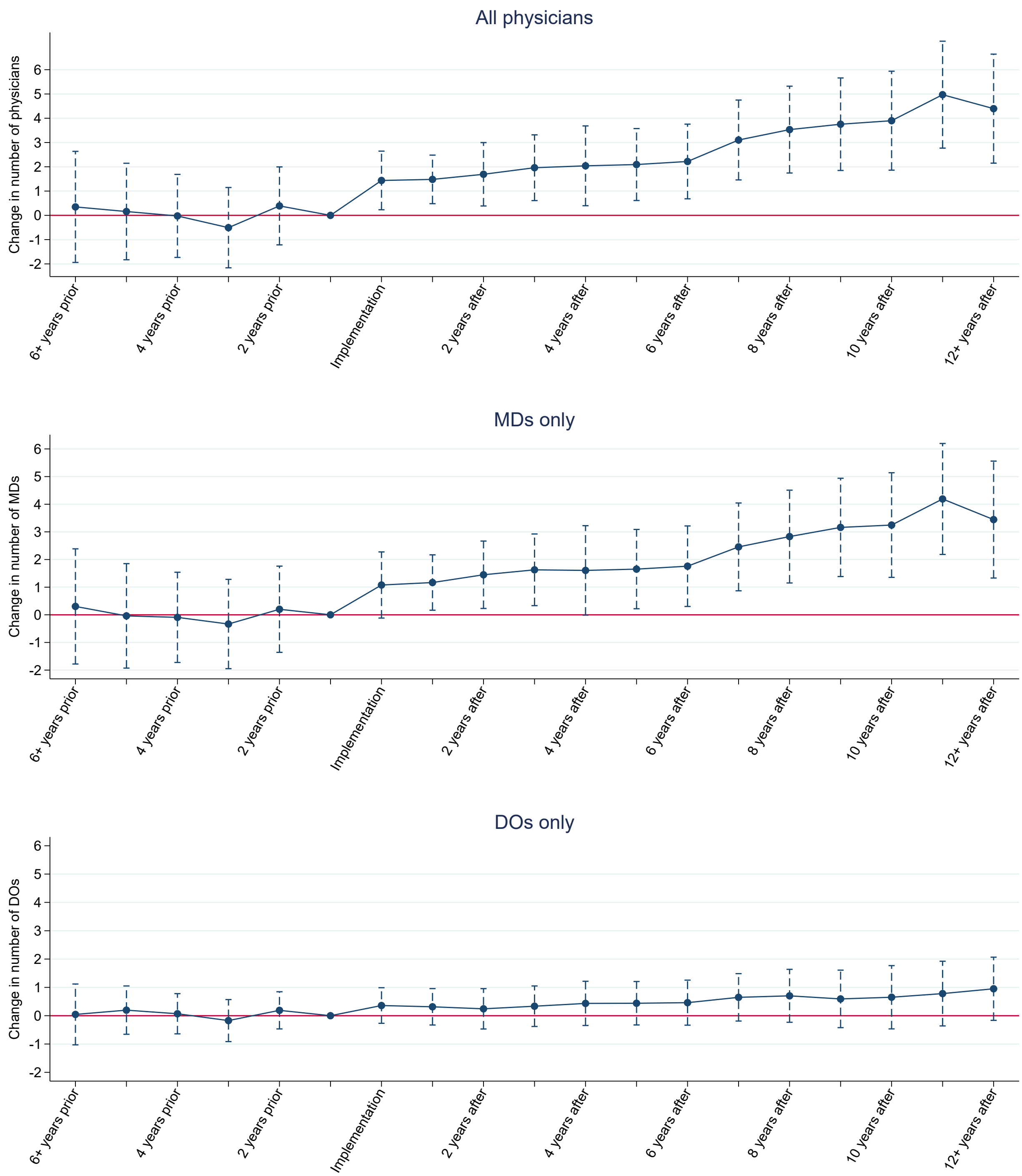
To further explore the dynamic effects of these programs, we use an event study framework to estimate the change in the number of physicians in each year after the introduction of the policy. This specification is given by

$$\text{Physicians}_{ct} = \sum_{k \in \mathcal{K}} \alpha^k \text{Loan}^k_{ct} + \boldsymbol{x}'_{ct}\boldsymbol{\beta} + \mu_s + \tau_t + \epsilon_{ct}$$

where $\mathcal{K} = \{-6, -5, \ldots, -2, 0, 1, \ldots, 12\}$. Note that the year prior to policy implementation ($k = -1$) is omitted as a normalization. In addition, $k = -6$ includes observations six or more years prior to implementation, and $k = 12$ includes observations twelve or more years after implementation. The variable Loan^k_{ct} is equal to one for counties in states with a loan forgiveness or repayment program in the kth year after the policy was implemented and is equal to zero otherwise. Thus, the effect of the loan program on the number of physicians practicing in each county in the kth year after the implementation of the policy (relative to the year prior to implementation) is given by α^k. This framework also allows us to test the parallel trends assumption for identification: if states that implemented loan forgiveness programs had similar trends in the number of physicians practicing in rural areas to states that did not implement loan forgiveness programs, then the estimated coefficients in the pre-treatment periods (i.e., α^k for $k < 0$) should be close to zero and statistically insignificant.

Figure 2.3 shows the results of this event study. As in the previous specification, we find that loan forgiveness programs immediately increase the number of physicians practicing in rural counties by about 1.5 physicians, and that this effect increases to about four physicians in the long run (i.e., in the twelfth year and beyond).[29] These estimated effects are statistically significant at the five percent level in all years. Again, we find that the majority of this overall effect is caused by an increase in the number of MDs; the estimated effects for

[29]Interestingly, there is a second distinct increase in the average treatment effect seven years after implementation. As mentioned above, the entire process of becoming a PCP takes seven years from the start of medical school; these results suggest that loan programs influence the physicians' decisions early in their medical education.



Notes: The figure displays the estimated coefficients and 95% confidence intervals from the event study specification. The top, middle, and bottom panels show the results for all physicians, MDs only, and DOs only, respectively. Each regression includes state and year fixed effects as well as the population, labor force, per capita income, and unemployment rate at the county-year level. Standard errors are clustered at the state level.

Figure 2.3: Dynamic effects of loan forgiveness programs on the number of physicians in rural counties

DOs are very small. Finally, the pre-treatment coefficients in each specification are close to zero and statistically insignificant, which supports the parallel trends assumption required for identification.

2.4.1 Robustness

In this subsection, we discuss several threats to identification and perform checks to address these concerns. The first potential concern is that our baseline results capture variation in the number of physicians across counties within states rather than the causal effect of the treatment. We address this concern by including county fixed effects in our analysis. The second potential concern is that the estimated treatment effects are a result of random noise rather than the causal effect of the treatment. We address this concern by conducting a placebo test with randomized program implementation years.

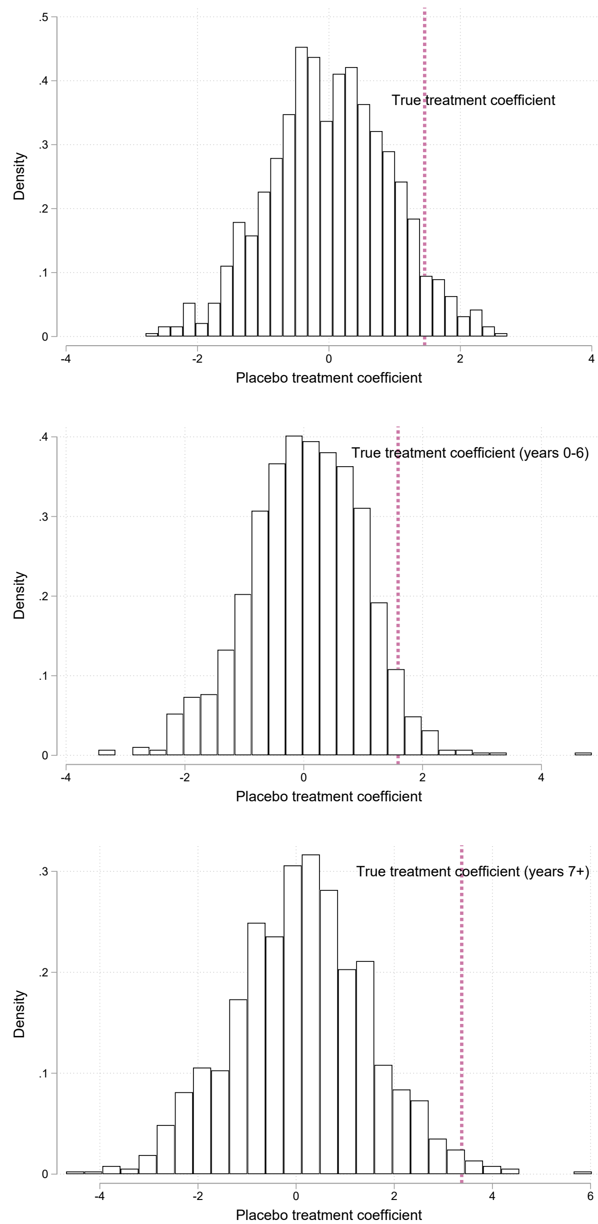
Since the loan forgiveness programs are introduced at the state level, we worry about confounding variation at the state level and our baseline empirical specification deals with these confounders by including state fixed effects that control for any time-invariant differences in the number of physicians at the state level. We also include time-varying controls at the county level to ensure that other differences occurring around the time of the introduction at the county level controlled for. Nevertheless, there is substantial variation in the number of physicians across counties within states. For example, some counties may have large hospitals that attract many physicians, while other counties may have hardly any healthcare facilities at all. To control for these time-invariant differences across counties within each state, we conduct a robustness check in which we replace state fixed effects with county fixed effects.

Table 2.A-2 in Appendix 2.A shows the results of this specification. If the treatment coefficient in Table 2.4 is driven by heterogeneity across counties, then we would expect the effect of the loan forgiveness program to disappear after county fixed effects are introduced. Table 2.A-2 shows that while the average treatment effect is much smaller than before (0.323)

and statistically insignificant, the long-run effect of the loan forgiveness program remains: column (4) shows a statistically significant increase of 1.59 physicians per county, and Figure 2.A-4 in Appendix 2.A shows that the effect of the policy is still statistically significant in years seven and beyond. We also note that the R^2 of this regression is has increased from 0.65 to 0.99, indicating that the fixed effects absorb much of the remaining variation in the number of physicians.

It is important to stress that this does not invalidate the original results that we find at the state level for treated compared to untreated states. It is possible that within a state, physicians incentivized by the loan forgiveness program are not evenly distributed across all counties so that the variation within county is noisy across the state. This is to be expected especially when one considers that different states define shortage areas differently for the purposes of the loan forgiveness program. Some states define shortage areas as all rural counties (e.g. Georgia and Oklahoma), some define them based on population (i.e. communities with fewer than 10,000 or 20,000 inhabitants such as South Dakota and Mississippi respectively) and some define them based on the HPSA shortage definition (Colorado) to name a few examples. Given this heterogeneity, we confirm that our preferred specification is the baseline specification and consider the persistent long-run effect even with county fixed effects as further evidence that we are indeed capturing causal effects of the loan forgiveness policy.

Finally, we run a series of placebo tests in which we randomize the implementation year for all states with a loan forgiveness policy. Specifically, for each state, we draw a random year from the interval that we observe programs being implemented in different states, i.e. 1978 to 2015. We then re-run our difference-in-difference and event study analyses. Figure 2.4 plots a histogram of estimated treatment coefficients for 1,000 draws of the placebo implementation year. The top panel shows the histogram for the baseline DiD coefficient. The estimated placebo treatment coefficients are centered around 0 as we would expect if the original design captures the true causal effect of the policy. 14% of the placebo coefficients



Notes: The figures display histograms of treatment coefficients for 1000 placebo draws of implementation years. The top panel shows the distribution of the baseline DiD coefficient for the placebo draws, the two lower panels show the distribution of the short-run (middle panel) and long-run DiD coefficients. We also plot the corresponding treatment coefficients for the true implementation years.

Figure 2.4: Placebo treatment effects

were statistically significant at at least the 10% level. The lower two panels plot histograms for the placebo treatment coefficients for the long-and short-run DiD coefficients. Both the placebo coefficient for 0-6 as well as the coefficient for at least 7 years after policy introduction are again centered around 0 and quite unlikely to be in the range of the coefficient we estimate on the true sample.

If our original design is capturing the causal effect of the policy introduction, we should not see any effect of the randomized implementation year in our results. Figure 2.A-5 in Appendix 2.A shows the event study with random introduction dates. None of the graphs show any systematic pattern and almost none of the coefficients are statistically different from zero. Table 2.A-3 in Appendix 2.A further confirms this. There is no difference in the long and short-run coefficients, and all coefficients are imprecisely estimated. These robustness checks give us confidence that our differences-in-differences design indeed captures the causal effect of the student loan policy.

2.4.2 Discussion

The results from our differences-in-differences analysis show that student loan forgiveness programs have a small effect on the number of physicians locating in rural counties. In the long run, our baseline specification with state fixed effects suggests a fourteen percent increase in the number of physicians per rural county. Our most conservative estimates with county fixed effects still imply a seven percent increase, mostly driven by the location choices of MDs. However, severe physician shortages persist across the country. Understanding why this is the case requires a more thorough analysis of physicians' preferences and responsiveness to incentives.

Loan forgiveness programs likely increase the number of rural PCPs in two ways: by enticing more physicians to choose primary care rather than another specialty and by attracting more primary care physicians to rural areas. We therefore develop a model of location and specialty choice that explicitly accounts for both the variation in characteristics across different locations (such as wages, rents, and amenities) as well as the heterogeneity in physicians' preferences based on individual characteristics (such as their gender and home state). Since current policies focus on monetary incentives, we pay particular attention to the estimation of physicians' sensitivity to wages.

This model will allow us to decompose the change in the number of rural physicians into those who switch specialties and those who switch locations in response to loan forgiveness programs. In addition, it will help us determine whether certain types of physicians are more responsive to monetary incentives than others. Furthermore, we can use the model to conduct counterfactual experiments in which we change the size of the monetary incentives and study how shortages in different locations evolve. We discuss this model in detail in the next section.

2.5 Model

This section introduces our model of physician specialty and location choice. It closely resembles the static discrete choice model in Berry, Levinsohn and Pakes (2004), who use data on individual purchase decisions to estimate the demand for automobiles. In the labor economics literature, this approach has been adopted by Diamond and McQuade (2019), Colas and Hutchinson (2018), and Piyapromdee (2019) to model worker migration decisions.

2.5.1 Setup

We model the joint decision of location and specialty as a static discrete choice problem. Physician i's indirect utility from practicing as a primary care physician in location $j \in \{1, 2, \ldots, J\}$ in year t is given by

$$u_{ijt} = \delta_{jt} + \eta_{ijt} + \epsilon_{ijt}$$

where

$$\delta_{jt} = \alpha_w \ln(w_{jt}) + \boldsymbol{x}'_{jt}\boldsymbol{\gamma} + \kappa_j + \lambda_t + \xi_{jt}$$

$$\begin{aligned}\eta_{ijt} = {} & \text{Female}_i \times [\beta_{Female} + \alpha_{Female} \ln(w_{jt})] \\ & + \text{DO}_i \times [\beta_{DO} + \alpha_{DO} \ln(w_{jt})] \\ & + \text{Foreign}_i \times [\beta_{Foreign} + \alpha_{Foreign} \ln(w_{jt})] \\ & + \beta_{State}\text{State}_{ij}\end{aligned}$$

$$\epsilon_{ijt} \overset{iid}{\sim} \text{T1EV}.$$

Indirect utility consists of three components. The first component, δ_{jt}, varies across locations and years but does not vary across physicians. The second and third components, η_{ijt} and ϵ_{ijt}, vary across physicians, locations, and years. Within δ_{jt}, w_{jt} is the average annual salary of primary care physicians in location j in year t, $\boldsymbol{x}_{jt}$ is a vector of time-varying observed amenities and housing expenditures, κ_j are location fixed effects that capture any time-invariant unobserved amenities in each location, λ_t are year fixed effects, and ξ_{jt} represents any remaining amenities that vary across locations and time and are unobserved by the econometrician. Within η_{ijt}, Female_i, DO_i, and Foreign_i are indicators for whether physician i is female, has a Doctor of Osteopathic Medicine degree, or is from a foreign country. These characteristics generate heterogeneity in preferences for primary care relative to other specialties and in sensitivity to wages. The variable State_{ij} indicates whether location j is in the same state as physician i's medical school, which captures home bias for domestic physicians. Since we do not observe birth or native states in our data, we instead use medical school states.[30] Finally, ϵ_{ijt} is an i.i.d. Type 1 Extreme Value idiosyncratic preference shock.

The indirect utility of choosing a specialty other than primary care (regardless of location[31]) is normalized to $u_{i0t} = \epsilon_{i0t}$, where ϵ_{i0t} is another i.i.d. Type 1 Extreme Value

[30] While this is not a perfect proxy for nativity, medical school location may itself be an important factor in location decisions; Burfield, Hough and Marder (1986) report that a large fraction of physicians practice medicine in the state in which they received their medical education.

[31] Since we are primarily interested in the location choices of primary care physicians, we do not model the

preference shock. Under the distributional assumptions on ϵ_{ijt}, the probability that physician i chooses location j in year t is

$$s_{ijt}(\boldsymbol{\delta}_t, \boldsymbol{\theta}) = \frac{\exp(\delta_{jt} + \eta_{ijt})}{1 + \sum_{k=1}^{J} \exp(\delta_{kt} + \eta_{ikt})}$$

where $\boldsymbol{\theta} = (\beta_{Female}, \alpha_{Female}, \beta_{DO}, \alpha_{DO}, \beta_{Foreign}, \alpha_{Foreign}, \beta_{State})'$. The share of all primary care physicians choosing to practice in location j in year t is calculated by aggregating over the distribution of physician characteristics (i.e., Female_i, DO_i, Foreign_i, and State_{ij}), so that

$$s_{jt}(\boldsymbol{\delta}_t, \boldsymbol{\theta}) = \sum_i s_{ijt}(\boldsymbol{\delta}_t, \boldsymbol{\theta}) \cdot \Pr(\text{Female}_i, \text{DO}_i, \text{Foreign}_i, \text{State}_{ij}),$$

and the share of specialists in year t is equal to $s_{0t}(\boldsymbol{\delta}_t, \boldsymbol{\theta}) = 1 - \sum_{j=1}^{J} s_{jt}(\boldsymbol{\delta}_t, \boldsymbol{\theta})$.

Finally, we acknowledge that a more intuitive and realistic way of modeling salaries would allow for individual heterogeneity in salaries in a given location over time: w_{ijt} rather than w_{jt}. While we would prefer this over our current specification, we are constrained by the data available to us. Precise individual-level income data for physicians that is also representative at a fine geographic level is very difficult to obtain. The Centers for Medicare and Medicaid Services (CMS) reports Medicare reimbursements by physician, but this data does not include income from private patients. Similarly, the American Community Survey (ACS) contains individual-level information but does not distinguish primary care physicians from surgeons and other specialists. For the purposes of this paper, the most important aspect of salaries is how they vary by location, so we use data on average salaries for primary care physicians by location from the Occupational Employment Statistics (OES) program of the Bureau of Labor Statistics. Though this program does not report salaries by demographic group, we are confident that it gives us a better description of the true compensation of primary care physicians than any other publicly available data.

location choice for specialists.

2.5.2 Estimation, Identification, and Results

This section describes our estimation procedure, which uses techniques developed in Berry, Levinsohn and Pakes (1995) and Berry, Levinsohn and Pakes (2004). Specifically, we use a two-step estimation procedure which has also been used by Goolsbee and Petrin (2004) and Train and Winston (2007) in the industrial organization literature as well as Diamond and McQuade (2019), Colas and Hutchinson (2018), and Piyapromdee (2019) in the migration literature.

Step One: Maximum Likelihood

In the first step of estimation, we recover the vector of common components of utility for each location in each year ($\boldsymbol{\delta}$) and the idiosyncratic preference parameters ($\boldsymbol{\theta}$) via maximum likelihood.

Let d_{ijt} be an indicator that takes the value one if physician i (observed in year t) makes practice choice $j \in \{0, 1, 2, \ldots, J\}$. The log-likelihood function could be formulated as

$$\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\theta}) = \frac{1}{N} \sum_{t=1}^{5} \sum_{i=1}^{N_t} \sum_{j=0}^{J} d_{ijt} \ln(s_{ijt}(\boldsymbol{\delta}_t, \boldsymbol{\theta})). \tag{2.1}$$

However, we do not maximize equation (2.1) directly because it would require a search over $5J + \dim(\boldsymbol{\theta})$ parameters, which is infeasible for $J = 200$. Instead, we use a strategy first proposed by Berry (1994), who showed that for every value of $\boldsymbol{\theta}$, there exists a unique vector $\boldsymbol{\delta}$ that matches the estimated aggregate location shares, $s_{jt}(\boldsymbol{\delta}_t, \boldsymbol{\theta})$, to the observed shares, s_{jt}. Berry, Levinsohn and Pakes (1995) show that $\boldsymbol{\delta}$ can be characterized as the unique fixed point of the contraction mapping

$$T(\delta_{jt}) = \delta_{jt} + [\ln(s_{jt}) - \ln(s_{jt}(\boldsymbol{\delta}_t, \boldsymbol{\theta}))]$$

for every j and t. This convenient contraction mapping allows us to reformulate the log-likelihood function as

$$\mathcal{L}(\boldsymbol{\delta}(\boldsymbol{\theta}), \boldsymbol{\theta}) = \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N_t} \sum_{j=0}^{J} d_{ijt} \ln(s_{ijt}(\boldsymbol{\delta}_t(\boldsymbol{\theta}), \boldsymbol{\theta}))$$

which allows us to maximize over only $\boldsymbol{\theta}$.

Step Two: Two-Stage Least Squares

Having recovered the common component of utility for each location and the idiosyncratic preference parameters in step one, we now estimate the parameters of mean utility. The common component of utility is given by

$$\delta_{jt} = \alpha_w \ln(w_{jt}) + \boldsymbol{x}'_{jt}\boldsymbol{\gamma} + \kappa_j + \lambda_t + \xi_{jt}$$

Estimation of this equation is relatively straightforward, with one exception: wages are likely to be correlated with unobserved amenities, creating an endogeneity problem between $\ln(w_{jt})$ and ξ_{jt}. Ordinary least squares estimation would therefore give us biased estimates of α_w, our key parameter.[32]

We use two techniques to address this endogeneity problem. First, we include location fixed effects in the regression, which absorb any time-invariant unobserved amenities in each location. These fixed effects limit the endogeneity concern to unobserved amenities that vary *within* locations over time. We use an instrumental variables technique to address this potential remaining endogeneity. Any instrument z_{jt} must be uncorrelated with unobserved amenities but correlated with wages. Formally, such an instrument z_{jt} must satisfy the

[32] To be precise, all location-specific amenities are potentially correlated with unobserved amenities, meaning the coefficients on these amenities in the utility function could be biased. Obtaining unbiased estimates would require at least as many instruments as endogenous amenities. The most important coefficient for our purposes is the sensitivity to wages, α_w. We therefore focus on getting an unbiased estimate of α_w by instrumenting for log salaries and controlling for other observed amenities.

following conditions for identification:

$$\mathbb{E}[z_{jt}\xi_{jt}|\boldsymbol{x}_{jt}] = 0 \tag{2.2}$$

$$\mathbb{E}[z_{jt}\ln(w_{jt})|\boldsymbol{x}_{jt}] \neq 0. \tag{2.3}$$

In the empirical industrial organization literature, cost shocks are commonly used as instruments to identify demand; exogenous shifts and rotations in supply allow the econometrician to separately identify demand. In our case, we use an exogenous shifter of demand in order to identify the labor supply curve for PCPs in each location. In particular, we use the log of total births as an instrument for log physician salaries. Since the location fixed effects absorb any time-invariant heterogeneity across locations, the remaining variation in our instrument and in salaries is *within* locations over time. The exogeneity condition given in equation (2.2) therefore requires that changes in log births must be uncorrelated with changes in unobserved amenities in that location. Similarly, the relevance condition in equation (2.3) requires that changes in log births are correlated with changes in log salaries for PCPs.

A typical pregnancy and birth in the United States involves significant interaction with the healthcare system. For most women, this includes frequent checkups with their PCP or obstetrician-gynecologist (OB-GYN) prior to delivery. Delivery itself is a major healthcare event, usually requiring a brief hospital stay. After birth, both mother and baby will have frequent appointments with their PCP and pediatrician.[33] Given these demands on the healthcare system, we should expect that changes in the number of births are positively correlated with changes in salaries of PCPs.

One might be concerned that changes in births might be correlated with shocks to unobserved amenities in a given location, such as natural disasters or changes in air pollution. For example, Evans, Hu and Zhao (2010) find a strong negative relationship

[33] For this reason, both OB-GYNs and pediatricians are often classified as primary care physicians; we also adopt this definition.

between hurricane severity and births, and Rogers, St. John and Coleman (2005) find an increase in fertility after the Oklahoma City bombing. On the other hand, Udry (1970) finds that births did not change after the New York City blackout of 1965. Similarly, while anecdote suggests that a local sports team winning a league championship is associated with a spike in births nine months later (e.g., see Eltagouri (2017)), the scientific evidence is mixed at best (Hayward and Rybińska, 2017). In any case, to address these remaining endogeneity concerns, we include a large set of time-varying observed amenities in $\boldsymbol{x}_{jt}$, including economic and demographic characteristics such as employment, per-capita income, population, net migration, and dummy variables for fifteen categories of federal disaster declarations.[34] After controlling for these time-varying characteristics, we are confident that the remaining variation in births is random and hence uncorrelated with unobserved amenities.[35]

Table 2.A-4 in Appendix 2.A displays the results of the first-stage regression of log salaries on log births, the other observed amenities, location fixed effects, and year fixed effects. The relevance condition is clearly satisfied; the coefficient on log births is statistically significant at the one percent level. The magnitude of the coefficient indicates that a one percent increase in total births is associated with a 0.98 percent increase in PCP salaries, on average.

Estimation Results

Table 2.5 displays the parameter estimates from our model. We report the estimated coefficent on log salaries (α_w) from both two-stage least squares (IV) and ordinary least squares (OLS) estimation. If unobserved amenities are negatively correlated with wages, we would expect the IV estimate of the wage coefficient to be larger than the OLS estimate. When we do not instrument for wages, the estimated coefficient is close to zero (0.083) and statistically insignificant. When we instrument for log wages using log births, the coefficient is equal to

[34] See Table 2.A-4 in Appendix 2.A for a full list of these variables.

[35] We cannot use the variation from the reduced-form analysis as an instrument because loan forgiveness programs have no direct effect on wages.

Table 2.5: Estimation results for structural model

	Step One	Step Two	
	MLE	IV	OLS
η_{ijt}:			
State$_{ij}$	3.311***		
	(0.274)		
Female$_i$	3.014***		
	(0.313)		
DO$_i$	−5.023***		
	(0.172)		
Foreign$_i$	3.601***		
	(0.281)		
Female$_i$ × ln(w_{jt})	−0.180		
	(3.807)		
DO$_i$ × ln(w_{jt})	0.454		
	(2.088)		
Foreign$_i$ × ln(w_{jt})	−0.187		
	(3.413)		
δ_{jt}:			
Log salary: ln(w_{jt})		2.037**	0.083
		(1.029)	(0.126)
Number of physicians	75,675		
Log-likelihood	−2.163		
Location-year observations		1,000	1,000
Number of locations		200	200
R^2		0.889	0.920
Other amenities		Yes	Yes
Location fixed effects		Yes	Yes
Year fixed effects		Yes	Yes

Notes: Standard errors are shown in parentheses; the OLS and IV standard errors are clustered at the location level. One, two, and three stars indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Other amenities include net domestic migration, net international migration, percentage of good air quality days, dummies for fifteen categories of federal disaster declarations, and the logarithms of annual average rent, population, population under 25, population over 65, population in poverty, female population of child-bearing age, labor force, unemployed workers, total college graduates, per-capita income, number of violent crimes, number of property crimes, and number of entertainment establishments at the location-year level.

2.037 and statistically significant at the five percent level.

The results show that physicians strongly prefer to practice in their home state — the magnitude of the coefficient on State$_{ij}$ indicates that physicians would need to be paid an additional \$9,338 per year on average in order to forgo living in their home state.[36] The point

[36]This number is calculated by finding the increase in salary for each physician that would hold expected utility constant if β_{State} were set to zero.

estimates also indicate that female and foreign physicians are less sensitive to differences in salaries across locations, while DOs are more sensitive; however, these coefficients are not statistically significant.

The results imply a median own-wage elasticity of 1.91, meaning a one percent increase in annual salaries in a given location yields a 1.91 percent increase in the fraction of PCPs who choose to practice medicine in that location. This estimate is similar to the estimates for college-educated workers reported by Diamond and McQuade (2019) and Piyapromdee (2019), which are 2.12 and 2.09, respectively. Our estimate is also consistent with the results from other papers that estimate models of physician location decisions. For example, Hurley (1991) finds an average income elasticity of 1.05, and Bolduc, Fortin and Fournier (1996) report an average income elasticity of 1.11, both of which are contained in the 95 percent confidence interval for our estimate.

Figure 2.A-6 in Appendix 2.A shows a histogram of own-wage elasticities for all 200 locations in 2016. Most locations have own-wage elasticities of around 1.9. All elasticities lie between 1.8 and 2.03, indicating that the shares of physicians in even the most elastic locations are fairly unresponsive to changes in salaries. Since large cities tend to have the fewest close substitutes, they also tend to have the smallest own-wage elasticities; Table 2.A-5 in Appendix 2.A lists the ten locations with the smallest own-wage elasticities in 2016.

2.5.3 Counterfactual Incentive Payment Policies

We now use the estimated model to simulate several policies intended to eliminate rural physician shortages. To supplement our differences-in-differences analysis, we first simulate the effects of current loan forgiveness programs and decompose the change in the number of rural physicians by specialty, gender, and medical school location. We then modify the loan forgiveness programs by allowing foreign physicians to receive loan forgiveness as well as by doubling the amount of forgiveness. Finally, we analyze the effects of revenue-neutral policies that tax primary care physicians who practice in urban areas and subsidize those who practice in rural areas.

To simulate the effects of loan forgiveness programs, we find the time-discounted increase in salaries that is equivalent to the lifetime savings generated by the loan forgiveness program. We assume that all physicians leave medical school with \$200,000 in outstanding student debt with an annual interest rate of 6.25 percent.[37] Since most physicians are able to defer payment until after the completion of their residency, we assume that physicians do not make any debt payments during residency but accrue interest during that time period.[38] We then assume that physicians adopt a standard ten-year repayment plan with flat payments each year, and that any loan forgiveness is realized at the beginning of this ten year window. Due to the effects of compound interest, \$100,000 of loan forgiveness generates \$137,482 in lifetime savings. This amount of savings is equivalent to a time-discounted increase in annual salaries of \$9,653 per year.[39] Similarly, \$200,000 of loan forgiveness generates \$329,809 in lifetime savings, which is equivalent to a \$23,157 increase in annual salaries.

Table 2.6 displays the number of new physicians by specialty and location before and after the introduction of the \$100,000 loan forgiveness program for rural primary care physicians.[40] A total of 53 additional physicians choose to be a rural primary care physician in response to the policy. Of these, 36 switch their specialty to primary care, and 17 PCPs switch from practicing in an urban area to practicing in a rural area. In percentage terms, the number of rural PCPs increases by 4.92 percent in response to this policy. This effect is very similar in magnitude to our reduced-form estimate; our event study indicated that the number of physicians in rural counties increased by about 1.5 physicians per county in the first year of the policy, which is about five percent of the average number of physicians per county.

[37] The Association of American Medical Colleges (AAMC) reports that the median amount of student loan debt among recent graduates is approximately \$200,000; see `https://students-residents.aamc.org/advocacy/article/student-debt/`.

[38] We assume all PCP residencies take three years.

[39] We assume that physicians will work for 30 years and that their annual discount factor is equal to $\frac{1}{1+r}$, where $r = 0.0625$. The choice of discount factor does not substantively affect our results.

[40] As in IHS Inc. (2016), we assume a total of 28,223 new physicians enter the workforce each year.

Table 2.6: Counterfactual number of new physicians by location and specialty

	No program	\$100,000 loan forgiveness	Policy effect (difference)	Policy effect (% change)
By location type (within primary care):				
Rural areas	1,077	1,130	53	4.92
Small metro areas	1,959	1,955	−4	−0.20
Large metro areas	7,054	7,041	−13	−0.18
By specialty:				
Primary care	10,090	10,126	36	0.36
Other	18,143	18,107	−36	−0.20

Table 2.7: New rural PCPs by gender, medical degree, and medical school location

	No program	\$100,000 loan forgiveness	Policy effect (difference)	Policy effect (% change)
By gender:				
Males	495	521	26	5.25
Females	582	609	27	4.64
By medical degree:				
MDs	897	938	41	4.57
DOs	180	192	12	6.67
By medical school location:				
In-state	229	249	20	8.73
Out-of-state	848	881	33	3.89

Table 2.7 shows the increase in the number of rural PCPs by gender, medical degree, and medical school location. The effects for males and females are very similar; the policy causes about a five percent increase in the number of rural PCPs of each gender. The results for DOs and MDs echo our differences-in-differences results; as there are many fewer DOs overall, MDs account for the majority of the increase in rural PCPs (41 out of 53 additional physicians). Finally, we find that physicians who attended an in-state medical school are more sensitive to loan forgiveness policies; the number of in-state graduates practicing in rural areas increases by 8.7 percent, compared to just 3.9 percent for out-of-state graduates. Since physicians strongly prefer to practice in the state in which they went to medical school, it is much more difficult to attract physicians from outside the state to practice in rural areas.

Figure 2.5 shows the percentage of the PCP shortage eliminated per year as a result of loan forgiveness policies for rural locations with a shortage of at least 200 PCPs in 2016.

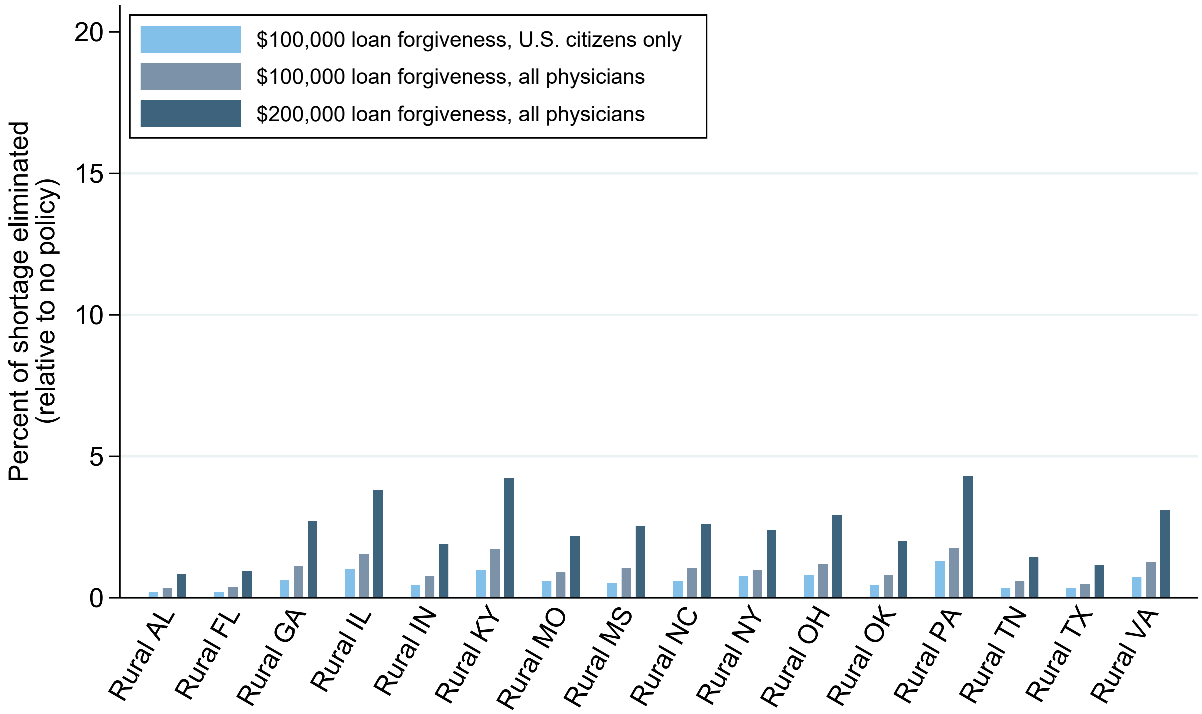


Figure 2.5: Effect of loan forgiveness programs on PCP shortages

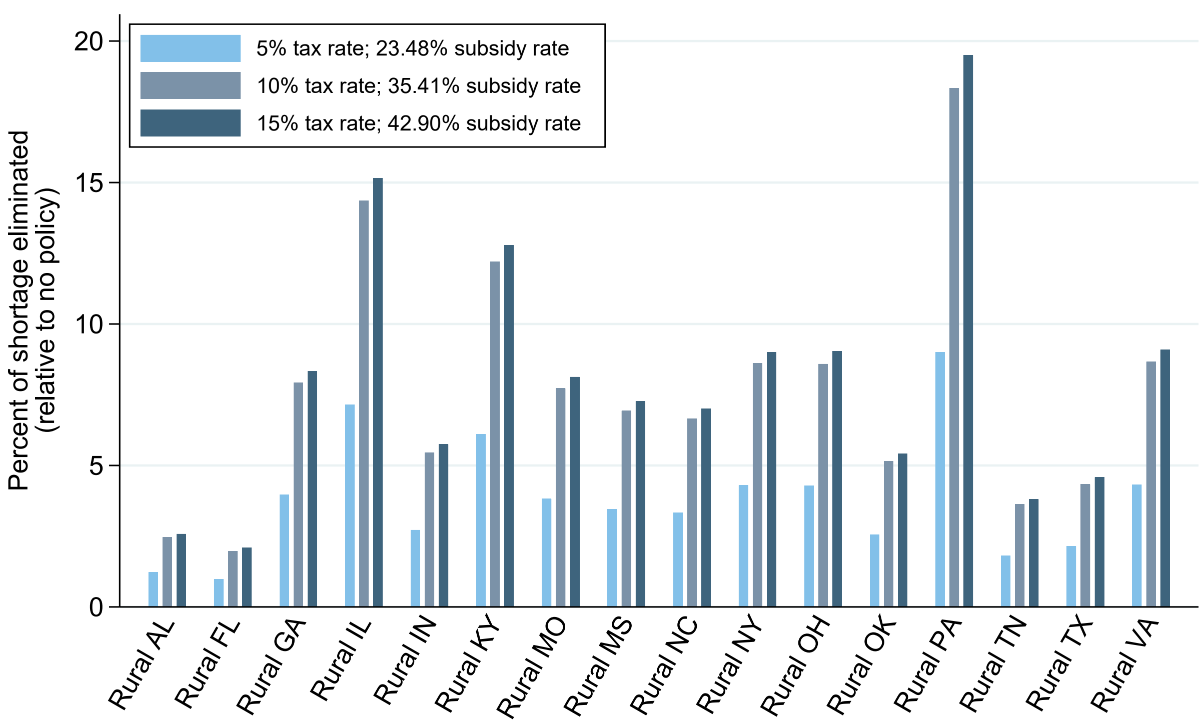


Figure 2.6: Effect of revenue-neutral policies on PCP shortages

In addition to the current \$100,000 loan forgiveness policies that are limited to U.S. citizens only, we also show the effects of two alternative policies that extend loan forgiveness to foreign physicians and double the amount of loan forgiveness to \$200,000. The results show that loan forgiveness programs do not have a substantial effect on shortages. The current loan forgiveness programs reduce shortages by less than one percent per year; allowing foreign doctors to receive loan forgiveness would have almost no additional impact on shortages, and doubling the amount of loan forgiveness to \$200,000 would reduce shortages by only about three percent per year in most locations.

Given the lack of effectiveness of loan forgiveness programs, we now simulate the effects of a nationwide revenue-neutral policy that taxes new PCPs who practice in urban areas and subsidizes those who practice in rural areas. In contrast to loan forgiveness programs, these polices make urban areas less attractive for PCPs in addition to making rural areas more attractive. Put simply, physicians who practice in rural areas get carrots, while those who do not get sticks. Since a much larger share of PCPs choose to practice in urban areas than in rural areas, a small tax rate implies a much larger revenue-neutral subsidy rate. For example, a five percent tax rate in urban areas allows a 23.48 percent subsidy rate in rural areas, which generates a \$46,862 increase in annual PCP salaries in rural areas; this is five times the present discounted salary increase implied by the \$100,000 loan forgiveness program.

Figure 2.6 displays the percentage of the PCP shortage eliminated after one year for the same rural locations as in Figure 2.5. These policies appear to be much more effective at reducing shortages; a ten percent tax rate and corresponding 35.41 percent revenue-neutral subsidy rate reduces shortages by between five and ten percent in most locations. However, there appear to be diminishing returns to these policies; the fifteen percent tax rate scheme generates very little additional reduction in shortages relative to the ten percent tax rate scheme. Nonetheless, these results suggest that a nationwide revenue-neutral policy could be more effective at reducing shortages than the current loan forgiveness policies.

2.6 Conclusion

In this paper, we provide both reduced-form and model-based evidence on the effects of incentive payment programs on the number of physicians practicing in rural areas. Using a differences-in-differences identification strategy, we find that student loan forgiveness programs increase the number of physicians by about three physicians per rural county in the long run. We then estimate a structural model of physician specialty and location decisions and use it to simulate the effects of counterfactual incentive policies. We find that physicians are relatively unresponsive to differences in salaries across locations and strongly prefer to practice medicine in their home state. As a result, current incentive payments are too small to eliminate shortages.

In light of these results, we suggest that policymakers who wish to reduce shortages of primary care physicians recruit more physicians who were raised in shortage areas to stay and practice medicine there. Given the strength of physicians' preferences for living close to their home location, attracting physicians from out-of-state to move to shortage areas in another state appears to be extremely difficult. In addition, a nationwide revenue-neutral policy that taxes physicians in urban areas and subsidizes those in rural areas might be more effective at reducing shortages than current policies.

There are other alternatives that might also improve access to healthcare in areas with physician shortages. One option is to entice more medical school graduates to choose a career in primary care rather than another specialty; we find that increasing the compensation of PCPs relative to specialists increases the number of physicians who choose primary care.[41] Another option is to allow other well-qualified but less-educated medical professionals such as nurse practitioners to perform more tasks usually completed by physicians. Traczynski and Udalova (2018) find that increasing the total number of primary care providers in this way improves health outcomes and decreases administrative costs. Finally, telemedicine (the use

[41] Schwartz et al. (2005) also note that more physicians might choose primary care if the pay gap between it and other specialties was narrowed.

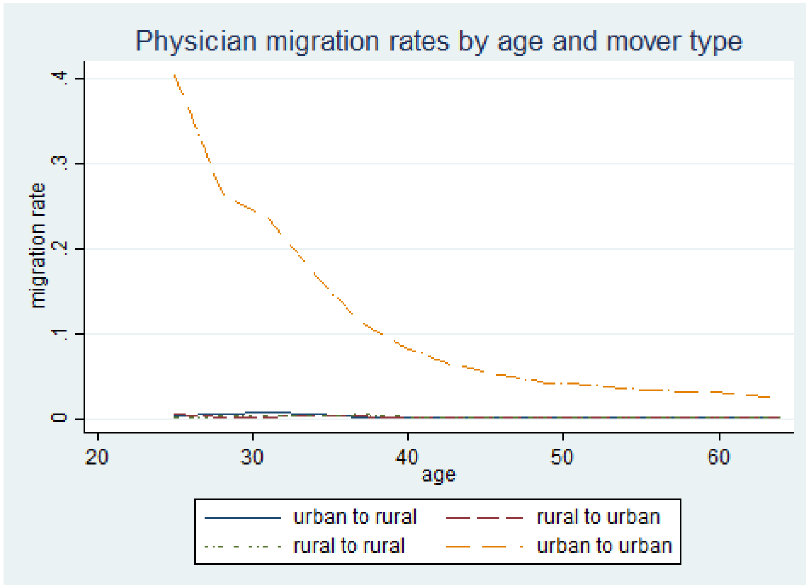
of remote communications technology to diagnose, monitor, and treat patients) may alleviate primary care shortages (Kvedar, Coye and Everett, 2014; Weinstein et al., 2014). However, there is a lack of research on the accuracy of diagnoses using such methods (Armfield et al., 2014), and many physicians seem reluctant to practice remotely. While there seems to be great potential for telemedicine for some medical procedures, there is also evidence that care provided through telemedicine is of lower quality than traditional care (Uscher-Pines et al., 2016). Additional research is needed to understand the effectiveness of these potential alternatives to traditional primary care.

There are several directions in which this research can be extended. While our model is static, migration is certainly a dynamic decision; physicians may choose to relocate many times over their careers for a variety of reasons. At the present time, we do not have a long enough panel of physicians to estimate a dynamic model of migration as in Kennan and Walker (2011).[42] In addition, more accurate data on the salaries of individual primary care physicians would allow us to explore the trade-offs between the choice of specialty and location in more detail. Finally, we note that our measure of the demand for physicians (and hence the shortage of physicians) only accounts for differences in population. The demand for primary care physicians likely varies according to differences in income, age, insurance status, and other factors across locations. Changes in healthcare demand in different locations over time may also drastically change the magnitude of shortages. For example, population decline may be so rapid in some areas that physician shortages may disappear without any further policy changes. On the other hand, an area with a rising proportion of elderly residents may see an increase in demand for healthcare in the short term but a decline over a longer time horizon as these residents die. Other factors such as competition in the insurance marketplace and idiosyncratic characteristics such as the population's lifestyle and overall attitudes toward healthcare may also play a role in demand. As a result, modeling the demand for primary care physicians will require a great deal of thought and presents

[42] This dynamic approach may be possible in the future if CMS continues to release the Physician Compare and Utilization and Payment datasets or if we obtain access to the AMA Physician Masterfile.

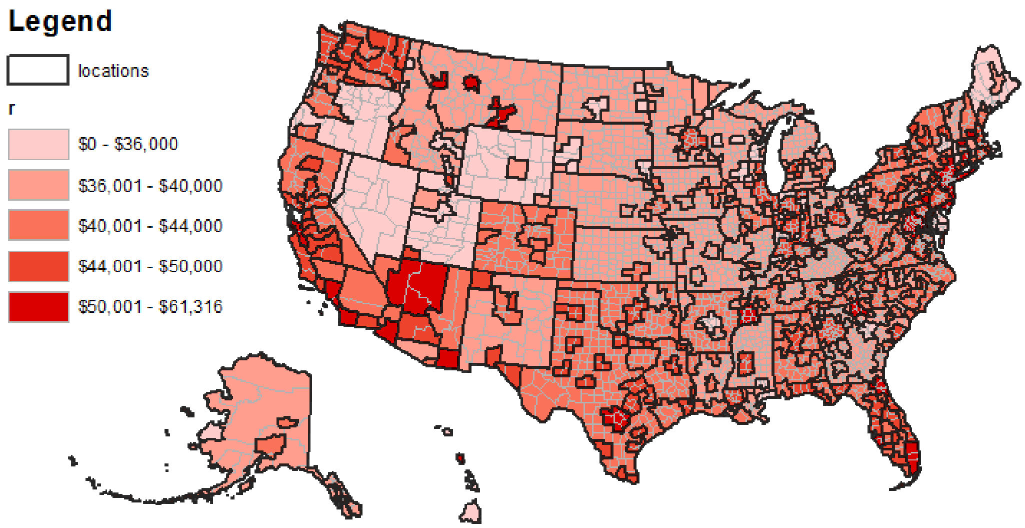
another opportunity for future work.

2.A Additional Tables and Figures



Source: ACS, 2010–2014

Figure 2.A-1: Physician migration rates by age and mover type



Source: ACS, 2010–2014

Figure 2.A-2: Annual housing expenditure for PCPs by location

Table 2.A-1: Health outcomes for shortage vs. non-shortage areas

	Non-shortage areas		Shortage areas	
	Mean	(Std. Dev.)	Mean	(Std. Dev.)
Rural areas (47 locations):	(12 locations)		(35 locations)	
Prev. hosp. per 1000	45.47	(9.42)	62.82†	(16.13)
% readmitted within 30 days	13.19	(1.10)	14.52†	(0.93)
% ER visit within 30 days	19.93	(1.44)	20.08	(1.71)
Nurses per 1000	2.58	(0.83)	2.31	(0.61)
Cardiovascular deaths per 1000	1.11	(0.17)	1.32†	(0.31)
Influenza deaths per 1000	0.26	(0.07)	0.30	(0.07)
Diabetes deaths per 1000	0.30	(0.07)	0.40†	(0.09)
Pop. in mental hospital per 1000	0.07	(0.11)	0.10	(0.15)
Small metro areas (49 locations):	(37 locations)		(12 locations)	
Prev. hosp. per 1000	47.18	(11.36)	53.56	(13.56)
% readmitted within 30 days	14.20	(0.78)	14.08	(1.52)
% ER visit within 30 days	19.99	(1.28)	19.63	(1.86)
Nurses per 1000	3.34	(1.47)	3.24	(0.94)
Cardiovascular deaths per 1000	1.02	(0.23)	1.11	(0.28)
Influenza deaths per 1000	0.18	(0.06)	0.19	(0.06)
Diabetes deaths per 1000	0.26	(0.07)	0.29	(0.07)
Pop. in mental hospital per 1000	0.13	(0.20)	0.05†	(0.06)
Large metro areas (104 locations):	(83 locations)		(21 locations)	
Prev. hosp. per 1000	49.43	(11.66)	49.41	(12.34)
% readmitted within 30 days	14.47	(0.97)	14.66	(1.33)
% ER visit within 30 days	19.77	(1.27)	19.96	(1.40)
Nurses per 1000	2.76	(1.01)	2.31	(0.80)
Cardiovascular deaths per 1000	0.98	(0.22)	0.90	(0.29)
Influenza deaths per 1000	0.17	(0.06)	0.15†	(0.05)
Diabetes deaths per 1000	0.25	(0.07)	0.25	(0.10)
Pop. in mental hospital per 1000	0.10	(0.16)	0.19	(0.63)
Overall (200 locations):	(132 locations)		(68 locations)	
Prev. hosp. per 1000	48.44	(11.40)	57.05†	(15.67)
% readmitted within 30 days	14.28	(0.99)	14.49	(1.17)
% ER visit within 30 days	19.84	(1.28)	19.96	(1.63)
Nurses per 1000	2.90	(1.17)	2.47†	(0.81)
Cardiovascular deaths per 1000	1.00	(0.22)	1.16†	(0.35)
Influenza deaths per 1000	0.18	(0.06)	0.23†	(0.09)
Diabetes deaths per 1000	0.25	(0.07)	0.33†	(0.11)
Pop. in mental hospital per 1000	0.10	(0.17)	0.12	(0.37)

Sources: Dartmouth Atlas of Healthcare and HRSA Area Health Resources File, 2014

† - Shortage area mean statistically different from non-shortage area mean at 5% level.

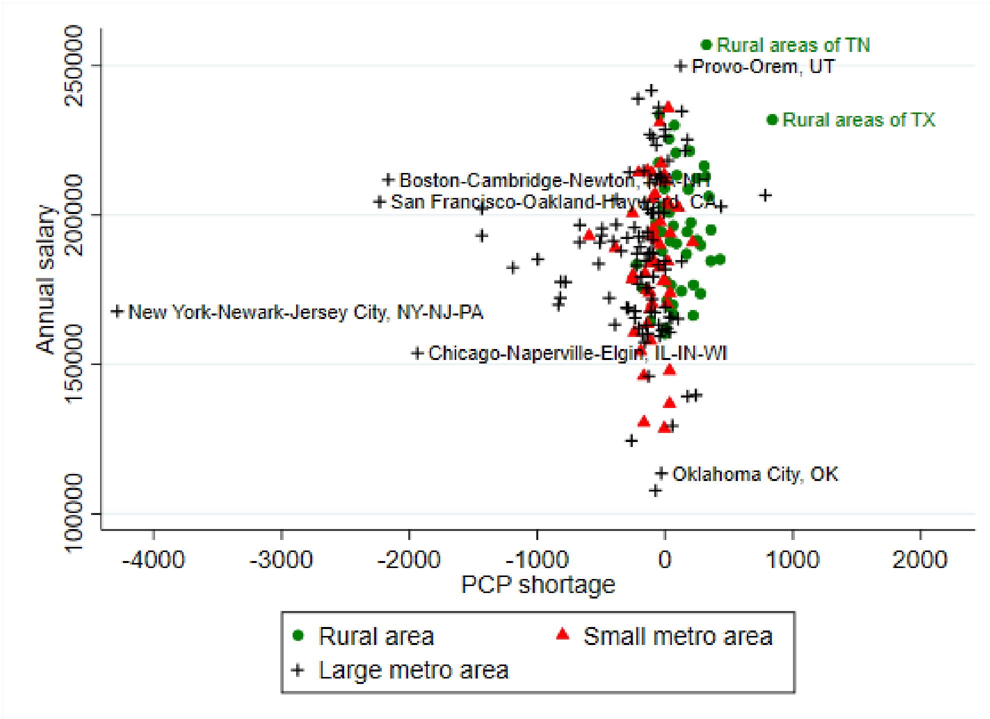


Figure 2.A-3: PCP shortage and average annual salaries by location and metropolitan status

Table 2.A-2: Robustness check — county fixed effects

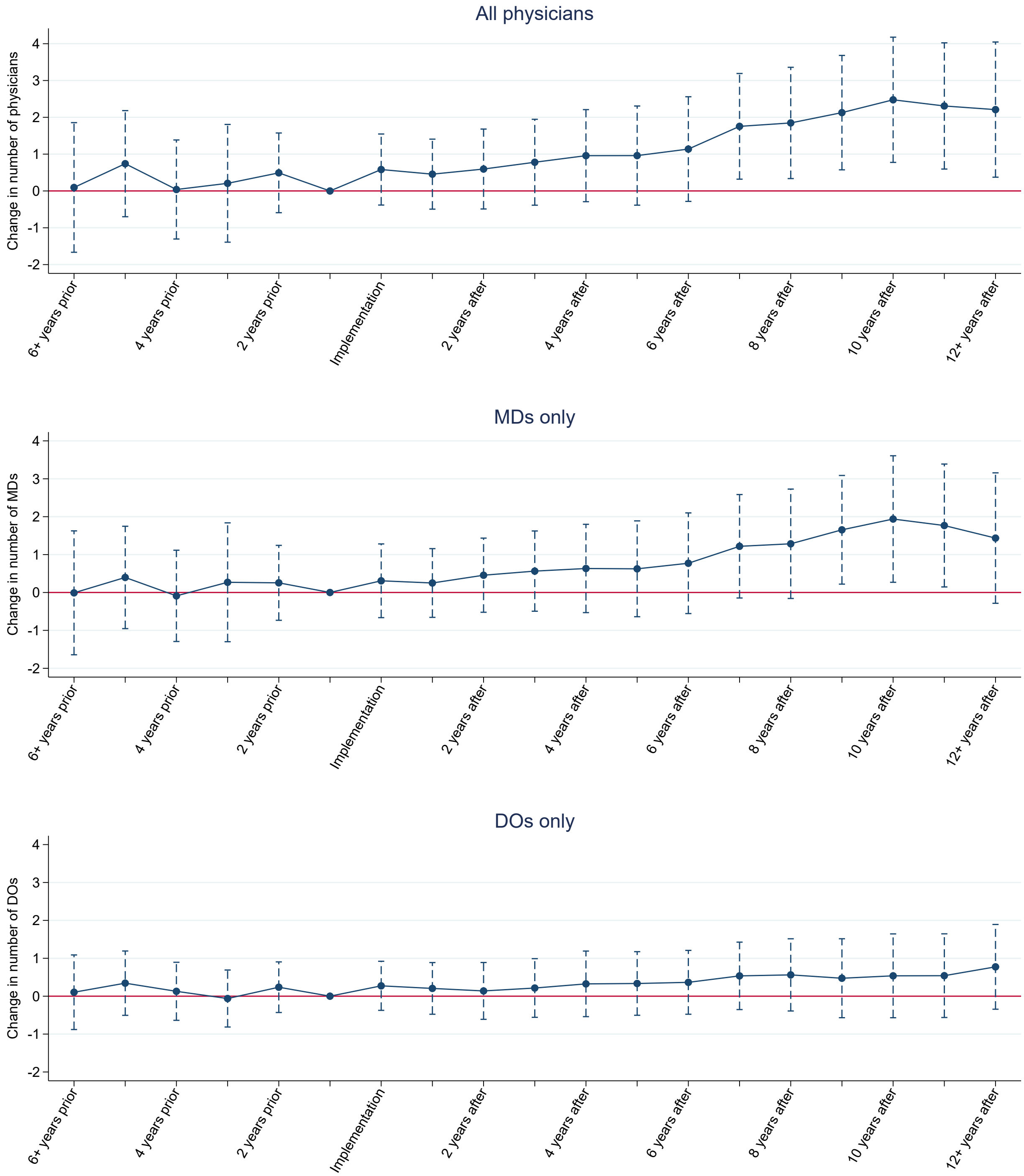
	(1) All physicians	(2) MDs only	(3) DOs only	(4) All physicians	(5) MDs only	(6) DOs only
Loan program	0.323	0.261	0.062			
	(0.482)	(0.424)	(0.164)			
Loan program (years 0–6)				0.405	0.324	0.081
				(0.465)	(0.422)	(0.156)
Loan program (years 7+)				1.593**	1.232*	0.360
				(0.689)	(0.633)	(0.255)
Observations	19,555	19,555	19,555	19,555	19,555	19,555
R^2	0.991	0.991	0.974	0.991	0.991	0.974
Equality of effects F-stat				12.630	9.667	2.181
Equality of effects p-value				0.001	0.003	0.147
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are shown in parentheses and are clustered at the state level. One, two, and three stars indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Control variables include the population, labor force, per capita income, and unemployment rate at the county-year level.

Table 2.A-3: Placebo test — random implementation years

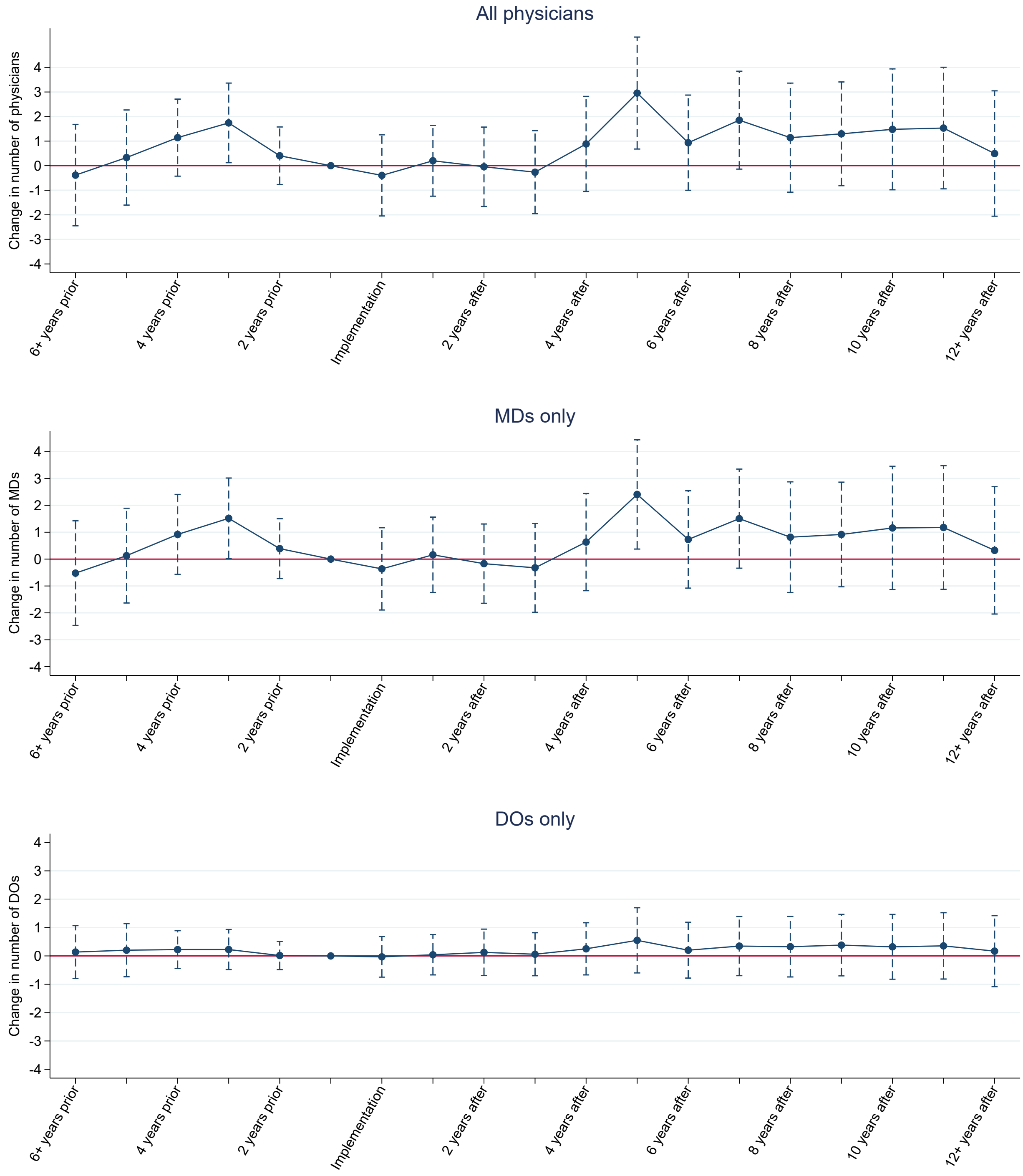
	(1) All physicians	(2) MDs only	(3) DOs only	(4) All physicians	(5) MDs only	(6) DOs only
Loan program	-0.015	-0.055	0.040			
	(0.764)	(0.653)	(0.161)			
Loan program (years 0–6)				0.018	-0.032	0.050
				(0.760)	(0.653)	(0.161)
Loan program (years 7+)				0.352	0.202	0.150
				(1.179)	(1.016)	(0.262)
Observations	19,555	19,555	19,555	19,555	19,555	19,555
R^2	0.648	0.630	0.435	0.648	0.630	0.435
Equality of effects F-stat				0.162	0.114	0.273
Equality of effects p-value				0.689	0.737	0.604
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are shown in parentheses and are clustered at the state level. One, two, and three stars indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Control variables include the population, labor force, per capita income, and unemployment rate at the county-year level.



Notes: The figure displays the estimated coefficients and 95% confidence intervals from the event study specification with county rather than state fixed effects. The top, middle, and bottom panels show the results for all physicians, MDs only, and DOs only, respectively. Each regression also includes year fixed effects as well as the population, labor force, per capita income, and unemployment rate at the county-year level. Standard errors are clustered at the state level.

Figure 2.A-4: Robustness check — event study with county fixed effects



Notes: The figure displays the estimated coefficients and 95% confidence intervals of the event study from the placebo test using randomly assigned implementation years. The top, middle, and bottom panels show the results for all physicians, MDs only, and DOs only, respectively. Each regression includes state and year fixed effects as well as the population, labor force, per capita income, and unemployment rate at the county-year level. Standard errors are clustered at the state level.

Figure 2.A-5: Placebo test — event study with random implementation years

Table 2.A-4: First-stage regression of log PCP salaries on log births and controls

	Log PCP salary	
Log births	0.983***	(0.342)
Log rent	0.314	(0.302)
Net domestic migration, thousands	0.000	(0.001)
Net international migration, thousands	−0.002	(0.002)
Good air quality days, percent	−0.001	(0.001)
Log population	−3.553	(2.236)
Log population under 25	1.188	(0.863)
Log population over 65	0.184	(0.358)
Log population in poverty	0.155	(0.130)
Log female population of child-bearing age	−0.415	(1.144)
Log labor force	0.106	(0.566)
Log unemployed workers	−0.065	(0.088)
Log college graduates	0.614	(0.535)
Log per capita income	0.151	(0.305)
Log number of violent crimes	−0.110*	(0.064)
Log number of property crimes	0.166**	(0.084)
Log number of entertainment establishments	−0.077	(0.397)
Chemical spill	−0.026	(0.075)
Tropical storm	0.062***	(0.022)
Earthquake	−0.043***	(0.010)
Fire	−0.017	(0.033)
Flood	0.010	(0.014)
Hurricane	−0.040**	(0.019)
Ice strom	−0.005	(0.027)
Landslide	−0.053*	(0.028)
Severe thunderstorm	−0.003	(0.009)
Snow storm	0.013	(0.023)
Terrorist attack	0.028	(0.028)
Tornado	0.097*	(0.056)
Toxic substance disaster	−0.005	(0.014)
Volcanic eruption	−0.042*	(0.022)
Other disaster	−0.142***	(0.020)
Constant	25.190***	(9.440)
Observations	1,000	
R^2	0.648	
Location fixed effects	Yes	
Year fixed effects	Yes	

Notes: Standard errors are shown in parentheses and are clustered at the location level. One, two, and three stars indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

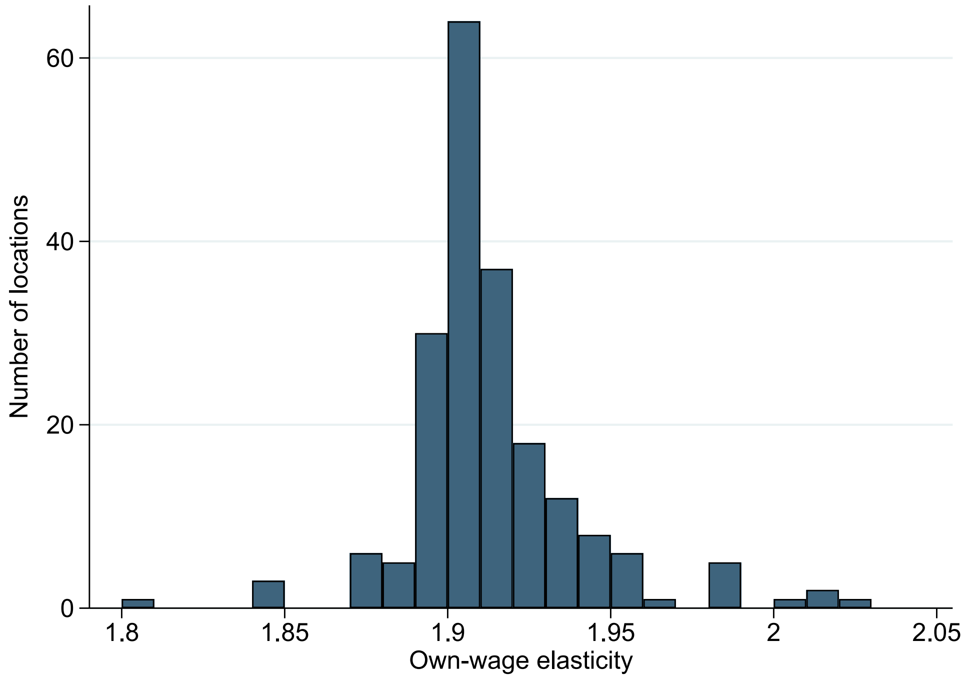


Figure 2.A-6: Frequency histogram of own-wage elasticities, 2016

Table 2.A-5: Locations with smallest own-wage elasticities, 2016

Rank	Location	Own-wage elasticity
1	New York-Newark-Jersey City, NY-NJ-PA	1.803
2	Los Angeles-Long Beach-Anaheim, CA	1.843
3	Washington-Arlington-Alexandria, DC-VA-MD-WV	1.848
4	Chicago-Naperville-Elgin, IL-IN-WI	1.849
5	Seattle-Tacoma-Bellevue, WA	1.876
6	Baltimore-Columbia-Towson, MD	1.876
7	Atlanta-Sandy Springs-Roswell, GA	1.877
8	Dallas-Fort Worth-Arlington, TX	1.877
9	Minneapolis-St. Paul-Bloomington, MN-WI	1.877
10	Detroit-Warren-Dearborn, MI	1.880

Note: All elasticities above represent the percentage change in the share of PCPs choosing the location that would result from a one percent increase in salaries at that location.

Table 2.6: Earliest loan forgiveness or repayment program implemented by state

State	Year	Program name	Amount*
AL	1982	Alabama Board of Medical Scholarship Awards	Varies†
AK	2010	Alaska Supporting Health Care Access through Loan Repayment Program	$70,000
AZ	2013	Arizona State Primary Care Provider Loan Repayment Program	$87,000
AR	1998	Arkansas Community Match Rural Physician Recruitment Program	$80,000
CA	1987	California State Loan Repayment Program	$50,000
CO	2009	Colorado Health Service Corps Provider Loan Repayment Program	$90,000
CT	1990	Connecticut State Loan Repayment Program	$60,000
DE	2000	Deleware State Loan Repayment Program	$100,000
DC	2006	DC Health Professional Loan Repayment Program	$120,000
GA	2010	Georgia Physicians for Rural Areas Assistance Program	$100,000
HI	2012	Hawaii State Loan Repayment Program	Varies‡
ID	2000	Idaho Rural Physician Incentive Program	$100,000
IL	2005	Illinois National Health Service Corps Loan Repayment Program	$50,000
IN	1993	Indiana Health Care Professional Recruitment and Retention Fund Program	$80,000
IA	1994	Iowa PRIMECARRE Loan Repayment Program	$50,000
KS	1991	Kansas Bridging Plan	$26,000
KY	2003	Kentucky State Loan Repayment Program	$80,000
LA	1991	Louisiana State Loan Repayment Program	$90,000
ME	2009	Doctors for Maine's Future Scholarship	$100,000
MD	1992	Maryland State Loan Repayment Program	$200,000
MA	1990	Massachusetts Loan Repayment Program for Health Professionals	$50,000
MI	1978	Michigan Essential Health Provider Recruitment Strategy	$200,000
MN	2003	Minnesota Rural Physician Loan Forgivenenss Program	$100,000
MS	2007	Mississippi Rural Physicians Scholarship Program	$120,000
MO	1988	Missouri Health Professional State Loan Repayment Program	$50,000
MT	1991	Montana Rural Physician Incentive Program	$150,000
NE	1991	Nebraska Loan Repayment Program for Rural Health Professionals	$180,000
NV	1989	Nevada Health Service Corps	Varies‡
NH	1979	New Hampshire State Loan Repayment Program	$115,000
NJ	1991	Primary Care Physician & Dentist Loan Redemption Program of New Jersey	$120,000
NM	1978	New Mexico Health Professional Loan Repayment Program	$50,000
NY	2008	Doctors Across New York Physician Loan Repayment Program	$150,000
NC	2011	North Carolina State Loan Repayment Program	$100,000
ND	1994	North Dakota State Loan Repayment Program	$100,000
OH	1995	Ohio Physician Loan Repayment Program	$120,000
OK	2010	Oklahoma Physician Loan Repayment Program	$160,000
OR	2013	Oregon Health Care Provider Loan Repayment Program	$105,000
PA	1992	Pennsylvania Primary Care Loan Repayment Program	$100,000
RI	1993	Rhode Island Educational Loan Repayment Program for Primary Care Providers	$80,000
SC	1989	South Carolina Rural Physician Incentive Grant Program	$100,000
SD	2012	South Dakota Recruitment Assistance Program	$219,000
TN	1988	Tennessee State Loan Repayment Program	$50,000
TX	1993	Texas Physician Education Loan Repayment Program	$160,000
UT	2015	Utah Rural Physician Loan Repayment Program	$30,000
VT	2005	Vermont Educational Loan Repayment Program for Health Care Professionals	$40,000
VA	1994	Virginia State Loan Repayment Program	$140,000
WA	1989	Washington State Health Professional Loan Repayment Program	$75,000
WV	1989	West Virginia State Loan Repayment Program	$90,000
WI	1989	Wisconsin Health Professions Loan Assistance Program	$100,000
WY	1993	Wyoming State Loan Repayment Program	$40,000
		Mean:	$98,447
		Median:	$100,000

* - Denotes the *maximum* amount of loan forgiveness a physician can receive in the program.

† - One year of service entitles the physician to repayment of one school year of student loans.

‡ - Amount of award is contingent on availability of funds.

Chapter 3

The Effect of Learning the Language on Job Market Outcomes: Evidence from "Integration Courses" in Germany

Chapter Summary

This paper evaluates the effects of Germany's "Integration course" - a 600 hour language class - on language learning, and uses the introduction of this class in 2004 as variation to investigate the causal effect of language proficiency on labor market outcomes for immigrants. I use data from the German Socio Economic Panel to estimate a fuzzy regression discontinuity model. The program is mandatory for some immigrant cohorts but it is open to almost all others as well on a voluntary basis. This allows me to separately study cohorts that are more and less exposed to the language policy. I find large and robust increases in the probability of full-time employment and in monthly income for immigrants that stem from increases in proficiency of spoken German. For refugees, I find that chances of being employed either part-time or full-time double for cohorts that are more exposed to the introductory language class. Finally, I find that among cohorts that were already in the country at introduction, the most recently arrived immigrants also gain language proficiency due to the language class.

3.1 Introduction

The large numbers of refugees arriving in Germany in 2015 challenged German administrative authorities in many different ways - one of them being the complete over-booking of introductory German language classes for immigrants (called "integration courses") offered by the Federal Office for Migration and Refugees (BAMF). The popularity of these classes points to the importance attributed to learning the local language for successful integration and assimilation into the culture and labor market of the host country.

This paper investigates the effects of language skills of immigrants to Germany on their job market outcomes, in particular, income and probability of employment. I focus on the change in language skills at the time of the introduction of the "Integrationskurs" (integration course) in 2005 - a 600 hour (mine months of intensive classes) language learning program for immigrants to Germany that also includes a (60 hour) cultural orientation class. The aim is to bring immigrants to a language proficiency level as classified under the Common European Framework as B1 (an independent user of the language who can handle day-to-day interactions in German easily as well as communicate on some more complex topics as well.) Integration courses are meant to ease participation in daily life in Germany on a personal as well as professional level.

The subject of language and immigrant assimilation has received increasing attention in policy making as well as economic literature over time. This paper is related to a growing literature on assimilation of immigrants to their host country, specifically assimilation in terms of job market outcomes, a literature on language and immigration and the beginning of a literature of evaluation of introduction programs for immigrants.

Since Chiswick (1978), economists have studied the assimilation of immigrants to natives in terms of their earnings. Borjas (1985), Borjas (1995) and Borjas (2015) build on the early literature by emphasizing the importance of cohort effects and return migration which may bias assimilation coefficients in crosssectional regressions. Borjas (2015) argues that the decline in the "quality" of immigrant cohorts to the US can be explained at least partially through lower rates of English proficiency. Card (2005) considers a different measure of assimilation: he focuses on the success of children of immigrants. In this paper, my aim is not to analyze assimilation of immigrants relative to natives, instead I focus on immigrants before and after a policy change as well as immigrants who were treated differently under the policy.

Various strategies have been employed to solve the endogeneity problem when studying the effect of language on earnings that arises since immigrants choosing to invest in language skills may differ systematically from their peers that choose not to. Chiswick and Miller (1995) model fluency in the language of the destination country as a function of economic incentives, exposure and efficiency and use other observable characteristics as instruments for language. Dustmann and Fabbri (2003) compare OLS, IV and propensity score estimates of the effect of language on employment probability. They find that OLS estimates of language on earnings significantly underestimate the actual effect. Dustmann and van Soest (2002) investigate the endogeneity of language skills of immigrants in two ways and discuss possible measurement error for the language variable measured in the German Socio Economic Panel (GSOEP) which I use as well. They suggest to instrument for language in various ways as a solution to this problem. Dustmann (1997) considers various determinants of both the writing and speaking proficiency of immigrants. Carliner (2000) investigates determinants of English fluency for immigrants to the US. Tubergen and Kalmijn (2005) consider immigrants of the same origin countries in different destination countries and study characteristics of the receiving societies and how these affect language proficiency of immigrants. This paper uses a new source of exogenous variation to estimate the causal effect of language proficiency on labor market outcomes, namely the introduction of large-scale language classes for immigrants in Germany. I demonstrate that the results are robust to various empirical specifications that deal with the age-period-cohort problem in immigration research. I am able to study the effects of language speaking and also provide evidence on differential impacts of language learning on refugees - a population that we know relatively little about.

The papers most closely related to this paper analyze introductory programs for immigrants in different countries. Evidence from an introduction program in Sweden (Svantesson and Aranki, 2006) suggests that the language component of the program has no effect while the job market component significantly increases the probability of being employed. A second Swedish program focuses on different intensities of labor market coaching that are randomly assigned to new and eligible immigrants (Joona and Nekby, 2012).

Most recently, Lochman, Rapoport and Speciale (2016) discuss the effects of a language and civil training program for non-European immigrants to France in 2007. They employ a regression discontinuity design using the score of an initial assessment test that the immigrants have to take. They find positive effects on the probability of employment and having a permanent contract as well as a negative effect on having a full time contract. They conjecture that most of the effect they find from language classes might come from a signalling effect rather than actually being more fluent in French.

The German integration course has not been the target of a lot of research by economists so far, probably due to the lack of data. In 2011 the Federal Office for Migration and Refugees (Schuller, Lochner and Rother, 2011) did a descriptive analysis focused on the effectiveness of language learning through the integration course. In a sample of participants in the integration course they find that language skills of participants do increase and remain high even a year after the class has been completed. In contrast to introductory classes for immigrants in other European countries, the German integration course focuses on language learning as opposed to specific training for the labor market. Analyzing this particular class therefore is an opportunity to study the effect of language proficiency without other confounding components.

I use data from the German Socio Economic Panel (GSOEP) from 1984-2017 to study the impact of the German integration classes and understand the effect of language proficiency on labor market outcomes. Since actual participation in the program of each individual is unknown, I use instruments that exploit the timing and eligibility criteria of the language policy to identify local average treatment effects. For immigrants arriving after 2004 I find substantial increases in language speaking proficiency (about 3%) due to the introduction of the program. These increases seem to be driven mostly by proficiency gains for non-EU citizens. Increases in language speaking proficiency are much higher - around 30% - for refugees. Cohorts that were less likely to be mandated to take the language class but arrived only a few years before the introduction of language classes also seem to have benefited. I find substantial increases both in the probability of full-time employment and monthly income as a result of the program. Increases in employment probability are about 12.5% for immigrants with around 10% increases in monthly household income. For refugees, the probability of being employed (either full-or part-time) doubles for cohorts arriving after 2005. The results highlight the importance of language classes for refugees and emphasize that gaining proficiency in language is a major contributing factor to the labor market success of immigrants that are mandated to take formal language classes.

The paper proceeds as follows: Section 2 discusses the introduction of integration classes in Germany, section 3 lays out the empirical strategy and challenges, section 4 describes sample selection and the data source, section 5 shows the baseline first stage results as well as results for subgroups and robustness, section 6 discusses second stage instrumental variable results again both for the baseline and by subgroups and section 7 concludes.

3.2 Introduction of integration classes in Germany

In 2004, Germany overhauled the "Aufenthaltsgesetz" (Residence Act) which defines different groups of immigrants and their residence status. The main difference is between immigrants from EU countries and immigrants from non-EU countries. In 2004 the Residence Act was restructured to reduce its previous complexity and to reflect free movement of labor across member states of the European Union. Citizens of EU member states no longer required any form of residence permit to live and work in Germany. An exception was made for the 10 new members of the European Union (Poland, Slovenia, Slovakia, Estonia, Latvia, Lithuania, Czech Republic, Hungary, Malta and Cyprus) who were admitted in 2004. Germany exercised the option to postpone free movement of labor for citizens of these countries until 2011. As a consequence, this particular change in the Residence Act had very little effect in 2004. Immigrants from Western European countries could already live and work in Germany easily but required a document which was no longer the case after 2004.

A bigger implication of the change in law in 2004 was the introduction of integration courses. These are 600 hour language classes and a 60 hour cultural class meant to help immigrants achieve a language proficiency level of B1 [1] in about 9 months of intensive classes. According to aggregate data from the BAMF about integration courses, 61% of participants who complete the course, finish with a B1 level, 31% finish with an A2 level and 8% finish with an A1 level. Over time the proportion of those finishing with B1 level has increased and those finishing with A1 level has decreased. In 2015, reflecting the Syrian refugee crisis, the 10 most represented countries among the participants were: Syria, Poland, Rumania, Bulgaria, Itlay, Turkey, Greece, Iraq, Spain and Hungary.

Eligibility is determined as follows: citizens from non EU countries, receiving their residence permit in or after 2005 (which is equivalent to immigrating to Germany in or after 2005) and with a residence permit of at least one year, were legally mandated to participate in the class unless they could demonstrate sufficient fluency in German. Immigrants who received their residence permit before 2005 and are citizens of non-EU countries or naturalized German citizens were not legally mandated to take the class but could enroll in it if there was sufficient capacity. Citizens of EU countries working in Germany were also not legally mandated or entitled to take the integration class but could register if there was space. Finally, old immigrants from EU and non-EU countries as well as naturalized German citizens who were born in a different country can be mandated to take the integration class if they

[1] The framework used to measure language here is the Common European Framework for Reference for Languages. It divides users into three categories: Basic user (level A), independent user (level B) and proficient user (level C). These categories are further subdivided into a lower and a higher level. The ranking of levels from lowest to highest is: A1, A2, B1, B2, C1, C2.

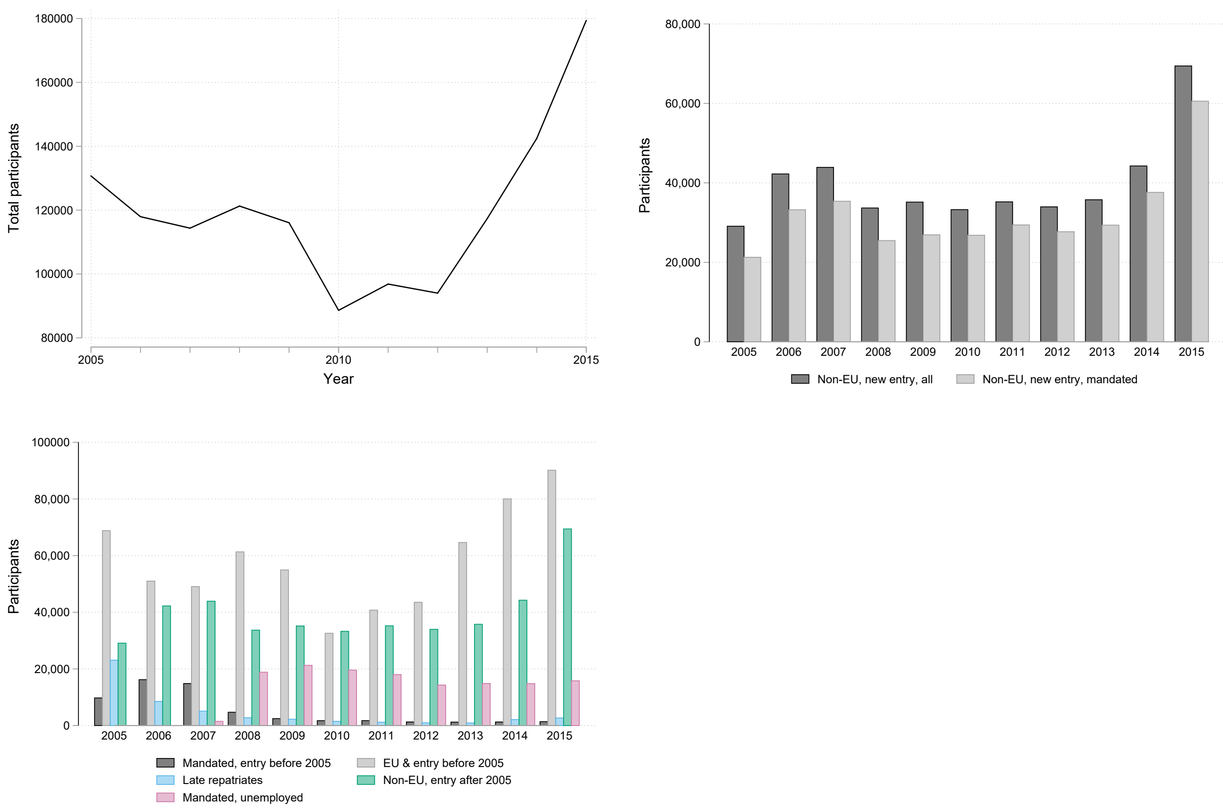
receive unemployment insurance. Not taking an integration class when mandated can have consequences related to renewal of the visa status.

New immigrants, when registering with local authorities would receive information about registering for integration classes. Other immigrants would have to fill up a form and could then be admitted to take the class. The law took effect in January 2005. Figure 3.1 shows the development of participation in integration courses since the introduction of the program. Initially the numbers were high. This is expected since many immigrants who were already in the country in 2005 would learn about the program and enroll. Over time we would expect the numbers to drop to the level of mostly the new immigrants and some old immigrants who were mandated by employment laws but not too many other old immigrants. By 2010, numbers dropped quite a bit. But by 2012 numbers picked up dramatically - mainly due to the increase in the number of refugees.

Not all immigrants who are issued a permission to take the class, actually take it. Over time the take up rate for those immigrants who were offered to take the integration class has been between 60 and 65%. There has been an upwards trend in take up in the past couple of years.

Panel 2 of Figure 3.1 shows a break down of the types of participants in the integration course over time. The biggest group throughout is the group of EU citizens and old immigrants who immigrated before 2005. They consistently make up around half of the participants. The numbers are initially high as old immigrants learn about the course and then drop down before picking up again in 2011. This is probably mostly due to EU citizens from countries that newly gained free movement of labor (10 Eastern European countries in 2011, Bulgaria and Rumania in 2014). This group is followed closely in size by the group of new immigrants who entered after 2005 and are from non-EU origin countries. Consistently, this group does not increase very much during the years when the new EU member countries gain access to the labor market, but it picks up afterwards in the wake of the refugee crisis. The three other groups are relatively small and consist of immigrants arriving before 2005 who were

Figure 3.1: Aggregate statistics of the integration course



Source: BAMF statistics

The top left figure shows the total number of participants over time. The top right figure shows the overall number of non-EU participants who immigrated to Germany in 2005 or later as well as the number of those participants that were mandated to take the class. The bottom left panel shows the overall number of participants broken down into the following groups: Non-EU citizens immigrating 2005 or later (further divided in the top right panel), EU and non-EU citizens immigrating before 2005 who participated in the class voluntarily, Late repatriates participating voluntarily and participants that were mandated to participate either by the Federal Employment Agency or for other reasons.

mandated to take the class most likely due to lacking language skills, immigrants entering at any time who are mandated to take the class due to being long term unemployed and late repatriates. [2]

Panel 3 of Figure 3.1 shows the break down of new immigrants from non-EU countries who participated in the course into the entire number and those among the entire group who were mandated to participate. The blue bars represent the total number of non-EU citizens immigrating in a given year and taking the integration course and the red bars represent the number of the total that was mandated to take the class. This graph shows that among the participants who are the main target of the policy, a high proportion of the participants is mandated to take the class.

The composition of the participating migrants is a mixture of mandated individuals and individuals joining voluntarily. If participation of individuals in the program was known, one would be worried about two types of biases since the participating group is not random. For those participating voluntarily, one might be worried about positive selection bias which would bias OLS estimates upward if not taken care of. For the mandated group, one might be worried about negative selection bias. Consequently one would have to employ an instrument to correct for omitted variable bias. Since I do not have any knowledge about the participation status of the individuals in the GSOEP, I hope to identify the effect on those individuals who changed their treatment status due to the introduction of the integration course using an instrumental variables approach.

3.3 Empirical Strategy

I want to estimate the effect of the introduction of integration courses in 2005 on labor market outcomes of immigrants to Germany. Since GSOEP does not contain information for individuals on whether the course was taken or not, the treatment effect that I focus on is

[2] Late repatriates are ethnic Germans from states of the former Soviet Union and other former eastern bloc states. Through a special acceptance process, they have a particular justification of living in Germany.

the effect of language skills on job market outcomes, rather than the direct effect of having taken the class on job market outcomes.[3]

As language classes were introduced in January 2005 (the new law was passed in 2004) there is a discontinuous jump in the probability of taking the class in 2005 from zero to a positive probability. One can think of this as a fuzzy regression discontinuity design. The course is taken either by immigrants who are mandated to do so or voluntarily. My empirical strategy revolves around using this information on the timing and rules of the policy.

Given the structure of the program and the broad definition of eligibility, focusing only on immigrants who arrive after 2004 would not capture the full population of immigrants who were affected by this policy change. Consequently, one could think about two different treatment groups: One can think of the entire population of immigrants in Germany as being treated with this policy change after 2004 and the new immigrants as also being treated separately because they are more likely to be mandated to take the class. In this paper, I mostly focus on the second group, i.e. immigrants entering Germany after 2004 that have a high chance of being mandated to take the language class. The composition of participants from the first group (i.e. immigrants to Germany arriving at any time before 2005) is likely confounded by strong selection effects that would require a different identification strategy than the one proposed here. Nevertheless, section 5.4 takes a closer look at pre-2005 cohorts.

A fuzzy regression discontinuity design can be treated like an instrumental variable set up where the discontinuity in the running variable is used as an instrument for the treatment outcome. Being on one side of the discontinuity is correlated with receiving the treatment and controlling for the running variable in a flexible way will bring out the effect of the discontinuity. My strategy uses the discontinuity that occurs in 2005 for immigrants arriving in or after 2005 (the possibility of being mandated to take the integration course).

The difference between an immigrant arriving in Germany in 2004 and one arriving in Germany in 2005 everything else equal would be that the latter immigrant could be

[3]This information is now collected (since 2015) but sample sizes are too small to perform any convincing analysis.

mandated to take a language class while the former one would have to voluntarily sign up. Only few immigrants arriving in Germany before 2005 were mandated to take the class. Hence, the probability of taking the class is higher for the population of immigrants who arrive after 2004 than for those who arrive before 2005 if these cohorts are similar in terms of characteristics.

Let C_i be the cohort that individual i belongs to. The indicator $\mathbb{1}\{C_i \geq 2005\}$ is an instrument for being in the group that is more likely to be mandated to take the class due to the timing of entrance. After controlling for demographic characteristics and the composition of a cohort through cohort effects and country of origin effects, $\mathbb{1}\{C_i \geq 2005\}$ captures the difference in language skills due to the introduction of the language program.

My baseline empirical model ideally would be the following:

$$Y_{it} = \beta_0 + \alpha_1 L_{it} + \beta_1 YSM_{it} + \gamma X_{it} + \theta_{c(i)} + \delta_t + \epsilon_{it} \tag{3.1}$$

where Y_{it} will be either log of monthly income or employment status, L_{it} is the language level of individual i in year t, YSM_{it} are the years since immigration of individual i at time t, X_{it} are demographic controls, $\theta_{c(i)}$ are cohort fixed effects and δ_t is a year fixed effect.[4]

The first stage establishes the relationship between the instrument and language proficiency, where the instrumental variable is $IV = \mathbb{1}\{C_{it} \geq 2005\}$.

$$L_{it} = \beta_0 + \alpha_1 \mathbb{1}\{C_i \geq 2005\} + \beta_1 YSM_{it} + \gamma X_{it} + \theta_{c(i)} + \delta_t + \epsilon_{it} \tag{3.2}$$

The way it is written, however, the model is not identified. A standard identification issue is to separately identify cohort, time and age effects (Heckman and Robb (1985)). A

[4]In this paper I focus on language skills measured by language speaking. If the language program affects labor market outcomes through other channels than language speaking proficiency, e.g. language writing proficiency, then the exogeneity condition of the instrument is not satisfied. Since GSOEP contains information on written language skills as well, a future version of this draft will aim to combine different language skills into one index and instrument for it.

variation of this problem in the context of migration is separately identifying cohort, time and time in country effects. Let T be the present period and C_i represent the immigration cohort, then the problem arises because of the identity:

$$YSM_{it} = T - C_i \tag{3.3}$$

If all three variables are included in the regression model, the parameters are not separately identified. Further, in crosssectional data, cohort and years in the country effects are not separately identified. Observing cohorts of immigrants over time, helps disentangle the effect of the cohort from the years in the country effects.

A second issue of this model is that the instrument is collinear with a set of cohort fixed effects.

$$\mathbb{1}\{C_i \geq 2005\} = \sum_{\tau=2005}^{2017} \mathbb{1}\{i = \tau\} \tag{3.4}$$

I solve these issues in several ways. To solve the second issue, I include a linear cohort trend instead of cohort fixed effects. This does not solve the first issue of multicollinearity of the time, period and cohort effects.

Due to perfect multicollinearity, it is not possible to include all three variables in a linear or categorical way. Borjas (2015) suggests to assume that year effects are identical for the immigrant and the native population and thus to estimate year effects from the native population and back out cohort and years in the country effects for immigrants using those estimates. Dustmann and van Soest (2002) and Lochman, Rapoport and Speciale (2016) estimate only 2 out of the 3 effects. I will break the multicollinearity in two ways. I estimate specifications with and without year fixed effects. The main purpose of including year effects is to capture economic conditions potentially affecting labor market outcomes. It is possible to control for these conditions by including macroeconomic variables for each year. Further, I also try controlling for years since immigration in a piecewise way. Figure 3.2 bears out the pattern that the increase in language gain flattens over time. I try to capture this feature in the data by including indicators for having spent less than 5 years in the country and between 5 and 10 years. This piecewise characterization breaks the multicollinearity between current year, cohort and years since immigration. I also interact these indicators with linear trends in years since immigration. Finally, instead of entering year of immigration linearly, I check for robustness when allowing it to enter as higher order polynomial in the regression.

The effects I identify using this IV strategy are local average treatment effects for those immigrants who were induced to take the class by the policy. On the one hand these are the immigrants who were mandated by the government to take the language class and would not have otherwise, i.e. those individuals who changed their treatment status induced by the policy on the other hand, for older cohorts these are immigrants who took up the class voluntarily.

3.4 Data

I use all available waves of the GSOEP (1984 - 2017). A migrant is defined as an individual who was not born in Germany or who immigrated to Germany after the end of the second world war in 1945. Those immigrating before 1945 are not marked as immigrants by GSOEP so I am unable to identify who was and was not born in Germany. Consequently I consider only immigrants who arrived in Germany after 1945. For the main analysis that focuses on differences between immigrants arriving before and after 2005, I restrict the sample to immigrants who have been in the country for at most 12 years and who arrived in 1990 or later. I pick 12 years as that is the longest I can potentially observe an immigrant arriving after 2005 if I observe that individual in the 2017 wave of GSOEP. Since the main focus of the paper is on language learning soon after arriving in the country, it makes sense to focus on relatively recently arrived immigrants. I perform some analyses with older cohorts as well to shed some light on how older immigrant cohorts are affected by the introduction of language classes. I choose immigrants arriving after 1990 as the baseline so that I only see

Table 3.1: Summary statistics

	Immigration before 2005		Immigration after 2005	
	Mean	SD	Mean	SD
German speaking skills (1-5)	3.59	1.00	2.96	1.00
German speaking skills (binary)	0.56	0.50	0.29	0.46
German writing skills (1-5)	3.23	1.16	2.86	1.10
Employed full-time	0.40	0.49	0.17	0.38
Employed	0.61	0.49	0.30	0.46
Monthly household income (EUR 2015)	2459.89	1331.31	1700.32	1150.80
Age at arrival	30.02	10.01	34.00	8.83
EU citizen	0.17	0.38	0.10	0.30
Refugee	0.03	0.18	0.50	0.50
Years of education	10.60	1.98	9.41	2.04
Experience	10.70	10.32	8.98	9.51
Experience squared	221.07	353.08	171.13	280.06
Married	0.85	0.36	0.72	0.45
Number of hh members	3.47	1.37	3.90	2.01
Female	0.57	0.50	0.45	0.50
Partner employed	0.72	0.45	0.64	0.48
Observations	3980		14402	

immigrants coming to Germany after reunification. In the sample that I use for the analysis I restrict the age of immigrants to 25-65 years. Since I am interested in the effect of language classes on job market outcomes, I want all the individuals in the sample to be of working age. Furthermore, the GSOEP includes information on refugees as well. I exclude refugees in the baseline specifications on the basis of having very different reasons for immigration than economic migrants. However, I also run the empirical analysis for refugees only.

Table 3.1 shows summary statistics for the sample I consider in this paper. The table breaks down the sample in terms of the instrument that I use in the empirical analysis, i.e. Table 3.1 shows the sample broken down by immigration date (immigration before and after 2005). Monthly household income is measured in 2015 Euros.

The raw data show some of the patterns that drive the results. The language variable is measured on a scale from 1 to 5 where 1 is not speaking German at all and 5 is speaking German very well. Table 3.1 shows that for immigrants arriving before 2005 the language mean is 3.59 and for those arriving after 2005 the mean is 2.96. Given that these means

Table 3.2: Immigrants and refugees

	Immigrants		Refugees	
	Mean	SD	Mean	SD
German speaking skills (1-5)	3.39	0.99	2.65	0.94
German speaking skills (binary)	0.46	0.50	0.18	0.39
German writing skills (1-5)	3.18	1.10	2.58	1.06
Employed full-time	0.33	0.47	0.06	0.23
Employed	0.52	0.50	0.14	0.35
Monthly household income (EUR 2015)	2251.93	1373.62	1289.86	645.94
Age at arrival	31.97	9.37	34.92	8.77
EU citizen	0.19	0.39	0.00	0.03
Years of education	10.14	2.04	8.93	1.94
Experience	9.79	9.74	8.68	9.66
Married	0.78	0.41	0.70	0.46
Number of hh members	3.57	1.70	4.16	2.12
Female	0.52	0.50	0.40	0.49
Partner employed	0.71	0.45	0.57	0.50
Observations	11088		7294	

are from the raw data and that immigrants arriving after 2005 should have worse language skills due to their recent arrival, the difference in average skills seems relatively small. One can also define speaking proficiency in a binary way, classifying it as good for values of the old language variable above 3. Again, skills for the more recently arrived immigrants are lower but the difference is of a similar magnitude as the difference for the original language variable, a little over half a standard deviation. The means of full employment, employment and monthly household income are all lower after 2005.

Most of the differences in demographic characteristics for immigrants arriving before and after 2005 are driven by the increase of refugees from 3% to 50% of the sample. Table 3.2 shows the difference between immigrants and refugees in the sample. On average, refugees have fewer years of education and experience and are less likely to be female or married. Language skills and employment outcomes are worse for refugees than they are for immigrants. In particular there is a large gap in employment rates and monthly household income is lower by around 1,000 Euros.

I am interested in assessing the effectiveness of integration courses on labor market

Figure 3.2: Language skills over time, before and after 2005

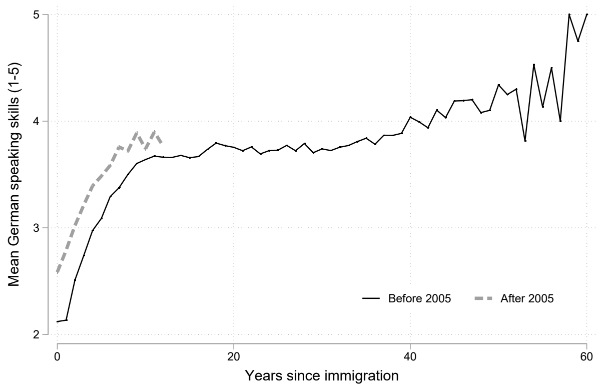
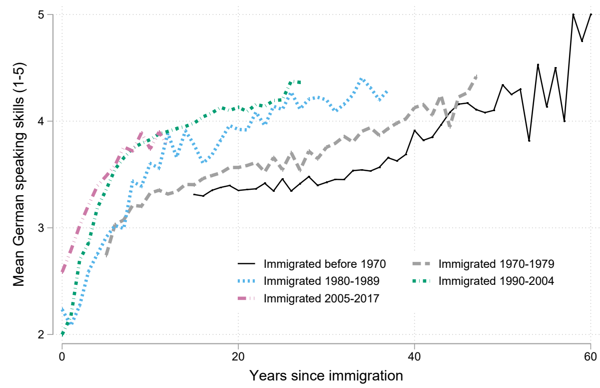


Figure 3.3: Language skills over time, for different cohorts



Notes: The left panel plots average German speaking proficiency (self-reported on a scale from 1-5) by year since immigration for immigrant cohorts arriving before and after 2005. Since I have access to GSOEP waves until 2017, the average for cohorts arriving after 2005 end at 12 year since immigration. The right panel plots average German speaking proficiency by year since immigration for various cohorts.

outcomes of immigrants through an improvement in language. Since I use an IV approach, exploiting a fuzzy discontinuity in the probability of receiving treatment in 2005, it is important that the instrument I use is correlated with the endogenous variable (language proficiency). Figures 3.2 and 3.3 attempt to look at this relationship in the raw data.

Figures 3.2 and 3.3 show the average of language speaking proficiency over time for different cohorts of immigrants. As expected, the average of language proficiency increases with the years since entry to Germany. Figure 3.2 shows the average over time for immigrants arriving before and immigrants arriving after 2005. In 2005, the integration course took effect. There is a big jump of over half a language category in the level of initial language proficiency in the group of immigrants arriving after 2005 in the raw data. In terms of the slope of language assimilation, the results from this graph are more ambiguous. For the first five years after immigration, the slope seems steeper for immigrants arriving after 2005 than before 2005, however this appears to vanish after that and there is some convergence to the other group.

Figure 3.3 shows the mean of language proficiency over time for different cohorts of immigrants. For the cohorts arriving before 1970, I have no information on means for immigrants with a low number of years since immigration and similarly for the cohort arriving between 1970 and 1979. However, the cohort arriving in the 70s seems to connect to the path of the cohorts arriving in the 80s. It appears that for the cohorts that I can observe from the beginning of their time in Germany (those immigrants arriving in the 80s, the 90s and the early 2000s) as compared to the immigrants arriving after 2005, I observe a similar (slightly smaller but still distinctive) gap in the levels of initial language proficiency as in Figure 3.2. Looking at the older cohorts, for example the cohorts arriving between 1980 and 1989, the youngest of those cohorts would have been in Germany for 16 years in 2005. Examining the line for that cohort between 10 and 16 years, there does seem to be an increase in the slope around that time which might be due to the introduction of integration classes. For the cohorts arriving between 1990 and 2004 the slope in the first years since immigration is quite steep, again possibly due to integration courses.

It is possible that the level effects I observe for the cohorts arriving in or after 2005 are cohort effects. It is also possible that the effect of the integration course on the new cohorts is more of a level effect and for the earlier cohorts is visible as a change in the slope of assimilation around the time of the policy change.

These graphs from the raw data indicate that language proficiency among immigrants is affected in some way after 2005 which I analyze further now.

3.5 Results: First stage

3.5.1 Basic results

The first stage examines the effect of the instrument on the endogenous variable, so the effect of arriving in Germany after the introduction of language classes on language skills. For the instrument to be valid it must be relevant and exogenous. As discussed above, the policy would not have been easy to anticipate and plan one's immigration date around. A possible concern is whether the instrument is capturing the effects of the introduction of integration courses as opposed to other policy changes occurring at the same time that may affect the immigrant population. However, in 2004, there was simultaneously no other policy change that would result in a sudden improvement of language skills. There is little reason to believe that the announcement of opening of borders for citizens of ten Eastern European countries seven years later (which was part of the change of the same law that also introduced language classes) would have an effect on language skills. I control for country of origin specific effects, time fixed effects and cohort trends to avoid that sudden unrelated changes in the composition of immigrants are captured through the instrument.

Tables 3.3 and 3.4 show the first stage results for immigrants arriving in Germany in or after 1990 who are between 25 and 65 years old and have been in the country for at most 12 years. The main variable of interest is the coefficient on the instrument.

Table 3.3 shows the baseline first stage results, controlling for years since immigration, cohort and period in different ways. Column 1 and 2 include year fixed effects, columns 3 and 4 include state-specific year trends and columns 5 and 6 do away with year effects altogether but instead include state specific unemployment rates. The different versions of controlling for years since immigration all indicate that spending more time in Germany increases the ability to speak German. There seems to be an overall downward trend in the language skills of new cohorts as indicated by the linear immigration year trend. Relative to this trend the instrument, an indicator for arriving in Germany after 2004 suggests that there is a positive change in language skills for those immigrants arriving after the introduction of the program who are more likely to be mandated to participate relative to immigrants arriving before. The effect of the instrument is quite robust across all specifications, hovering around a tenth of a language category. Relative to the mean of the language variable (3.39) the coefficients suggest a 3-3.5% increase in German speaking skills for immigrants arriving after 2004. Since no information is known as to whether an immigrant actually participated in a language class or not, these effects are likely to be attenuated relative to the effect of treatment on a participating individual.

Table 3.3: First stage - basic specification

	(1) German speaking	(2) German speaking	(3) German speaking	(4) German speaking	(5) German speaking	(6) German speaking
Entry after 2004	0.109**	0.111**	0.0846*	0.0836*	0.116**	0.118**
	(0.0396)	(0.0394)	(0.0372)	(0.0374)	(0.0361)	(0.0362)
Immigration year	-0.0730***	-0.0717***	-0.0707***	-0.0756***	-0.00679**	-0.00481
	(0.00819)	(0.00516)	(0.00500)	(0.00806)	(0.00255)	(0.00329)
Years since immigration (YSM)					0.0604***	0.0623***
					(0.00383)	(0.00432)
Less than 5 years in country	0.0558			0.105		
	(0.0637)			(0.0633)		
Between 5 and 10 years in country	0.0957*			0.131***		
	(0.0385)			(0.0380)		
Less than 5 years × YSM		0.0199*	0.0267**			
		(0.00967)	(0.00962)			
Between 5 and 10 years × YSM		0.0123***	0.0154***			
		(0.00330)	(0.00322)			
Years of education	0.132***	0.132***	0.129***	0.129***	0.130***	0.130***
	(0.00478)	(0.00478)	(0.00476)	(0.00476)	(0.00476)	(0.00476)
Experience	0.00768**	0.00759**	0.00836**	0.00846**	0.00822**	0.00823**
	(0.00264)	(0.00264)	(0.00265)	(0.00265)	(0.00265)	(0.00265)
Experience squared	0.0000767	0.0000776	0.0000662	0.0000648	0.0000730	0.0000729
	(0.0000802)	(0.0000803)	(0.0000807)	(0.0000807)	(0.0000808)	(0.0000809)
EU citizen	0.987***	0.984***	0.968***	0.971***	0.984***	0.981***
	(0.131)	(0.131)	(0.132)	(0.132)	(0.132)	(0.133)
Married	-0.0348	-0.0346	-0.0377	-0.0381	-0.0302	-0.0304
	(0.0230)	(0.0230)	(0.0230)	(0.0230)	(0.0231)	(0.0231)
Number of hh members	-0.0490***	-0.0491***	-0.0469***	-0.0468***	-0.0483***	-0.0482***
	(0.00578)	(0.00578)	(0.00578)	(0.00579)	(0.00582)	(0.00581)
Female	0.0196	0.0198	0.0220	0.0218	0.0211	0.0212
	(0.0197)	(0.0197)	(0.0197)	(0.0197)	(0.0197)	(0.0197)
Age at arrival	-0.0292***	-0.0291***	-0.0291***	-0.0291***	-0.0292***	-0.0292***
	(0.00139)	(0.00139)	(0.00139)	(0.00139)	(0.00140)	(0.00140)
State level controls						✓
Year f.e.	✓	✓				
State year trend			✓	✓		
State f.e.	✓	✓	✓	✓	✓	✓
Country of origin f.e.	✓	✓	✓	✓	✓	✓
Observations	10485	10485	10485	10485	10485	10485
R^2	0.315	0.315	0.314	0.313	0.306	0.306

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The control variables that I use are commonly used in the literature in the context of immigrants and job market outcomes. Their signs and sizes remain relatively constant throughout all specifications of the model and will not be reported in further tables. Most of the variables are related to language skills in the way that we would expect: Immigrants with more years of education and experience are better at speaking the language, as are EU citizens which might reflect a cultural and linguistic closeness. Being female or being married affects language skills in a positive but statistically not significant way while the number of household members has a strong negative impact on speaking skills. The larger a family is, the more they might be speaking a language other than German at home which may prevent them from learning German. Immigrant parents may also use their children to translate when in situations where German is required. Finally, immigrants' age at arrival in Germany is correlated negatively with speaking skills, possibly as it gets more difficult to learn a new language with age.

While I do not see each individual's participation in the language class, the policy itself provides some guidance. All non-EU immigrants arriving after 2004 were mandated to participate in this language class unless they could already demonstrate sufficient language skills on arrival. The upper right panel of Figure 3.1 confirms that a large majority of non-EU immigrants arriving after 2005 and participating in the class were mandated to do so. To explore this difference in the regulation by country of origin, specifically by EU citizenship status, panel A and B of table 3.4 perform the baseline analysis for only EU citizens (panel A) and only non-EU citizens (panel B).[5]

Comparing panel A and B, the coefficients for non-EU citizens are all estimated more precisely than those for only EU citizens and comparing the preferred specifications in columns 9-12 to columns 3-6 effect sizes are mostly larger for non-EU citizens than EU citizens. Relative to means in these sub-groups, the increase in German speaking skills suggested by columns 5-6 and 11/12 for cohorts entering after 2004 is roughly 4% for non-EU

[5]Note that the definition of being an EU citizen changes over time as more countries join the EU.

Table 3.4: First stage - subgroup analysis

			A. Only EU citizens			
	(1)	(2)	(3)	(4)	(5)	(6)
Entry after 2004	0.152	0.152	0.0974	0.0885	0.0710	0.0988
	(0.0788)	(0.0816)	(0.0777)	(0.0777)	(0.0760)	(0.0760)
Observations	1850	1850	1850	1850	1850	1850
R^2	0.323	0.323	0.321	0.322	0.307	0.309
			B. Only non-EU citizens			
	(7)	(8)	(9)	(10)	(11)	(12)
Entry after 2004	0.0978*	0.0970*	0.0971*	0.0989*	0.128**	0.128**
	(0.0459)	(0.0457)	(0.0429)	(0.0431)	(0.0417)	(0.0417)
Observations	8463	8463	8463	8463	8463	8463
R^2	0.310	0.311	0.309	0.308	0.301	0.301
			C. Refugees			
	(1)	(2)	(3)	(4)	(5)	(6)
Entry after 2004	0.791*	0.00806	0.0634	0.222	0.738***	0.775***
	(0.357)	(0.375)	(0.233)	(0.225)	(0.207)	(0.205)
Observations	6576	6576	6576	6576	6576	6576
R^2	0.305	0.311	0.313	0.312	0.301	0.303
Years since immigration	Piecewise	Piecewise × linear	Piecewise × linear	Piecewise	Linear	Linear
Cohort	Linear	Linear	Linear	Linear	Linear	Linear
State level controls						✓
Demographic controls	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓				
State year trend			✓	✓		
State f.e.	✓	✓	✓	✓	✓	✓
Country of origin f.e.	✓	✓	✓	✓	✓	✓

Notes: Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Demographic controls include years of education, experience, experience squared, indicators for being a citizen of an EU member state, for being married, for being female, the number of household members and age at arrival in Germany. The dependent variable is German speaking proficiency (measured on a scale from 1-5, 1 being not at all, 5 being native).

citizens and roughly 2.7% for EU citizens. It seems that the baseline results were driven by larger increases in language skills for non-EU citizens who were more likely to be mandated to take the class. Once again, the effects are quite robust with respect to changes in the specification.

It is possible that the larger standard errors in Panel A stem from a much smaller sample size. While non-EU citizens were mandated to take the language class, there is for the most part selection in language class participation among EU citizens. If selection on unobservables is positive, then the estimated first stage coefficient could have been much larger for EU than non-EU citizens.

3.5.2 Refugees

The analysis so far has excluded refugees since they are a fundamentally different type of immigrant. As they are less selected for their destination country than economic migrants, they are an interesting subgroup to study. In the wake of the Syrian refugee crisis, Germany accepted a large number of asylum seekers from Syria. A shortage of spots in language classes was among the early challenges that the German government faced.

Panel C of Table 3.4 shows first stage results for the refugee sample in GSOEP. The results for refugees are not as robust across specifications as they are for the other samples and sub-samples. In particular, piecewise controlling for years since migration (columns 2-4) leads to very small first stage coefficients compared to the other specifications. This is likely due to the imbalance of the refugee sample, with over 90% arriving after 2005 and two thirds of the refugees in the sample arriving between 2014 and 2016. There are very few refugees that are there for longer than 5 years. Therefore, there may be a multicollinearity issue when in the piecewise specifications in columns 2,3 and 4.

The effect sizes in columns 5 and 6 suggest an increase in language speaking ability of 28% relative to the mean of 2.64 for refugees arriving after 2005 compared to those arriving before. The results indicate that having access to language classes soon after arriving in the country, can substantially improve the language proficiency of refugees who have on average fewer years of experience and education than do economic migrants. In addition, their social networks could likely include mostly non-German speakers so that participating in a language program provides better access to learning the language.

3.5.3 Robustness

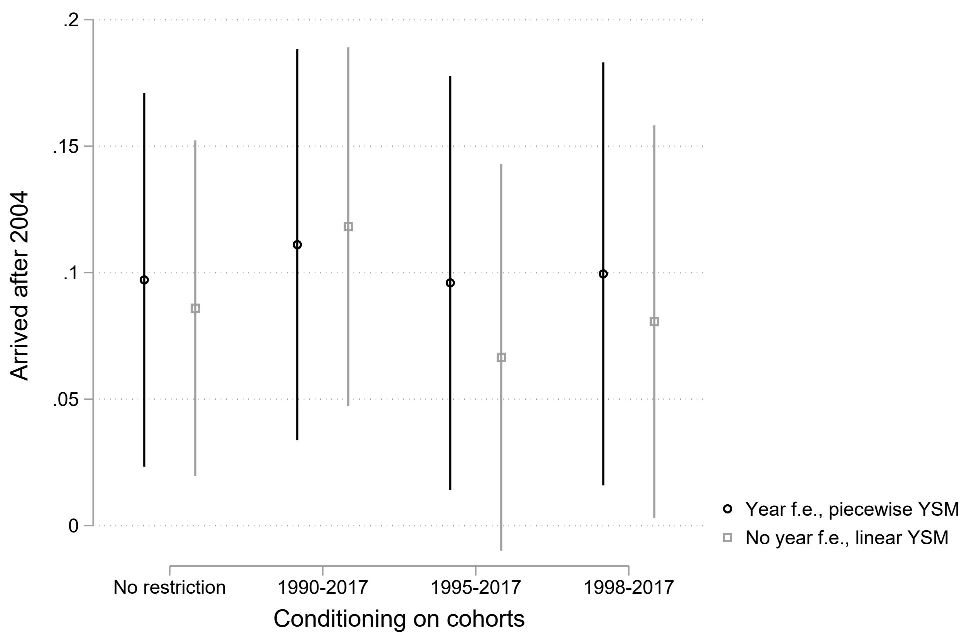
For the main analysis I restrict myself to immigrants arriving in 1990 or later, i.e. arriving after German reunification. This subsection relaxes that restriction and runs the basic analysis for sub-samples restricted differently by immigration year. I run two specifications from Table 3.3, the specification used in column 2 which estimates year fixed effects and piecewise controls for years since immigration as well as the specification used in column 6 which includes state controls but does not estimate time fixed effects. I run these two specifications for a sample that does not restrict immigration year at all, the sample used in the baseline (i.e. immigration year restricted to 1990 or later), restricting immigration year to 1995 or later and restricting immigration year to 1998 or later.

Figure 3.4 plots the first stage coefficients for each of these sample restrictions. The second pair of coefficients comes from the baseline sample. The coefficients are very similar and it is not possible to reject that they are not equal to each other. This provides further evidence that the first stage patterns do not arise from the way the sample is restricted based on cohorts.

3.5.4 Treatment of pre-2005 cohorts

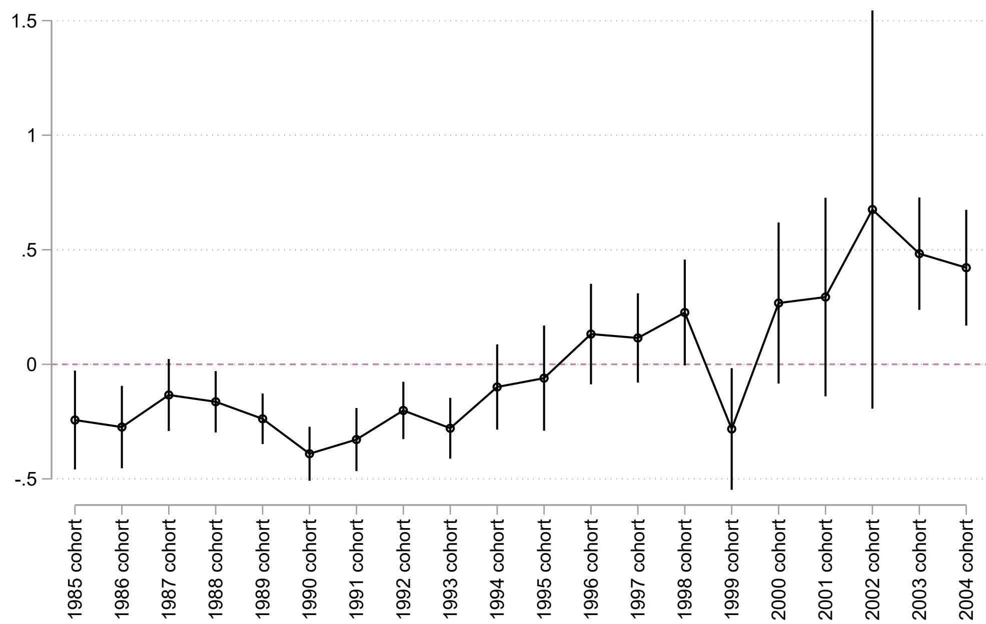
The analysis until now has focused on comparing cohorts of immigrants arriving after 2005 to those that arrive before 2005. However, even immigrants from cohorts that arrive before 2005 may enroll in this language class if there is capacity. Some immigrants from earlier cohorts are also mandated to attend by the Federal Employment Agency. Essentially, each cohort arriving before 2005 is exposed to the language program at a different point in their time in

Figure 3.4: First stage - varying cohort restrictions



Notes: This figure reports first stage coefficients and 95% confidence intervals for the instrument for all available immigration cohorts, cohorts arriving between 1990 and 2017 (baseline sample restriction), cohorts arriving between 1995-2017 and cohorts arriving 1998-2017. The first coefficient for each sample comes from a regression specification as in Table 3.3 column 2, including year fixed effects and piecewise controls for years since immigration. The second coefficient comes for a regression specification as in Table 3.3 column 6, i.e. without year effects and with a linear years since immigration trend.

Figure 3.5: Cohort treatment effects



Notes: This figure reports cohort treatment effects T_c and 95% confidence intervals from Equation 3.5. Each immigrant cohort arriving before the introduction of language classes is treated at a different point in their tenure in Germany. The regression contains controls for years since immigration, years a cohort was exposed to the program, state unemployment rate, state fixed effects, cohort fixed effects, country of origin fixed effects, years of education, experience, experience squared, indicators for being a citizen of an EU member state, for being married, for being female, the number of household members and age at arrival in Germany.

the country, i.e. the 2004 cohort is treated after one year in the country, the 2000 cohort is treated after 5 years in the country etc. To analyze the effect of the language program on the cohorts already in the country at the time of introduction, I run the following specification:

$$L_{it} = \alpha_0 + \beta_{c(i)}T_{c(i)} + \gamma X_{it} + \tau YSM_{it} + \theta unemploymentrate_t + \nu state_{(i)} + \epsilon_{it} \qquad (3.5)$$

where $T_{c(i)}$ is a set of treatment indicators for each cohort, that takes the value of 1 for each cohort after 2005 and $\delta_{c(i)}$ is a set of cohort fixed effects. The demographic controls, X_{it} are defined in the same way as before and years since immigration enter linearly. I also control for the number of years of exposure to the program which will be common to all cohorts. The $T_{c(i)}$ capture the effect of the program being introduced at a particular point in time in the tenure of a given cohort in Germany.

Figure 3.5 plots the cohort treatment effects $\beta_{c(i)}$ and 95% confidence intervals. The size of the coefficients increases almost monotonly. The most recent cohorts have the largest treatment coefficients, of the magnitude of almost half a language category, so about four times the size of the effect for cohorts entering after 2005. It is reasonable that more recently arrived cohorts are more likely to take advantage of language classes and enroll compared to older cohorts. If there is positive selection for the participants, this could explain why the magnitude of the effect is large.

Tables 3.A-1 and 3.B-3 show further robustness checks for the first stage. Table 3.A-1 shows the robustness of first stage results to defining language proficiency using an indicator for language proficiency being good. The results are qualitatively the same as results for the original language variable. Table 3.B-3 shows robustness with respect to controlling for cohort effects using a 3rd order polynomial instead of a linear trend.

3.6 Results: Second Stage

The first stage regressions show that introducing language classes for immigrants robustly increased the German language proficiency of immigrant cohorts more likely to have attended such a class. Now I turn to the second stage results that estimate a local average treatment effect of language proficiency on job market outcomes for those immigrants that were induced by the program to take a language class. I examine three outcomes: whether an individual is employed full-time, whether an individual is employed at all (full- or part-time) and the log of monthly household income. From now on I restrict attention to specifications of the empirical model that control for state level labor market controls but do not estimate time fixed effects. I keep those specifications where the multicollinearity between years in the country, cohort and period is least concerning, i.e. either year fixed effects are not estimated or there are state year trends with piecewise control for years since immigration. The previous section has shown that first stage results are robust across these specifications. The treatment coefficient can be determined as the change in the outcome variable for a one unit increase in the language proficiency variable.

3.6.1 Cohorts after 2004

Table 3.5 reports the results of the second stage regressions. Panel A focuses on cohorts from 1990-2017. Columns 1-3 show that there is a large increase in the probability of full-time employment for an increase in language speaking proficiency by one unit. A base full-time employment rate in the sample of 37% implies that an increase in speaking proficiency by one unit doubles the probability of full-time employment. The first stage estimate implies a reduced form estimate of the language program on the probability of full-time employment of 0.046 to 0.06, i.e. an increase of about 12.5% relative to the average in the sample. This magnitude is quite comparable to the literature on language and job market outcomes that tends to find increases around 15-20% in the probability of employment.

Columns 4-6 in Panel A suggest positive effects of language proficiency on the probability of being employed in any way (part-time or full-time) but accompanied by large standard errors. Combined with the results from columns 1-3 this could suggest that language classes are beneficial for finding a full-time job or switching from part-time to full-time work. These findings are contrary to Lochman, Rapoport and Speciale (2016) who find that taking a language class is negatively related to finding full time employment, perhaps since it is difficult to be working at the same time as having a job. As I don't observe when exactly someone participates in the class, it is reasonable to find positive effects if participation improves employment chances over time.

Columns 7-9 show the effects on log monthly household income. These effects are extremely large and indicate increases in household income by 90-100% for an increase in language proficiency by 1. The reduced form effect of the policy on household income is roughly a 10% increase. I control for the partner's employment status in these regressions to control for there being a second earner and for the fact that job searching behavior might be different in a dual-earner household. Together, the results in Panel A tell the story of substantial increases both in the chances of full-time employment and potentially employment in higher-paying jobs. Unfortunately, data on occupation and job characteristics in GSOEP is too sparse to analyze whether there are changes in the type of employment that immigrants find after improving language proficiency.

Panel B in Table 3.5 shows results for the same specifications for only immigrants from non-EU countries. The patterns remain very similar to those in Panel A, with large and mostly robust effects on full-time employment and monthly income, but imprecisely estimated effects on the overall probability of being employed. While the first stage effects for non-EU citizens were larger than those for citizens from EU member states, the reverse seems to be true in the second stage. This may reflect the fact that higher education systems are fairly integrated within the EU but it is often much more difficult to transfer to working in Germany with degrees from outside the EU.

Table 3.5: Second stage results

	Employed full-time			Employed			Log monthly household inc		
A. Baseline									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
German speaking skills	0.469*	0.512*	0.388**	0.232	0.282	0.252	1.004**	1.022**	0.878***
	(0.280)	(0.298)	(0.180)	(0.240)	(0.252)	(0.169)	(0.462)	(0.474)	(0.296)
Observations	10485	10485	10485	10485	10485	10485	9797	9797	9797
B. Only non-EU citizens									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
German speaking skills	0.291	0.303	0.321*	0.142	0.164	0.249	0.889**	0.868**	0.891***
	(0.234)	(0.234)	(0.178)	(0.231)	(0.229)	(0.179)	(0.410)	(0.394)	(0.315)
Observations	8463	8463	8463	8463	8463	8463	7877	7877	7877
C. Refugees									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
German speaking skills	0.103	0.0496	0.0359	0.196*	0.176**	0.122	0.313*	-0.0337	-0.198
	(0.0940)	(0.0634)	(0.116)	(0.118)	(0.0839)	(0.149)	(0.183)	(0.112)	(0.188)
Observations	5052	5052	5052	5052	5052	5052	4759	4759	4759
Years since immigration	Piecewise × linear	Piecewise	Linear	Piecewise × linear	Piecewise	Linear	Piecewise × linear	Piecewise	Linear
Cohort	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
State level controls			✓			✓			✓
Demographic controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year f.e.									
State year trend	✓	✓		✓	✓		✓	✓	
State f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country of origin f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Robust standard errors in parentheses, * $p < 0.01$, ** $p < 0.05$, *** $p < 0.01$. This table reports the results from an IV regression of labor market outcomes on language skills, instrumenting for language skills using an indicator for immigration cohorts arriving in Germany in 2005 or later. The outcomes are indicators for whether an individual is employed full-time, employed full-or part-time and the log of monthly household income. Panel A and C include all immigrant cohorts starting in 1990, Panel B restricts attention to citizens of non-EU countries and Panel C focuses on refugees. Demographic controls include years of education, experience, experience squared, indicators for being a citizen of an EU member state, married, female, the number of household members, partner's employment status and age at arrival in Germany.

Table 3.6: OLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Employed full-time			Employed			Log monthly hh income		
German speaking skills	0.0218***	0.0220***	0.0218***	0.0564***	0.0568***	0.0570***	0.0434***	0.0434***	0.0437***
	(0.00465)	(0.00466)	(0.00463)	(0.00502)	(0.00503)	(0.00500)	(0.00517)	(0.00517)	(0.00516)
Years since immigration	Piecewise × linear	Piecewise	Linear	Piecewise × linear	Piecewise	Linear	Piecewise × linear	Piecewise	Linear
Cohort	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
State level controls			✓			✓			✓
Demographic controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year f.e.									
State year trend	✓	✓		✓	✓		✓	✓	
State f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country of origin f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	10485	10485	10485	10485	10485	10485	9797	9797	9797

Notes: Robust standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table reports the results from OLS regressions of labor market outcomes on German speaking proficiency (measured on a scale from 1-5, 1 being not at all, 5 being native). The outcomes are indicators for whether an individual is employed full-time, employed full-or part-time and the log of monthly household income. Demographic controls include years of education, experience, experience squared, indicators for being a citizen of an EU member state, married, female, the number of household members, partner's employment status and age at arrival in Germany.

Comparing the IV results to OLS results informs us further about how to interpret the local average treatment effect of the IV estimate. Table 3.6 reports OLS regressions for the same specifications and dependent variables as in Table 3.5. Better language proficiency is strongly positively associated with better employment outcomes and higher monthly income. One might think that the OLS results are upward biased if individuals investing in language skills more are otherwise also more likely to do well in the labor market. Comparing Tables 3.5 and 3.6 however, the estimated IV effects of language proficiency are much larger than the effects estimated using OLS. An explanation might be that the program induces immigrants to learn German who may otherwise not have invested as much in their language skills and that the local average treatment effect of learning German is high for this group. In other words, language proficiency is a key component to finding a job in Germany.

Panel C in Table 3.5 shows results for the refugee sample. Overall, these estimates are much more noisy than the immigrant sample but the most precisely estimated effects exhibit the opposite pattern of the immigrant sample. The estimated effects on full-time employment are very small and imprecise and the effects on household income are even more volatile. The effects on the probability of employment overall are the most robust, implying a doubling in employment chances (relative to a low baseline average of 14.4% employed) for an increase in the language variable by 1 unit. The reduced form is very similar to this effect (Table 3.4, Panel C shows that first stage effects for refugees are almost an entire unit of the language variable) and suggests increases in employment chances by 90%. Given that refugees come from different education and skill backgrounds and are also likely to be much less prepared to enter the German labor market than immigrants, it is reasonable that participating in a mandatory language program boosts the chances of employment significantly. It does not seem though that the employment that refugees find who have taken the language class is higher paying than the employment that refugee cohorts before 2005 find.

3.6.2 Pre-2005 cohorts

Table 3.7 shows the second stage results for the older cohorts, instrumenting for language proficiency using the set of cohort treatment effects. An increase in language speaking proficiency by one unit leads to an increase in employment probability of roughly 23% (relative to an average of 60%), an increase in the probability of full-time employment of 30% (relative to an average of 45%) and an increase in monthly income by 11%.

The first stage results for older cohorts suggest that the average treatment effect probably masks larger effects for cohorts arriving close to 2004 and smaller effects for cohorts that arrived in Germany in the early 90s and the 80s. Older immigrants cohorts already have much stronger ties to the labor market and are employed at higher rates than their more recent counterparts. Those who enroll in this language class are either enrolling voluntarily or are mandated to enroll by the Federal Employment Agency (see Figure 3.1). Compared to their more recently immigrated counterparts, language classes have a smaller effect on older cohorts that are already more established in the labor market and familiar with the language. However, the effect sizes are still larger than the OLS estimates, possibly because one major reason for enrolling is being told to do so by the Federal Employment Agency. For those participants, improvements in employment status and income are almost mechanically expected.

Table 3.7: Second stage - pre-2005 cohorts

	(1) Employed	(2) Employed	(3) Employed full-time	(4) Employed full-time	(5) Log monthly hh inc	(6) Log monthly hh inc
German speaking skills	0.208***	0.143***	0.159***	0.142***	0.111**	0.114**
	(0.0582)	(0.0546)	(0.0532)	(0.0499)	(0.0508)	(0.0478)
Years since immigration	Linear	Linear	Linear	Linear	Linear	Linear
Cohort	Linear	Linear	Linear	Linear	Linear	Linear
State level controls	✓	✓	✓	✓	✓	✓
Demographic controls	✓	✓	✓	✓	✓	✓
Year f.e.						
State year trend						
State f.e.	✓	✓	✓	✓	✓	✓
Country of origin f.e.	✓	✓	✓	✓	✓	✓
Observations	10251	13137	10251	13137	9781	12510

Notes: Robust standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table reports second stage IV estimates for cohorts already in Germany at the time of the policy change. The instruments are a set of cohort treatment effects (see Equation 3.5). Each immigrant cohort arriving before the introduction of language classes is treated at a different point in their tenure in Germany. The outcomes are indicators for whether an individual is employed full-time, employed full-or part-time and the log of monthly household income. The regression contains controls for years since immigration, years a cohort was exposed to the program, state unemployment rate, state fixed effects, cohort fixed effects, country of origin fixed effects, years of education, experience, experience squared, indicators for being a citizen of an EU member state, for being married, for being female, the number of household members and age at arrival in Germany.

3.7 Conclusion

This paper studies the effect that proficiency in the language of the destination country has on job market outcomes of immigrants. I use the variation from the introduction of a large-scale language program in Germany that mandated certain groups of immigrants to enroll in language classes after 2004 and allowed other groups to voluntarily enroll. I find that German speaking skills increase by 3% which masks increases of 4% for non-EU citizens and smaller increases for EU-citizens. Refugees - who are much less selected for their destination country than economic migrations, see increases of 30% in language speaking proficiency.

The increases in language proficiency are robust to various empirical models that deal with the multicollinearity of time in the country, time period and cohort.

The second stage results show that the increases in language proficiency caused by the integration classes resulted in improved labor market outcomes. For immigrants, I find increases in the chance of full-time employment of between 12 and 13% but no increases in the probability of being employed in any way (either part-time or full-time). For refugees, the probability of being employed in any way doubles for cohorts that are mandated to take a language class relative to those that were not.

To my knowledge, this is the first economic analysis of the integration course in Germany. According to my results, the continued popularity of this course is justified as it seems to boost chances of integration into the German labor market substantially - particularly for refugees and non-EU citizens who often struggle more with finding employment in Germany. My results match with the findings of the evaluation of similar programs in other European countries. I want to end with some suggestions about possible extensions of this paper.

A drawback of the GSOEP data is that the current sample of individuals for whom there is information on actual participation is quite small. In some years, this sample should be large enough to allow a better analysis of the treatment effect on the treated. At the moment, this paper focuses on the effect of language speaking on job market outcomes. It is possible that language classes affect other aspects of language proficiency such as German writing proficiency. In future versions of this paper I will consider a measure of language proficiency that combines writing and reading skills and study the impact of the language program on job market outcomes through overall language proficiency. Another path for extension lies in studying whether immigrant labor becomes more substitutable with native labor and whether immigrants pick different jobs after learning the local language. The GSOEP gives data on occupations of its respondents, however the sample of those who answered is quite small. One can use O*NET data to match occupations with skills that are related to speaking the language, for example communicativeness, persuasiveness or how important the language is for the job. The size of the sample makes it difficult to estimate the same empirical models that I have been estimating in this paper, hence I leave it for an extension to be thought of in the future.

3.A Binary language learning indicator

Tables 3.A-1 and 3.A-2 show the first and second stage results when language proficiency is defined in a binary way: The indicator equals 1 when the original language variable is 4 or greater, i.e. self-reported German speaking skills are labeled as good or very good.

Table 3.A-1: First stage - language binary

	(1) German speaking	(2) German speaking	(3) German speaking	(4) German speaking	(5) German speaking	(6) German speaking
Entry after 2004	0.0548**	0.0445*	0.0286	0.0268	0.0393*	0.0397*
	(0.0211)	(0.0216)	(0.0200)	(0.0201)	(0.0194)	(0.0195)
Years since immigration	Piecewise	Piecewise × linear	Piecewise × linear	Piecewise	Linear	Linear
Cohort	Linear	Linear	Linear	Linear	Linear	Linear
State level controls						✓
Demographic controls	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓				
State year trend			✓	✓		
State f.e.	✓	✓	✓	✓	✓	✓
Country of origin f.e.	✓	✓	✓	✓	✓	✓
Observations	10485	10485	10485	10485	10485	10485
R^2	0.234	0.235	0.235	0.236	0.230	0.230

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.A-2: Second stage - language binary

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Employed full-time			Employed			Log monthly hh income		
German speaking skills	1.388	1.600	1.154*	0.685	0.883	0.750	2.909	3.114	2.557**
	(1.107)	(1.322)	(0.687)	(0.794)	(0.939)	(0.573)	(1.993)	(2.261)	(1.274)
Years since immigration	Piecewise × linear	Piecewise	Linear	Piecewise × linear	Piecewise	Linear	Piecewise × linear	Piecewise	Linear
Cohort	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
State level controls			✓			✓			✓
Demographic controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year f.e.									
State year trend	✓	✓		✓	✓		✓	✓	
State f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country of origin f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	10485	10485	10485	10485	10485	10485	9797	9797	9797

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.B Flexible immigration year

Table 3.B-3 shows the first stage results when controlling for immigration cohort as third order polynomial.

Table 3.B-3: Flexible immigration year

	(1) German speaking	(2) German speaking	(3) German speaking	(4) German speaking	(5) German speaking	(6) German speaking
Entry after 2004	0.139***	0.108**	0.0646	0.0613	0.107**	0.108**
	(0.0387)	(0.0398)	(0.0379)	(0.0382)	(0.0366)	(0.0366)
Years since immigration	Piecewise	Piecewise × linear	Piecewise × linear	Piecewise	Linear	Linear
Cohort	3rd order polynomial	3rd order polynomial	3rd order polynomial	3rd order polynomial	3rd order polynomial	3rd order polynomial
State level controls						✓
Demographic controls	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓				
State year trend			✓	✓		
State f.e.	✓	✓	✓	✓	✓	✓
Country of origin f.e.	✓	✓	✓	✓	✓	✓
Observations	10485	10485	10485	10485	10485	10485
R^2	0.314	0.315	0.314	0.314	0.306	0.306

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Bibliography

Aaronson, Daniel, Daniel Hartley and Bhashkar Mazumder. 2017. "The Effects of the 1930s HOLC "Redlining" Maps." Working Paper.

Anenberg, Elliot and Edward Kung. 2018. "Can more housing supply solve the affordability crisis? Evidence from a neighborhood choice model." *Regional Science and Urban Economics* pp. 1–12.

Angrist, Joshua and Joern-Steffen Pischke. 2008. *Mostly Harmless Econometrics. An Empiricist's Companion.* Princeton University Press.

Angrist, Joshua and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *American Economic Review* 94(5):1613–1634.

Ansari, Zahid, James N. Laditka and Sarah B. Laditka. 2006. "Access to Health Care and Hospitalization for Ambulatory Care Sensitive Conditions." *Medical Care Research and Review* 63(6):719–741.

Appel, Ian and Jordan Nickerson. 2017. "Pockets of Poverty: The Long-Term Effects of Redlining." Working Paper.

Armfield, Nigel R., Sisira K. Edirippulige, Natalie Bradford and Anthony C. Smith. 2014. "Telemedicine — Is the Cart Being Put Before the Horse?" *Medical Journal of Australia* 200(9):530–533.

Banzhaf, H. Spencer and Kyle Mangum. 2018. "Capitalization as a Two-Part Tariff: The Role of Zoning." Working Paper.

Baum-Snow, Nathaniel and Byron F. Lutz. 2011. "School Desegregation, School Choice, and Changes in Residential Location Patterns by Race." *American Economic Review* 101(7):3019–3046.

Bayer, Patrick, Fernando Ferreira and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy* 115(4):588–638.

Bayer, Patrick, Nathaniel Keohane and Christopher Timmins. 2009. "Migration and hedonic valuation: The case of air quality." *Journal of Environmental Economics and Management* 58(1):1–14.

Bayer, Patrick, Robert McMillan, Alvin Murphy and Christopher Timmins. 2016. "A dynamic model of demand for houses and neighborhoods." *Econometrica* 84(3).

Bayer, Patrick, Robert McMillan and Kim Rueben. 2004. "An Equilibrium Model of Sorting in an Urban Housing Market." Working Paper.

Berry, Steven. 1994. "Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics* 25(2):242–262.

Berry, Steven, James Levinsohn and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63(4):841–890.

Berry, Steven, James Levinsohn and Ariel Pakes. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy* 112(1):68–105.

Bhattacharya, Debopam. 2014. "Nonparametric Welfare Analysis for Discrete Choice." *Econometrica* 83(2):617–649.

Billings, Stephen B. 2018. "Technical Summary - Merging Home Mortgage Disclosure Data to Property Records from Zillow (Ztrax) 1995-2016." Technical Report.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *The Quarterly Journal of Economics* 114(2):577–599.

Bodenheimer, Thomas S. and Mark D. Smith. 2013. "Primary Care: Proposed Solutions to the Physician Shortage Without Training More Physicians." *Health Affairs* 32(11):1881–1886.

Bolduc, Denis, Bernard Fortin and Marc-André Fournier. 1996. "The Effect of Incentive Policies on the Practice Location of Doctors: A Multinomial Probit Analysis." *Journal of Labor Economics* 14(4):703–732.

Borjas, George J. 1985. "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants." *Journal of Labor Economics* 3(4):463–489.

Borjas, George J. 1995. "Assimilation and Changes in Cohort Quality Revisted: What Happened to Immigrant Earnings in the 1980s?" *Journal of Labor Economics* 13(2):201–245.

Borjas, George J. 2015. "The Slowdown in Economic Assimilation of Immigrants: Aging and Cohort Effects Revisted Again." *Journal of Human Capital* 9(4):483–517.

Bronstein, Janet M., Victoria A. Johnson and Crayton A. Fargason Jr. 1997. "How Rural Physicians Compare on Cost and Quality Measures for Medicaid Ambulatory Care Episodes." *Journal of Rural Health* 13(2):126–135.

Bunten, Devin Michelle. 2017. "Is the Rent Too High? Aggegate Implications of Local Land-Use Regulation." Working Paper.

Burfield, W. Bradley, Douglas E. Hough and William D. Marder. 1986. "Location of Medical Education and Choice of Location of Practice." *Academic Medicine* 61(7):545–554.

Calabrese, Stephen, Dennis Epple and Richard Romano. 2007. "On the political economy of zoning." *Journal of Public Economics* 91(1-2):25–49.

Card, David. 2005. "Is the New Immigration Really So Bad?" *The Economic Journal* 115(507):F300–F323.

Card, David and Jesse Rothstein. 2007. "Racial segregation and the black-white test score gap." *Journal of Public Economics* 91(11-12):2158–2184.

Carliner, Geoffrey. 2000. "The Language Ability of U.S. Immigrants: Assimilation and Cohort Effects." *International Migration Review* 34(1):158–182.

Chan, Leighton, L. Gary Hart and David C. Goodman. 2005. "Geographic Access to Health Care for Rural Medicare Beneficiaries." WWAMI Rural Health Research Center Working Paper #97.

Chen, Frederick, Meredith Fordyce, Steve Andes and L. Gary Hart. 2010. "Which Medical Schools Produce Rural Physicians? A 15-Year Update." *Academic Medicine* 85(4):594–598.

Chiswick, Barry R. 1978. "The Effect of Americanization on the Earnings of Foreign-born Men." *Journal of Political Economy* 86(5):897–921.

Chiswick, Barry R. and Paul W. Miller. 1995. "The Endogeneity between Language and Earnings: International Analyses." *Journal of Labor Economics* 13(2):246–288.

Clemens, Jeffrey and Joshua D. Gottlieb. 2017. "In the Shadow of a Giant: Medicare's Influence on Private Physician Payments." *Journal of Political Economy* 125(1):1–39.

Colas, Mark and Kevin Hutchinson. 2018. "Heterogeneous Workers and Federal Income Taxes in a Spatial Equilibrium." Working Paper.

Colas, Mark and Kevin Hutchinson. 2019. "Heterogeneous Workers and Federal Income Taxes in a Spatial Equilibrium." *American Economic Journal: Economic Policy* . Revise and Resubmit.

Cutler, David M. and Edward L. Glaeser. 1997. “Are Ghettos Good or Bad?” *The Quarterly Journal of Economics* 112(3):827–872.

Davis, Morris A., Jesse Gregory, Daniel A. Hartley and Kegon T.K.Tan. 2017. “Neighborhood Choices, Neighborhood Effects and Housing Vouchers.” Working Paper.

Davis, Morris, William D. Larson, Stephen D. Oliner and Jessica Shui. 2019. “The Price of Residential Land for Counties, ZIP Codes, and Census Tracts in the United States.” FHFA Staff Working Paper Series.

Diamond, Rebecca. 2016. “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill:1980-2000.” *American Economic Review* 106(3):479–524.

Diamond, Rebecca and Tim McQuade. 2019. “Who Wants Affordable Housing in their Backyard? An Equilibrium Analysis of Low Income Property Development.” *Journal of Political Economy* 127(3):1063–1117.

Dube, Jean-Pierre, Jeremy T. Fox and Che-Lin Su. 2012. “Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation.” *Econometrica* 80(5):2231–2267.

Dustmann, Christian. 1997. “The Effects of Education, Parental Background and Ethnic Concentration on Language.” *The Quarterly Review of Economics and Finance* 37(Special Issue 1997):245–262.

Dustmann, Christian and Arthur van Soest. 2002. “Language and Earnings of Immigrants.” *Industrial and Labor Relations Review* 55(3):473–492.

Dustmann, Christian and Francesca Fabbri. 2003. “Language Proficiency and Labour Market Performance of Immigrants in the UK.” *The Economic Journal* 113:695–717.

Eltagouri, Marwa. 2017. "A Cubs World Series Baby Boom? Some Parents and Hospitals Think So." *Chicago Tribune* . 26 July 2017.

Epple, Dennis and Holger Sieg. 1999. "Estimating Equilibrium Models of Local Jurisdictions." *Journal of Political Economy* 107(4):645–681.

Evans, Richard W., Yingyao Hu and Zhong Zhao. 2010. "The Fertility Effect of Catastrophe: U.S. Hurricane Births." *Journal of Population Economics* 23(1):1–36.

Falcettoni, Elena. 2018. "The Determinants of Physicians' Location Choice: Understanding the Rural Shortage." Working Paper.

Fernandez, Raquel and Richard Rogerson. 1997*a*. "Education Finance Reform: A Dynamic Perspective." *Journal of Policy Analysis and Management* 16(1):67–84.

Fernandez, Raquel and Richard Rogerson. 1997*b*. "Keeping People Out: Income Distribution, Zoning and the Quality of Public Education." *International Economic Review* 38(1):23–42.

Fernandez, Raquel and Richard Rogerson. 2003. "Equity and Resources: An Analysis of Education Finance Systems." *Journal of Political Economy* 111(4):858–897.

Fischel, William A. 2001. "An Economic History of Zoning and a Cure for its Exclusionary Effects." Working Paper.

Galiani, Sebastian, Alvin Murphy and Juan Pantano. 2015. "Estimating Neighborhood Choice Models: Lessons from a Housing Assistance Experiment." *American Economic Review* 105(11):3385–3415.

Ganong, Peter and Daniel Shoag. 2017. "Why has regional income converge in the US declined?" *Journal of Urban Economics* 102(November 2017):76–90.

Geyer, Judy and Holger Sieg. 2013. "Estimating a model of excess demand for public housing." *Quantitative Economics* 4:483–513.

Glaeser, Edward L. and Bryce A. Ward. 2009. "The causes and consequences of land use regulation: Evidence from Greater Boston." *Journal of Urban Economics* 65:265–278.

Glaeser, Edward L. and Joseph Gyourko. 2002. "The Impact of Zoning on Housing Affordability." Working Paper.

Glaeser, Edward L., Joseph Gyourko and Raven Saks. 2005. "Why Is Manhatten So Expensive? Regulation and the Rise in Housing Prices." *The Journal of Law and Economics* 48(2):331–369.

Goodman, David C., Ellisott S. Fisher, Tomas A. Bubolz, Jack E. Mohr, James F. Poage and John E. Wennberg. 1996. "Benchmarking the US Physician Workforce: An Alternative to Needs-Based or Demand-Based Planning." *Journal of the American Medical Association* 276(22):1811–1817.

Goolsbee, Austan and Amil Petrin. 2004. "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV." *Econometrica* 72(2):351–381.

Green, David A. 1999. "Immigrant Occupational Attainment: Assimilation and Mobility over Time." *Journal of Labor Economics* 17(1):49–79.

Guryan, Jonathan. 2004. "Desegregation and Black Dropout Rates." *American Economic Review* 94(4):919–943.

Gyourko, Joseph, Albert Saiz and Anita Summers. 2008. "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index." *Urban Studies* 45(3):693–729.

Hamilton, Bruce W. 1975. "Zoning and property taxation in a system of local governments." *Urban Studies* 12:205–211.

Hanushek, Eric A., John F. Kain and Steven G. Rivkin. 2009. "New Evidence about Brown v.

Board of Education: The Complex Effects of School Racial Composition on Achievement." *Journal of Labor Economics* 27(3):349–383.

Hanushek, Eric A. and Kuzey Yilmaz. 2015. "Land-use Controls, Fiscal Zoning, and the Local Provision of Education." *Public Finance Review* 43(5):559–585.

Hayward, George M. and Anna Rybińska. 2017. ""Super Bowl Babies": Do Counties with Super Bowl Winning Teams Experience Increases in Births Nine Months Later?" *Socius* 3:1–14.

Heckman, James and Richard Robb. 1985. *Cohort Analysis In Social Research*. Springer-Verlag chapter 5. Using Longitudinal Data to Estimate Age, Period and Cohort Effects in Earnings Equations, pp. 137–150.

Hoxby, Caroline M. 1996. "Are Efficiency and Equity in School Finance Substitutes or Complements?" *Journal of Economic Perspectives* 10(4):51–72.

Hoxby, Caroline M. 2001. "All School Finance Equalization are Not Created Equal." *Journal of Economics* 116(4):1189–1231.

Hsieh, Chang-Tai and Enrico Moretti. 2015. "Why Do Cities Matter? Local Growth and Aggregate Growth." Working Paper.

Hurley, Jeremiah E. 1991. "Physicians' Choices of Specialty, Location, and Mode: A Reexamination within an Interdependent Decision Framework." *Journal of Human Resources* 26(1):47–71.

IHS Inc. 2016. "The Complexities of Physician Supply and Demand: Projections from 2014 to 2025." Association of American Medical Colleges.

Joona, Pernilla Andersson and Lena Nekby. 2012. "Inensive Coaching of New Immigrants: An Evaluation Based on Random Program Assignment." *Journal of Economics* 114(2):575–600.

Kennan, John and James R. Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica* 79(1):211–251.

Kotera, Tomoaki and Ananth Seshadri. 2017. "Educational policy and intergenerational mobility." *Review of Economic Dynamics* 25:187–207.

Kvedar, Joseph, Molly Joel Coye and Wendy Everett. 2014. "Connected Health: A Review of Technologies And Strategies To Improve Patient Care With Telemedicine and Telehealth." *Health Affairs* 2:194–199.

Laditka, James N., Sarah B. Laditka and Janice C. Probst. 2009. "Health Care Access in Rural Areas: Evidence that Hospitalization for Ambulatory Care-Sensitive Conditions in the United States May Increase with the Level or Rurality." *Health & Place* 15(3):761–770.

Li, Phillip. 2012. "Speeding Up the BLP Contraction Mapping." Working Paper.

Liu, Jiexin (Jason), Gail Bellamy, Beth Barnet and Shuhe Weng. 2008. "Bypass of Local Primary Care in Rural Counties: Effect of Patient and Community Characteristics." *Annals of Family Medicine* 6(2):124–130.

Lochman, Alexia, Hillel Rapoport and Biagio Speciale. 2016. "The Effect of Language Training on Immigrants' Economic Integration. Empirical Evidence from France.".

McConnell, Virginia and Margaret Walls. 2005. "The Value of Open Space: Evidence from Studies of Nonmarket Benefits." Resources for the Future Report 30, Washington, DC.

Murphy, Alvin. 2018. "A Dynamic Model of Housing Supply." *American Economic Journal: Economic Policy* . forthcoming.

Newhouse, Joseph P., Albert P. Williams, Bruce W. Bennett and William B. Schwartz. 1982. "Does the Geographical Distribution of Physicians Reflect Market Failure?" *Bell Journal of Economics* 13(2):493–505.

Pathman, Donald E., Thomas R. Konrad, Tonya S. King, Donald H. Taylor Jr. and Gary G. Koch. 2004. "Outcomes of States' Scholarship, Loan Repayment, and Related Programs for Physicians." *Medical Care* 42(6):560–568.

Petterson, Stephen M., William F. Rayburn and Winston R. Liaw. 2016. "When Do Primary Care Physicians Retire? Implications for Workforce Projections." *Annals of Family Medicine* 14(4):344–349.

Petterson, Stephen M., Winston R. Liaw, Robert L. Phillips, David L. Rabin, David S. Meyers and Andrew W. Bazemore. 2012. "Projecting US Primary Care Physician Workforce Needs: 2010-2025." *Annals of Family Medicine* 10(6):503–509.

Piyapromdee, Suphanit. 2017. "The Impact of Immigration on Wages, Internal Migration and Welfare." Working Paper.

Piyapromdee, Suphanit. 2019. "The Impact of Immigration on Wages, Internal Migration and Welfare." Working Paper.

Reschovsky, James D. and Andrea B. Staiti. 2005. "Access And Quality: Does Rural America Lag Behind?" *Health Affairs* 24(4):1128–1139.

Ricketts III, Thomas C., Laurie J. Goldsmith, George M. Holmes, Randy Randolph, Richard Lee, Donald H. Taylor Jr. and Jan Ostermann. 2007. "Designating Places and Populations as Medically Underserved: A Proposal for a New Approach." *Journal of Health Care for the Poor and Underserved* 18(3):567–589.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90(6):1257–1278.

Rogers, Joseph Lee, Craig A. St. John and Ronnie Coleman. 2005. "Did Fertility Go Up After the Oklahoma City Bombing? An Analysis of Births in Metropolitan Counties in Oklahoma, 1990–1999." *Demography* 42(4):675–692.

Rosen, Sherwin. 1979. “Wages-based Indexes of Urban Quality of Life.” *Current Issues in Urban Economics* pp. 74–104.

Rothwell, Jonathan. 2012. “Housing Costs, Zoning, and Access to High-Scoring Schools.” Brookings Metropolitan Policy Program Report.

Saiz, Albert. 2010. “The Geographic Determinants of Housing Supply.” *The Quarterly Journal of Economics* 125(3):1253–1296.

Schuller, Karin, Susanne Lochner and Nina Rother. 2011. Das Integrationspanel - Ergebnisse einer Laengsschnittstudie zur Wirksamkeit und Nachhaltigkeit von Integrationskursen. Technical report Bundesamt fuer Migration und Fluechtlinge.

Schwartz, Mark D. 2011. “Health Care Reform and the Primary Care Workforce Bottleneck.” *Journal of General Internal Medicine* 27(4):469–472.

Schwartz, Mark D., William T. Basco Jr., Michael R. Grey, Joann G. Elmore and Arthur Rubenstein. 2005. “Rekindling Student Interest in Generalist Careers.” *Annals of Internal Medicine* 142(8):715–724.

Segarra, Elan. 2017. “Sanctuaries For Immigrants or Sanctuaries For Criminals?” Working Paper.

Sloan, Frank A. 1970. “Lifetime Earnings and Physicians’ Choice of Specialty.” *Industrial and Labor Relations Review* 24(1):47–56.

Sommers, Anna S., Julia Paradise and Carolyn Miller. 2011. “Physician Willingness and Resources to Serve More Medicaid Patients: Perspectives from Primary Care Physicians.” *Medicare & Medicaid Research Review* 2(1):E1–E17.

Svantesson, Elisabeth and Ted Aranki. 2006. “Do Introduction Programs Affect the Probability of Immigrants getting Work?”.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64(5):416–424.

Traczynski, Jeffrey and Victoria Udalova. 2018. "Nurse Practitioner Independence, Health Care Utilization, and Health Outcomes." *Journal of Health Economics* 58:90–109.

Train, Kenneth E. and Clifford Winston. 2007. "Vehicle Choice Behavior and the Declining Market Share of U.S. Automakers." *International Economic Review* 48(4):1469–1496.

Tubergen, Frank Van and Matthijs Kalmijn. 2005. "Destination-Language Proficiency in Cross-National Perspective: A Study of Immigrant Groups in Nine Western Countries." *Journal of American Sociology* 110(5):1412–1457.

Turner, Matthew A., Andrew Haughwout and Wilbert van der Klaauw. 2014. "Land Use Regulation and Welfare." *Econometrica* 82(4):1341–1403.

Udry, J. Richard. 1970. "The Effect of the Great Blackout of 1965 on Births in New York City." *Demography* 7(3):325–327.

Uscher-Pines, Lori, Andrew Mulcahy, David Cowling, Gerald Hunter, Rachel Burns and Ateev Mehrortra. 2016. "Access and Quality of Care in Direct-to-Consumer Telemedicine." *Telemedicine and e-Health* 22(4):282–287.

Weinstein, Ronald S., Ana Maria Lopez, Bellal A. Joseph, Kristine A. Erps, Michael Holcomb, Gail P. Barker and Elizabeth A. Krupinski. 2014. "Telemedicine, Telehealth and Mobile Health Applications That Work: Opportunities and Barriers." *The American Journal of Medicine* 127(3):183–187.

Wong, Timothy. 2015. "Choice Set Aggregation and Consistent Standard Errors in BLP Models for Micro- and Macro-level Data." Working Paper.

Zhou, Teresa Jing. 2017. "The Doctor is In/Out: Determinants of Physician Labor Supply Dynamics." Working Paper.