NBER WORKING PAPER SERIES

GENTRIFICATION AND NEIGHBORHOOD CHANGE: EVIDENCE FROM YELP

Edward L. Glaeser
Michael Luca
Erica Moszkowski

Working Paper 28271
http://www.nber.org/papers/w28271

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2020

We are grateful for financial support to Harvard University, Harvard Business School, and the Star Grant Fund. We are grateful to Yelp for providing data. Hyunjin Kim provided valuable feedback on this project. Luca has done consulting for tech companies, including Yelp. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Gentrification and Neighborhood Change: Evidence from Yelp
Edward L. Glaeser, Michael Luca, and Erica Moszkowski
NBER Working Paper No. 28271
December 2020
JEL No. H75,J15,L81,R30

ABSTRACT

How does gentrification transform neighborhoods? Gentrification can harm current residents by increasing rental costs and by eliminating old amenities, including distinctive local stores. Rising rents represent redistribution from tenants to landlords and can therefore be offset with targeted transfers, but the destruction of neighborhood character can – in principle – reduce overall social surplus. Using Census and Yelp data from five cities, we document that while gentrification is associated with an increase in the number of retail establishments overall, it is also associated with higher rates of business closure and higher rates of transition to higher price points. In Chicago and Los Angeles especially, non-gentrifying poorer communities have dramatically lower turnover than richer or gentrifying communities. However, the primary transitions appear to the replacement of stores that sell tradable goods with stores that sell non-tradable services. That transition just seems to be slower in poor communities that do not gentrify. Consequently, the business closures that come with gentrification seem to reflect the global impact of electronic commerce more than the replacement of idiosyncratic neighborhood services with generic luxury goods.

Edward L. Glaeser
Department of Economics
315A Littauer Center
Harvard University
Cambridge, MA 02138
and NBER
eglaeser@harvard.edu

Michael Luca
Harvard Business School
Soldiers Field
Boston, MA 02163
and NBER
mluca@hbs.edu

Erica Moszkowski
Department of Economics
Harvard University
Cambridge, MA 02138
emoszkowski@g.harvard.edu

I. Introduction

Does gentrification destroy neighborhoods and generate negative externalities for existing neighborhood residents? Vigdor (2002 and 2010) defines gentrification as an increase in demand to live in a formerly high poverty neighborhood. Rising rents will cause long-term tenants to lose and landlords to gain, but there is no larger welfare gain from preventing neighborhood change. According to this view, if renters were compensated for rising rents, through community land trusts or other social programs, then gentrification could benefit everyone.

Yet protesters argue that the adverse impacts of gentrification go beyond higher rents and include the destruction of community assets, such as ethnic restaurants and social cohesion.[1] This view suggests that gentrification might generate harmful externalities that more than offset the benefits that accrue to landlords.[2] In this paper, we present a model in which gentrification can reduce overall social welfare through an endogenous change in retail amenities, in which less profitable, idiosyncratic stores that generate more consumer surplus are replaced by more profitable, but generic, stores that generate less surplus. The key distinction is that a reduction in the number idiosyncratic stores is like a drop in the number of product varieties, while an increase in the number of generic luxury stores is just an outward shift in the supply curve of a single commodity.

The model shows how welfare-reducing gentrification could happen, but does not imply that welfare is actually being reduced. The model also shows an improvement in electronic commerce will also leads to business closures, as shops that sell tradable goods are replaced by shops that sell non-tradable services. That transition will also increase rents, because the services generate more local surplus, but it will not reduce welfare for either the rich or the poor.

---

[1] Newman and Wyly (2006) discuss the displacement created by gentrification and community opposition that the gentrification process. Betancur (2011) identifies a negative effect of gentrification on "community fabric" in Chicago.

[2] Sullivan (2007) finds a generally positive view of gentrification using survey data, but renters and minorities are more negative than the average respondent. These findings are compatible with the view that rising rents are a primary negative effect of gentrification.

In this paper, we test whether gentrification is associated with changes in the nature of retail trade in five U.S. cities. We use data from Yelp to analyze whether gentrification is associated with a greater number of retail closures, a shift from tradable goods to non-tradable services, and an increase in prices charged by local retail.

We focus on the three largest U.S. cities (New York, Los Angeles, and Chicago) and two smaller cities that are known for their high levels of education and limited local housing supply--San Francisco and Boston. We use Yelp data from 2012 to 2017 to measure store closings, the transition from selling tradable goods to non-tradable goods and changes in price point. We define gentrifying areas as places with high initial poverty, relative to the city as a whole, and high rent growth. In our Online Appendix, we show results using two additional gentrification measures: increasing education and increasing share of white residents. All of the results are broadly similar.

In our large cities – Chicago, New York and Los Angeles – there are significantly more store closings in gentrifying areas than in poor areas which did not gentrify. For example, one-fifth of all retail establishments present in our sample in gentrifying parts of Chicago in 2012 closed between 2012 and 2017. Only eight percent of retail establishments present in non-gentrifying neighborhoods in Chicago in 2012 close over the same time period. Gentrification brings the closure rates in poor areas closer in line with the generally high closure rate in richer areas. It is the closure rates in poor places with low rent growth that are unusual.

We had anticipated that gentrification would have an even larger impact on store closures in the smaller and land constrained cities of Boston and San Francisco, but in those cities we find no meaningful difference in closure rates between gentrifying and non-gentrifying areas. The poorer parts of these cities were already reasonably well educated and expensive in 2012, and all areas in these cities experienced significant increases in rents. One interpretation is that all neighborhoods in Boston and San Francisco are gentrifying and so there are fewer meaningful differences between neighborhoods.

We take two strategies to try to isolate the impact of gentrification. First, we use proximity to the central business district as an instrument. Areas that are close to the downtown, like the Boyle Heights neighborhood of Los Angeles, are gentrifying more quickly, but if central location has a direct impact on store closures then the exclusion restriction will be violated.

Consequently, we also use a matching estimator that compares stores in gentrifying neighborhoods with stores in non-gentrifying neighborhoods that have similar price points, product category, ratings and number of reviews and are located in neighborhoods that are matched on poverty, and proximity to downtown.

Using both a continuous and a discrete definition of gentrification, the instrumental variables estimate of the impact of gentrification on closure is larger in magnitude. For example, our ordinary least squares specification indicates that the closure rate increases by 2.3 percentage points in gentrifying areas, relative to a mean of 17 percent for poor areas overall. The instrumental variables estimate finds that the closure rate increases by 6.1 percentage points. The matching estimator finds a three percentage point increase in the closure rate, which is not statistically distinct from zero. These are meaningful differences, but they are modest in magnitude.

As total establishment growth is higher in gentrifying areas, the closed businesses are being replaced by other firms. Gentrification can cause a meaningful decline in neighborhood amenities if the new establishments are both different and inferior to the old establishments. We look first at transition matrices that enable us to look at specific retail categories, and run regressions focusing on the transition from selling tradables, like clothes and appliances, to selling non-tradable experiences, like a night out.

Over the past 10 years, our cities appear to have responded to the rise of e-commerce with a significant change in the composition of urban retail. Stores that sold goods, which are tradable, closed. Stores that sold experiences, which are not tradable, opened. In our data, the modal new store is a restaurant and even 27 percent of groceries that close are replaced by restaurants.

When we split closures up by the initial store category, we find that gentrifying areas are more likely to see closures of stores that sell tradable goods, but not more closure of stores that non-tradable services. In other words, rising rents appear to be speeding up a process that is going on everywhere. Gentrification is not associated with an increasing more closures among stores that sell non-tradable services. These results are consistent across ordinary least squares, instrumental variables and the matching estimator, but as before the coefficients are larger when using our instrument.

We then test whether, conditional upon closure, gentrification is associated with more transitions in either direction. We do not find that to be the case. Gentrification is changing the composition of local retail by disproportionately closing the stores that sell goods, but not because stores that close are more like to switch to non-tradable services in gentrifying areas. A reasonable interpretation is that stores that sell tradable goods are more likely to be on the margin of survival, so when gentrification pushes up prices, they are the first to close.

The Yelp data does not enable us to look at price changes within an establishment over time, but we can look at price changes when an establishment closes and is replaced by another establishment. We found only modest evidence that Yelp's price measure – expressed by 1, 2, 3, or 4 dollar signs — increased disproportionately in gentrifying areas. For example, 52 percent of establishments in the lowest price category remain in that category when they are replaced in gentrifying areas. Fifty-six percent of establishments in that category remain in that category when they are replaced in poor non-gentrifying areas.

Taken together, these results suggest that rent growth is correlated with slightly higher store closure rates, but little change in the composition or price point of retail stores. We also find no evidence that the number of establishments is declining with gentrification. Gentrification does increase closures, but primarily by hastening the demise of stores selling goods, which appear to be an endangered species in our cities.

Our results do not imply that gentrification comes without costs. Residents of poorer areas who are used to long run stability may be understandably surprised and troubled by the rate of change. We define gentrification by rising rental costs, and these higher costs are likely to hurt most long-term renters. Moreover, the Yelp data is coarse and it is certainly possible that some stores are subtly changing their character in important ways. However, the Yelp data does not suggest the sort of retail Armageddon that is sometimes suggested by anti-gentrification advocates. While our model admits the possibility that strong planning controls that limit change could be welfare improving, our results suggest that the losers from gentrification are more likely to benefit from standard income redistribution rather than from retail related restrictions aimed at preserving neighborhood character.

II. Gentrification and Retail: A Model of Redistribution and Externalities

In this section, we present a simple model of gentrification that is meant to clarify the core question of this paper: does gentrification just redistribute wealth or does it create negative externalities? We highlight two forces for neighborhood change, which appear to be occurring simultaneously in many places. First, more rich people want to move into previously poor neighborhoods. Second, an increased ability to supply traded goods through online retailers has caused stores that sell tradable goods to close.

We assume that there is a single neighborhood with a fixed number of identical homes, denoted H, that are split between rich and poor people. Each person rents exactly one home, so H also captures the population size of the neighborhood. We let r denote the endogenous share of homes that are rented by the rich. The neighborhood has a fixed stock of S stores, which may sell tradable goods (T), high-cost generic non-tradable services (G) or idiosyncratic non-tradable lower-cost services (L). High-end generic services are meant to include expensive coffee chains and restaurants. Idiosyncratic non-tradable stores refer to more unique local shops. We let σ_T, σ_L and σ_G denote the endogenous share of stores that sell traded goods, low cost idiosyncratic non-traded services and high cost generic non-traded services respectively.

Households may buy tradable goods directly (off the internet) or from a local store. We assume that local stores are capacity constrained and can each sell at most one unit of the good. Therefore, the internet price (which we normalize to 1) will determine the local market price of the traded good. Local stores are however able to acquire the goods at a cost of $1 - \tau + \varphi\sigma_T$, where τ represents last-mile shipping premium (which makes it more costly to ship to consumers than to stores). This cost is increasing in the total number of traded goods being sold locally because of limited shipping capacity at the neighborhood level. Profits per store equal $\tau - \varphi\sigma_T$. We will model improvements in e-commerce for consumers as an equivalent increase in the last-mile costs for stores, which is decrease in τ.[3] Since tradable goods are available everywhere,

[3] This assumption enables us to keep the cost of the traded good at one, which is convenient since it is the natural numeraire good.

these stores do not generate social surplus for local residents. Furthermore, all profits are competed away paying for commercial rents. Therefore, these stores are welfare-neutral.

The idiosyncratic stores sell differentiated services and face a marginal cost per unit of service *k.* They have monopoly power over their idiosyncratic experience and charge a price that precludes them from reaching their capacity constraint. The generic luxury stores, however, sell a non-differentiated product, which is meant to reflect complaints about the homogeneity of luxury. They compete on price, and sell to their capacity of 1. Their marginal cost is k_G. We assume that both rich and poor favor the idiosyncratic stores equally, but the rich have a stronger taste for the luxury store. Consequently, the number of luxury stores increases as the rich move into the neighborhood.

Including all three types of stores allows the model to capture both the replacement of interesting local stores with expensive chains and the closing of stores that sell goods. The model treats these as separate phenomena with different welfare consequences. When chain stores replace idiosyncratic stores, consumer surplus can fall because the idiosyncratic stores were each producing a unique good. When generic tradable goods stores (for example, appliance stores or clothing stores) are replaced by an increasingly efficient internet that sells a wider range of products, then social welfare does not decline (at least in the model).

The welfare of low income residents of the neighborhood equals $\underline{Y} + B_L\sigma_L + .5b(A - P_H)^2 + \theta_i - R$, where $\underline{Y}$ is income, B_L is the net benefit generated by each low cost non-tradable good store for each household, *A* and *b* are taste parameters, P_G is the price of the generic non-tradable good, θ_i represents a resident-specific taste shock for living in the community, and *R* represents the rental cost.[4] The taste for living in the community among is drawn from a uniform distribution with a low of 0, a high of θ_0, and a total mass of *H*. Therefore, if the share of the community that is rich equals *r* (and so there are *(1-r)H* poor people in the community) then the

[4] To arrive at this consumer welfare expression, we assume that the utility of the poor is $Q_T + \frac{\vartheta}{1-\epsilon}\sum_j X_j^{1-\epsilon} + AQ_G - .5bQ_G^2 \quad + \theta_i$, where *j* indexes the low cost service stores, Q_T represents the quantity of the traded good, and Q_G represents the total quantity of higher end luxury services. The utility of the rich is $Q_T + \frac{\vartheta}{1-\epsilon}\sum_j X_j^{1-\epsilon} + (A + \omega)Q_G - .5b{Q_G}^2$, with $\omega > 0$ We also assume that all individuals always have enough wealth to consume some of the traded good that is imported from outside the area.

marginal poor resident has a taste for living in the community of $\theta_0 r$. The welfare of poorer individuals if they leave the neighborhood is denoted $\underline{Y} + \underline{U}$.

The welfare of higher income individuals who live in the neighborhood is $\overline{Y} + B_L\sigma_L + .5b(A + \omega - P_G)^2 - R$, where $\overline{Y} > \underline{Y}$ reflects the higher income level and ω reflects the stronger tastes of the rich for luxury services. The utility of higher income individuals if they leave the neighborhood is denoted $\overline{Y} + \underline{U} + \Delta$. The value of Δ reflects the outside options available to the rich. We will model gentrification as a reduction Δ (perhaps due to an increase in prices elsewhere) which pushes more wealthy people into the neighborhood.

Given the utility functions described in footnotes 3 and 4, firms set optimal prices, and earn profits. [5] The equilibrium is described in Appendix A. We assume that both rich and poor live in the neighborhood and that all three types of commerce are present. Proposition 1 describes the impact of gentrification. Proposition 2 describes the impact of increasingly efficient e-commerce.

Proposition 1: Gentrification (a decrease in Δ) will lead to:

1. an increase in the number of rich residents,
2. an increase in the number of generic luxury services,
3. a decrease in the number of idiosyncratic local services,
4. an increase in residential rents if and only if $2b\theta_0 > \omega HB_L$,
5. an increase in the sum of landlord profits and welfare of the poorer residents if and only if $2rb\theta_0 > \omega HB_L$, and
6. no impact on the number of stores that sell tradable goods.

---

[5] The profit maximizing price for idiosyncratic stores is $\frac{k}{1-\epsilon}$ and hence each idiosyncratic store generates profit of $\pi_L = H\epsilon(1-\epsilon)^{\frac{1-\epsilon}{\epsilon}}\vartheta^{\frac{1}{\epsilon}}k^{\frac{\epsilon-1}{\epsilon}}$ and welfare denoted B_L equals to $\frac{1}{1-\epsilon}$ times that amount. For the generic luxury service, the poor will consume $\frac{\sigma_G S}{H} - r\frac{\omega}{2b}$ and the rich will consume $\frac{\sigma_G S}{H} + (1-r)\frac{\omega}{2b}$. The price of the luxury service is $A - b\frac{\sigma_G S}{H} + r\frac{\omega}{2}$, so profits per store equal $A - b\frac{\sigma_G S}{H} + r\frac{\omega}{2} - k_G$. We assume all expressions are positive. The welfare that the rich receive from the luxury service is $\frac{1}{2b}\left(\frac{\sigma_G S}{H} - r\frac{\omega}{2b}\right)^2$ and the welfare that the rich receive from these services is $\frac{1}{2b}\left(\frac{\sigma_G S}{H} + (1-r)\frac{\omega}{2b}\right)^2$

Unsurprisingly, an increase in the demand of the rich to live in this area leads to an increasing number of rich people in the area. Since the rich have a greater taste for generic luxury services, the number of stores selling those products goes up and that crowds out the number of stores selling idiosyncratic services. In the real world, firms selling tradable goods might also be crowded out, but we have assumed a more inelastic supply of tradable goods stores. This assumption highlights the concerns of lost culture raised by anti-gentrification protesters.

That shift in local retail means that gentrification might not even raise local rents. If the shift in retail destroys enough social surplus generated by these idiosyncratic stores, which requires both a high value of idiosyncratic retail to the consumers and a large shift in the number of stores, which are together guaranteed by $2b\theta_0 < \omega HB_L$, then gentrification can actually reduce rents. If by $2b\theta_0 > \omega HB_L$ then rents do go up, and that represents redistribution between the tenants (who get effectively poorer) and the landlords.

In addition to higher rents, gentrification also causes poor tenants to lose the benefits of the closed idiosyncratic stores. Total consumer welfare for the poor is made up of three parts: idiosyncratic preferences for living in the neighborhood, benefits from the idiosyncratic stores, and rent payments. This metric rises with gentrification if and only if $2rb\theta_0 > \omega HB_L$, which is a more stringent condition. If that condition fails to hold, then gentrification is not just redistribution, it is inefficient.

This claim does not prove that gentrification causes efficiency losses. It is a statement of conditions under which it could be inefficient, as well as inequitable. In the model, gentrification is inefficient only if it replaces enough idiosyncratic stores (which generate more consumer surplus) with generic stores (which generate less consumer surplus). This claim motivates the empirical work that follows.

We are wary of confusing the churning of stores because of gentrification with the churning of stores because of the evolution of online retailing. Proposition 2 highlights the impact of improvements in delivering goods directly to consumers.

Proposition 2: Improvements in e-commerce (a decrease in τ) will lead to:

1. a decrease in the number of stores that sell goods,
2. an increase in the number of stores that sell idiosyncratic services,

3. an increase in residential rents, and
4. no impact on the welfare of the poor, the share of the population that is rich or the number of luxury service stores.

In the model, more efficient delivery to consumers (which is equivalent in the model to less efficient delivery to stores) primarily reduces the profitability of stores that compete with online commerce. These shops sell generic traded products and so there is little social loss, at least in the model, from their disappearance. As the idiosyncratic stores respond more readily than the luxury stores (they are more elastically supplied as discussed above), they open when the traded product stores close. The result is that residential rents go up, because the neighborhood has become more fun. The welfare of the poor and rich are unchanged, and the composition of the neighborhood's population doesn't move.

As we turn to the data, we will focus both on overall changes in closure rates and changes in price points. We will also look at whether transitions appear to primarily represent the replacement of local stores with more expensive ones, which is the process predicted by gentrification, or the replacement of stores that sell tradable goods with stores that sell non-tradable services, which is the process predicted by the technological march of electronic commerce.

III. Measuring Gentrification in Five Cities

Our model generates an empirical question: in areas where gentrification is occurring, is there a corresponding and contemporaneous change in the nature of neighborhood retail stores? This empirical question requires us to measure which neighborhoods within a city are gentrifying, and then to measure the correlation between the degree of gentrification and different forms of retail change. There are almost as many ways to measure gentrification as there are papers about the subject, and we cannot hope to use every definition. In this section, we introduce our measure of gentrification, and discuss the extent and geography of gentrification in our five cities: Boston, Chicago, Los Angeles, New York, and San Francisco.

Our measure of gentrification in this paper is based on data from the American Community Survey (ACS). These data are only available at the Census Zip Code Tabulation Area (ZTCA) level in 5-year windows. To take advantage of the most recent data possible, we compare ZIP codes across the 2008-2012 and 2013-2017 vintages of the ACS.

In all five cities, we define the set of neighborhoods that *could* gentrify as those Zip Codes with poverty rates greater than the city's median poverty rate in the 2008-2012 ACS. We made this restriction because the term gentrification is generally associated with a rapid inflow of wealthier residents who push out poorer long-term residents, and wanted to specifically look at neighborhoods where there was a significant share of poor residents that could be displaced. This classification gave us thirteen potential gentrifying Zip Codes in San Francisco and fifteen in Boston. The same definition implied that there were thirty Zip Code that could gentrify in Chicago, fifty-four in Los Angeles and sixty-seven in New York City.

Gentrification is also a dynamic process that requires a dynamic definition. Both the literature and our model suggest that gentrifying neighborhoods can be identified as those with rising housing costs, which reflects significant inflows of richer residents. The flurry of new neighborhood-based housing prices indices issued by the Federal Housing Finance Agency (FHFA) would seem to provide ideal measures for housing cost changes (Larson et al., 2018). Unfortunately, however, these Zip Code and even tract level indices rely on repeat sales of single-family housing, and such housing is relatively limited in high poverty urban areas. For example, less than ten percent of the homes in the Boston Zip Codes that are eligible for gentrification are single family detached and the share is even lower in New York City. Consequently, we use growth in median Zip Code rents between the 2008-2012 and 2013-2017 ACS vintages rather than housing prices as our measure of housing costs. Where we use a discrete measure of gentrification, we label a neighborhood as gentrifying if its rents grow faster than the median growth rate among the potentially gentrifying zip codes in a city (as defined in the previous paragraph). This labeling scheme is illustrated in Figure 1.

Rent growth need not accompany demographic change. An area can change its demographics without becoming more expensive and an area can become more expensive without changing its demographics. Yet rent growth is certainly correlated with neighborhood change and it is a particularly cited cost of neighborhood change. Furthermore, in our model, the poor residents of a gentrifying neighborhood lose both from rising rents and from changing composition of neighborhood retail. The model motivates us to ask whether rent growth is associated with changing composition of retail stores.

We recognize that our measure of gentrification (initial poverty combined with rising rents) is in some ways limited. In popular discourse, gentrification is a highly multifaceted term and can refer to a wide array of demographic and cultural changes. We chose this statistical definition for two reasons. First, it focuses on the demand for residential space and the economic costs faced by residents of these communities. Second, this simple definition makes it easier to interpret and replicate our analysis. We also use other variables from the ACS to control for other factors that could be considered part of gentrification, including demographic variables (including median income and the percentage of residents aged 25 to 34) and information about the housing stock (for instance, the share of housing units with 0-2 bedrooms and the share of single-family homes).

In our online appendix, we show how our rent-based definition differs from two alternative definitions. The first alternative definition of gentrification is based on growth in the share of a neighborhood which is college-educated. Specifically, we take the same set of zip codes that *could* gentrify and order them by percentage-point increase in share college educated rather than rent growth. We define the Zip Codes in top half of this group as gentrifying. This education-based measure of gentrification is in some ways more compatible with the word gentrification itself and its history.[6] Glass (1964) coined the term gentrification to refer to population changes that were happening in the London neighborhood of Islington, where more educated urban professionals were replacing the area's historically working-class population. "Gentry" and "college educated" may not be synonymous, but they are as close as is possible within the heterogeneous American population.

The second alternative definition of gentrification is based on growth in the share of neighborhood residents who are white. We chose this definition because of the common public association of gentrification and ethnic conflict. Of course, there can be gentrification without

[6] McKinnish, Walsh and White (2010) find that gentrification often takes the form of better educated minorities replacing older residents.

ethnic change. For example, Boyle Heights is a gentrifying neighborhood in Los Angeles where both long-term poorer residents and richer new residents are Latino.

There is considerable overlap between the three definitions of gentrification, but there are some discernable spatial differences. Rent-based gentrification is strongly associated with proximity to the city center (city hall in our analysis), while education-based gentrification is more likely to occur in far flung areas. Over the past 20 years, price increases have concentrated close to the urban core (Hipsman, 2017) presumably because rising incomes have increased the demand for short commutes and because of increased demand for urban amenities (Glaeser, Kolko and Saiz, 2001). Our rent-based definition matches areas that are typically described as gentrifying. While the model suggested the gentrification could lower rents by destroying local amenities, choosing only those areas with rising rents allows us to focus on areas where there are winners and losers. All of our three measures are correlated with other measures of demographic change that are highlighted by the Urban Displacement Project (2019). Baum-Snow and Hartley (2020), Meltzer & Capperis (2017) and Meltzer & Schuetz (2012) all provide a deeper analysis of how different measures of demographic change correlate with retail turnover.

Figure 2 provides maps of gentrification based on our rent growth definition. For each city, the lightly colored Zip codes show areas that were above the median city-wide poverty rate in 2012, the dark red zip codes show gentrifying areas, and the orange areas show areas that are poor but not gentrifying. A black circle surrounds the centroid of the zip code containing city hall.

In Chicago, rent-based gentrification occurs closer to the central business district, in areas that are west and south of the city. Those areas were once industrial, and many also had housing projects. In some cases, they are the same streets that the pioneering urban sociologist Robert Park described as "deteriorating" almost 100 years ago in *The City*. The more distant parts of the south and west sides remain poor and non-gentrifying. In Los Angeles, rent based gentrification is also more centered, but in north L.A. the anchor appears to be proximity to eastern suburbs rather than downtown. The Central Business District Zip Code itself is classified as gentrifying, as is the area due east across the river, which includes Boyle Heights. In New York, rent based gentrification occurs predominantly in northern Manhattan and the South Bronx, which are areas with easy north-south subway access to downtown. Brooklyn is also a center for rent-based gentrification. In our smaller cities (which are smaller in both geographical and population terms), gentrification is more spread out through the city. This is intuitive: when the entire city lies within a few miles of city center, distance matters less.

Table 1 shows the characteristics of high poverty areas that do and don't gentrify across the five cities. Our definition of gentrification split the higher poverty tracks into two equal groups based on rent growth. We use one, two, and three stars to denote cases in which the gentrifying and non-gentrifying poorer areas differ in a way that is statistically significant at 10, 5, and 1 percent level. These differences are quite rare in the small cities, even though magnitudes may differ, because the overall sample of zip codes are quite small.

The first two rows show that the share of college-educated residents is generally higher in gentrifying areas, which corresponds closely to standard ideas about the association of rent-based gentrification and demographic gentrification. Education growth is also higher in gentrifying areas in the three large cities, but not in the two smaller cities. In Boston and San Francisco, education levels are already quite high at the start of our sample period even in the poor non-gentrifying areas.

The next four rows show the shares of the population that are white and aged between 25 and 34. In almost every case, the gentrifying areas are both whiter and younger than the non-gentrifying areas. The one exception is that the gentrifying areas of San Francisco have slightly fewer young residents. These differences again suggest that our rent-based definition is close to common conceptions of gentrification.

There are no significant income level differences between gentrifying and non-gentrifying areas at the start of our period, but the gentrifying areas do experience faster income growth. Rising incomes often accompany rising rents, as our model suggests, so this correlation is unsurprising. Nonetheless, it suggests that this definition is working well.

In all cases, gentrifying areas are closer to the central business district, as the maps suggest. Gentrifying areas are experiencing much faster rent growth, which is how they were defined. Perhaps more surprisingly, there were no significant rent differences in the initial time period. The small differences in initial rent partially reflect the fact that housing units are typically smaller in the gentrifying areas. There is little difference in the use of public transportation between gentrifying and non-gentrifying areas.

IV. Data

To investigate the nature of retail change in a given neighborhood, we use establishment-level data from Yelp. Our primary outcome of interest is a binary variable indicating, for each establishment that was present in 2012, whether that establishment closed between 2013 and 2017. Our dependent variables of interest include the establishment's industry (restaurant, café, grocery store, etc), price level on a scale of $ (least expensive) to $$$$ (most expensive) and average numerical rating. We can also see the establishment's name, address, and the number of reviews posted each year.

Our dataset includes a sample of establishments within each city's political boundaries, up through the end of 2018. Yelp obtains its listings either directly from platform users (either business owners or consumers) or by acquiring listing data from other companies and data partnerships. As a result, industry categorization does not follow any particular protocol and does not correspond to any government datasets. Our dataset contains 235 different categorizations, which are not mutually exclusive (for example, "restaurant" and "French restaurant" are both categories reported in the data) and vary tremendously in popularity (some categories, such as "rotisserie chicken", list only one establishment, but others, like "restaurant", list over ten thousand). To simplify analysis and improve power, we aggregate these categories into broad categories for most of the analysis. These broad categories are sit-down restaurants, cafés and coffee shops, dessert places (including bakeries and ice cream shops), hair salons/barbers, fast food (including "fast casual" establishments and chains), bars, groceries (including supermarkets, butchers, vegetable stands, etc.), and convenience stores. These categories represent the most popular categories on Yelp, so while we do not observe every retail establishment, they are likely to have the most representative coverage within our dataset.

In order to incorporate zip-code level information from the American Community Survey, our analysis focuses on the nature of retail change over the 5-year period from 2012-2017. Therefore, we need to know whether the establishment opened before 2013, and whether it closed between 2013 and 2017. Since Yelp does not monitor every storefront at all times, the data we observe about the timing of openings and closings have limited precision. We observe the date that an establishment was added to Yelp's database, not the date it first opened its doors. Furthermore, we have no direct information about when an establishment closes. We therefore proxy for the year of closure with the last year in which it received a review. This imprecise information on timing is a limitation, but since we are looking at the nature of retail change over a 5-year period, the precise date on which an establishment opens or closes should not matter much.

We drop establishments that Yelp adds via data partnerships and acquisitions, because in that case our proxy for establishment entry is not accurate. Therefore, our analysis relies only on crowdsourced data, and there is reason to be concerned that these establishments constitute a selected sample.

Validating the Yelp Dataset

To address concerns about selection onto Yelp platform, we benchmark our Yelp data against the County Business Patterns (CBP). The County Business Patterns use the North American Industrial Classification System (NAICS), which is standard but does not correspond exactly to the classification given in Yelp. To effectively compare the Yelp data and the CBP data, we need to create a crosswalk from Yelp categories to NAICS industry categories (which is what the County Business Patterns does). Our crosswalk is displayed in Table 2.

Overall, the correlation between the number of Yelp establishments and the number of CBP establishments is generally quite high. The basic correlation between stores as measured by Yelp and stores measured by County Business Patterns is strong. Figure 3 shows the correlation between the number of listings on Yelp and the number of corresponding businesses in the County Business Patterns. The correlation coefficient is 0.801 and the estimated coefficient from a regression line is 0.99. As described by Glaeser, Kim and Luca (2018), the match between Yelp establishments and county business patterns is not perfect, but it is still quite close.

Table 3 then breaks out the ratio of Yelp establishments to County Business Patterns establishments by year and category, for gentrifying, poor non-gentrifying, and rich zip codes. The average of the ratio of number Yelp establishments to number of CBP establishments is usually less than one. After all, the County Business Patterns count all businesses while Yelp's establishments are mostly crowdsourced. Sometimes, however, the ratio is greater than one, and we attribute this to the fact that our crosswalk is imperfect and because CBP may miss the smallest establishments. We expect Yelp to have the best coverage of restaurants compared to other categories, and indeed we see that the Yelp establishment count to CBP establishment count ratio is very close to 1 in all cases.

However, the correlation between Yelp establishment counts and CBP establishment counts is lower for poor, non-gentrifying neighborhoods relative to gentrifying and rich neighborhoods. So there is a possibility that Yelp measurement may artificially rise due to gentrification. It could also be that Yelp measurement is more accurate for non-tradable services establishments than for retailers of tradable goods, and we do find a stronger correlation between the number of Yelp and County Business Patterns establishments in a given zip code for these categories.

We perform most of our analysis with the Yelp data because the County Business Patterns does not track individual establishments over time. Rather, the County Business Patterns reports only the total number of establishments in each NAICS category in a given year at the zip code level. Therefore, it can only be used to examine the net change in number of establishments, and is not suitable for separately analyzing opening and closing rates. It also does not report other features of the establishments, and it is not possible to follow a given establishment over time.

Overall Establishment Growth and Gentrification in the Five Cities

In Tables 4a and 4b we show the differential growth rates of the number of establishments across categories in all five cities, using the establishment counts from the County Business Patterns. Once again, we use stars to denote the statistically significant differences between gentrifying areas and the other two classes of Zip codes. The results for Chicago are most dramatic. The non-gentrifying poor areas see an overall decline in the number of establishments of eleven percent, and much larger declines in the numbers of bars, cafes, grocery stores and hair salons.[7] These are sharply different from gentrifying areas where the establishment numbers stayed far more constant.

In Los Angeles, bars were the only category in which shrinkage was dramatically higher in non-gentrifying poor areas. In New York, both gentrifying and non-gentrifying poor areas experienced more establishment growth than rich areas. Gentrifying areas saw faster growth in

[7] These results can be compatible with the low closure rates reported by Yelp data for these areas if stores with Yelp reviews are more successful.

bars and cafes, and non-gentrifying areas had a slightly increase in the number of groceries, while the number of groceries shrank in gentrifying areas.

In Boston and San Francisco, there are essentially no statistically significant differences between gentrifying areas and either non-gentrifying poor areas or rich areas. Fast food restaurants grew more quickly in non-gentrifying areas of Boston. Hair salons grew in gentrifying areas of Boston and San Francisco. These establishment change figures show little difference between gentrifying and non-gentrifying areas outside of Chicago.

With the County Business Patterns data, we are able to compare the percentage change in number of establishments in our study period (2012-2017) with the 5 years prior (2006-2011). There are fewer statistically significant differences in the percent change in number of establishments across gentrifying, poor, and rich neighborhoods in the pre-period, particularly for the larger cities.

V. Is Gentrification Associated with Higher Closure Rates?

As we have discussed above, the Yelp data is reasonably good at identifying businesses that once existed and that now have closed, although there will be some stores that have closed that have not been reported. We focus on closings in this section. As the Yelp data can also measure stores replacements at the same address, we can see whether a new store opens and classify the category of store and its price level. We will focus on replacement in the next section.

The Yelp data is less good at identifying business openings or the overall change in retail activity within an area. New stores may be added to Yelp because of increased reporting, not because of a new store opening. The danger of false openings is particularly dangerous when looking at gentrifying areas, since new, more educated residents may be more likely to report their retail experiences on Yelp. Consequently, in Section VII, we will use County Business Patterns to measure the overall changes in the level of business activity.

We perform three exercises within this section. First, for all five cities separately, we look at closure rates by retail category. We then compare closure rates between poor gentrifying, poor non-gentrifying and non-poor areas. We perform our regressions at the store level, and we regress closing rates on changes in rental price level at the zip code level and the interaction between that rate and the initial poverty level. We also control for other characteristics in the area. For these regressions, we treat gentrification as a continuous variable and focus on the interaction between the initial poverty rate and the growth in rental prices. The continuous version of the gentrification variable avoids the arbitrariness of relying on discrete cutoffs, but our results are largely unchanged when we perform regressions with discrete gentrification measures. We will return to these discrete measures, when we compare the ordinary least squares and instrumental variables results with those obtained using a matching estimator.

We instrument for rising prices levels with distance to City Hall and a polynomial of that distance. As we discussed in Section III, distance to City Hall is significantly correlated with price growth during this time period. These regressions cannot separately identify whether rental price growth or proximity itself are causing closings, but they can address the fear that the exogenous changes in the retail mix are themselves leading to rising price levels.

Closures in Gentrifying and Non-Gentrifying Areas

In Table 5, we show closure rates across the five cities for gentrifying poor areas, non-gentrifying poor areas and rich areas. The closure rate is defined as the share of establishments open before 2013 that closed between 2013 and 2017. As discussed above, gentrification is defined as having rent growth in the zip code that was over the city-level median. We use one star to indicate that the difference between closure rates was significant at the ten percent level, two stars to indicate significance at the five percent level, and three stars to indicate significance at the one percent level. The stars in the middle column indicate statistical difference between gentrifying and stable poor areas. The stars in the left column indicate statistical difference between gentrifying and rich areas.[8]

Part a of the table shows our results for Chicago, Los Angeles and New York. Part b shows results for the smaller cities of Boston and San Francisco.

[8] These differences in statistical significance was established using a city-by-city linear probability models including only stores in initially poor areas where closure was regressed on a dummy variable indicating gentrification.

The first three columns show our results for Chicago. In every category, closure rates are higher in gentrifying areas than in poor non-gentrifying areas. In all but two categories (bars and convenience stores), the differences are statistically significant. In all but one category (restaurants), the closure rate is rate is twice as high in gentrifying areas than in non-gentrifying poor areas. In most cases, gentrifying areas do not have more closures than rich areas, and overall the closure rates are higher in rich areas.

Chicago is remarkable in that the retail stores in its non-gentrifying poor areas are extremely stable. Every sector, except for restaurants, had a closure rate below ten percent in poor areas that didn't gentrify. Chicago has particularly low commercial rents in its poor areas, and this may keep businesses in operation for long periods of time. Consequently, residents of poorer parts of Chicago may have had an expectation of permanence that broke down when gentrification occurred. Yet when poor areas of Chicago gentrify, the closure rates converge to the city-wide norm, not to some exceptionally high level.

The middle panel shows results for Los Angeles. The results are similar to those in Chicago in that closure rates are higher in gentrifying areas than in non-gentrifying poor areas for all of the categories. In five of the categories and for all of the establishment types together, the difference in closure rates is statistically significant. In Los Angeles, just like Chicago, the gentrifying areas do not have higher closure areas than the rich areas. Once again, the non-gentrifying poor areas uniquely have low closure rates.

The third panel shows the weaker results for New York City. While closure rates are always higher in gentrifying areas in New York, the differences are statistically significant for only three of the categories and for the overall closure rates. In New York, the closure rates are typically highest in the rich areas. In most cases, gentrifying areas have significantly lower closure rates. Once again, the rich areas are less stable than the poor areas and gentrifying areas lie between these extremes. In New York, gentrifying areas fall more clearly in the middle, whereas in Chicago and Los Angeles, gentrifying areas have closure rates that are closer to those in rich areas.

The bottom panel shows our results for Boston and San Francisco. In these smaller places, there are virtually no statistically distinct differences in the closure rates between the three types of areas. There are, of course, fewer areas and so we should expect to see less statistical significance, but the point estimates are also quite close. We expected to see more closures in gentrifying areas in these cities because of the paucity of land, but instead we found that there were higher closure rates everywhere.

Regression Analysis of Closure Rates

We now turn to our regression analysis of closure rates across all five cities. Our basic regression treats a business as a unit of observation and the model predicts whether a business that was open on December 31, 2012 closes between January 1, 2013 and December 31, 2017. Our key independent variable is the interaction between the demeaned rent growth between 2012 and 2017 and the demeaned initial poverty rate. We demeaned to ease the interpretation of our controls for the initial poverty rate and the growth in rents.

More formally, our main probit specification is given by:

$$\begin{aligned}\Pr(Closed_{iz}^{2013-2017}) \\ &= \Phi(\beta_0 + \beta_1 RentGrowth_z^{2012-2017} + \beta_2 PovertyRate_z^{2012} \\ &+ \beta_3 (PovertyRate_z^{2012} \times RentGrowth_z^{2012-2017}) + \delta X_{iz})\end{aligned}$$

where *i* indexes establishments and *z* indexes zip codes. As we have chosen to define gentrification as high rent growth in initially poor areas, we are primarily interested in the effect of the interaction between initial poverty and subsequent rent growth on closure probability. This effect is captured by the parameter is β_3. The interaction term will be high for areas which have high initial poverty and high rent growth over the study period) and low for both rich areas (which have low poverty rates) and poor non-gentrifying areas (which have low rent growth).

The expression X_{iz} denotes a vector of controls. We control for the initial density of Yelp establishments in the Zip Code and the density of that category of establishment in the Zip Code, as well as initial percent college educated, the initial share of the population aged between 25 and 34, the initial median income and the initial percent white. We also include fixed effects for the different retail sectors, different cities and the price level of the establishment as it is categorized in the Yelp data. All standard errors are clustered at the Zip Code level.

Table 6 presents the results of a probit model following this specification. The first regression shows the impact of gentrification on the overall closure rate for our entire sample of establishments. The initial poverty rate has a slight negative coefficient on the closure rate, capturing the lower turnover rates in high poverty areas that were shown in Table 5. The overall impact of change in rental price on closure is positive, but not statistically distinct from zero. The lowest cost establishments had a higher closure rate, and none of the other control variables were statistically significant.

The interaction between rent growth and the initial poverty level, which is our measure of gentrification is positive, but small and on the margins of statistical significance. This coefficient has the interpretation that a one percentage point increase in median rents is associated with a .15 percentage point reduction in the closure rate in a neighborhood with a five percent poverty rate. Furthermore, the same rent growth is associated with a .09 percentage point increase in the closure rate in a neighborhood with a 30 percent poverty rate. The difference is small, but it does indicate that gentrification is associated with a modest increase in closures, relative to neighborhoods that are not gentrifying (either because they were rich to begin with or because rents didn't rise much). We will return to this point when we discuss our regression specifications using our discrete measure of gentrification.

In Appendix Table B.1, we reproduce this regression using only our three larger cities (Chicago, Los Angeles, and New York) where we have already observed a larger link between gentrification and closures. In these cities, the interaction is larger and more strongly significant. Nonetheless, controlling for other factors, we still only find that a one percentage point increase in rents is associated with a 0.2 percentage point increase in the closure rate for a neighborhood with a 30 percent poverty rate.

The most significant control is the initial share of the population with a college degree, and more education areas appear to be correlated with higher closure rates.[9] In Appendix Table B.1, we find that the interaction is again stronger and more statistically significant if we restrict our analysis to the largest three cities.

---

[9] We cannot rule out that this represents higher reporting of closures among more educated citizens.

The model suggested that the welfare costs from shutting highly idiosyncratic non-tradable businesses might be significant but that closures of stores that focus on ordinary tradable goods might have little welfare effect. The model also noted that we might confuse the impact of gentrification, which should replace idiosyncratic non-tradable service stores with luxury service stores, and electronic commerce, which should replace all tradable goods stores with non-tradable service stores. To examine this possibility, we now split our sample into "tradable goods selling" and "non-tradable goods selling" establishments.

We include convenience stores, delis, and grocery stores in the tradable goods category. We include all restaurants (including fast food), bars, dessert stores, cafes, and hair salons as non-tradable service-selling establishments. This distinction does not capture the division between higher-end and idiosyncratic service stores that we highlighted in the model. Presumably, both types of stores are included in the group of non-tradable services. It is, of course, possible that some of the "tradable" stores do create special consumer value (certainly locals might be attached to their corner convenience store), but it is reasonable to expect that long-term residents are more attached to local stores that supply non-tradable services relative to appliance stores that sell wares that can increasingly be bought on the internet.

Regression (2) of Table 6 shows the results for establishments that sell tradable goods. The coefficient is much larger than for all closures, but the effect remains only marginally significant because of the smaller sample. Because the coefficient on rent growth has become more negative, we find that a one percentage point increase in median rents is associated with a 0.16 percentage-point reduction in the closure rate in a neighborhood with a 30 percent poverty rate. In a neighborhood with a five percent poverty rate, the same percentage point change in rents is associated with a 0.6 percentage point reduction in the closure rate.

Regression (3) shows results for establishments that sell non-tradable services. The coefficients are smaller than for all establishments and indistinguishable from zero. The comparable coefficients in Appendix Table B.1 are also not significantly different from zero. Gentrification appears to be associated with more closures primarily for establishments that sell tradable goods, perhaps because the newer residents are more mobile and able to buy goods elsewhere. Within the framework of the model, these findings are more compatible with the impact of electronic commerce (which caused closures of tradable goods stores) than gentrification. Yet it is certainly possible that gentrification is causing a shift within the tradable goods category, and it is possible that the closure of non-tradable service stores is harming some long-term residents.

Robustness Checks

In this section, we test the robustness of the correlations from Table 6 in a number of ways. First, we run alternative regression specifications (a linear model and a linear IV) with the same controls as in table 4. These results are displayed in Table 7. We find similar patterns: the coefficients on the interaction of the poverty rate and rent growth are marginally statistically significant for the overall sample and for tradable-selling establishments. The core interaction coefficient is not significant for the non-tradable service sample.

Linear Regression Specification

In the first 3 columns of Table 7, we test the robustness of the store closure regression results using a linear probability model:

$$
\begin{aligned}
Closed_{iz}^{2013-2017} &= \beta_0 + \beta_1 RentGrowth_z^{2012-2017} + \beta_2 PovertyRate_z^{2012} \\
&\quad + \beta_3 (PovertyRate_z^{2012} \times RentGrowth_z^{2012-2017}) + \delta X_{iz} + \varepsilon_{iz}
\end{aligned}
$$

This makes the coefficients easier to interpret and more directly comparable for the instrumental variables estimates that follow. The results are again similar to the probit model. The gentrification interaction coefficient is small and modestly significant for all establishments. The coefficient is larger for tradable goods selling establishments and smaller for non-tradable service selling establishments. Again, the coefficient for non-tradable selling establishments is not significantly different from zero.

Instrumental Variables Estimates

One challenge with interpreting the previous results is that the changing mix of establishments could itself change rent levels and influence gentrification. Consequently, we turn now to a linear instrumental variables strategy that relies on well-known spatial patterns in gentrification. In the second 3 columns of Table 7, we use non-linear transformations of distance to downtown, and distance to downtown interacted with initial poverty rates, as an instrument for rent growth and for the interaction between rent growth and the initial poverty levels.

This instrumental variables strategy is imperfect, because the instrument could fail to satisfy the exclusion restriction. Distance to downtown could have an independent impact on restaurant closures, especially if commuters have become more likely to shop near their workplace. Moreover, we cannot know that the gentrification effects of being downtown work solely through rising rents. For example, places that are closer to downtown might experience demographics that shift more than their rents rise. We consider this issue a lesser problem because we are interpreting our results not as a causal estimate of the impact of rising rents, but rather as an attempt to statistically capture the correlation between gentrification and store closures in higher poverty areas.

Column 4 in Table 7 shows the IV results for all establishments. We see that the coefficient is larger than the linear probability shown in column 1, and statistically significant. It is also significant in Appendix Table B.2, which reproduces this exercise just for the large cities. The coefficients for goods producing stores is significant both in Table 7 and Appendix Table B.2. The coefficients for experience producing stores is insignificant in Table 7, but it becomes significant in Appendix Table B.2. The magnitudes are all somewhat larger than in the ordinary least squares regression.

Overall, our instrumental variables results largely corroborate the results found using ordinary least squares. There does appear to be a positive impact of gentrification on closures, especially for goods producing stores. As with Table 4, the complaints about store closings have some basis in the data. Yet these effects are small, and seem unlikely to generate the large welfare losses that were suggested as possibilities in our model.

Discrete Analysis

In Table 8, we perform similar analyses as in Table 6, but we use a discrete definition of gentrification and limit the sample to only high-poverty neighborhoods. This allows us to compare poor non-gentrifying neighborhoods with gentrifying neighborhoods, rather than lumping rich neighborhoods and poor non-gentrifying neighborhoods together as we did in our continuous analysis. We find that in our linear regression and linear IV specifications, gentrifying neighborhoods have significantly higher closure rates than poor non-gentrifying neighborhoods. We also perform a matching exercise in which we match observably similar restaurants in different areas of the same city, but find no significant effect of gentrification in that analysis. Combining these results with those from our continuous analysis, it appears that gentrifying neighborhoods do not have unusually high closure rates relative to other richer neighborhoods. Rather, if anything, poor non-gentrifying neighborhoods have unusually *low* closure rates relative to all other neighborhoods.

Matching Estimator

The instrumental variables estimate is meant to address the possibility that closing restaurants were pushing rents up rather than the reverse. Yet that estimator focused on areas that were clearly defined by having different locations. We now take a different approach and attempt to compare gentrifying areas that were quite similar to them at the start of our sample period.

We use a matching estimator to compare pairs of establishments that are observably similar, except that one is located in a gentrifying area and one is located in a non-gentrifying area. We match establishments exactly on city, category (restaurant, grocery, etc), Yelp price point ($-$$$$) and year added to the Yelp dataset. Establishments which have no exact match for any of these variables are dropped. We match coarsely on initial poverty rate, Yelp rating, number of reviews, and distance to city hall.

We use a coarsened-exact match for establishments in different neighborhoods. We start with the sample of all establishments open in 2012 in gentrifying and poor, non-gentrifying areas (a total of 16,338 establishments), of which 6,077 are located in non-gentrifying areas and 10,261 are located in gentrifying areas. We then match these establishments without replacement.

We follow Austin (2011) and drop any observations for which we cannot find a match within 0.2 standard deviations on any given dimension, except for the initial poverty rate. For this variable, since it varies only at the zip code level, if we use 0.2 standard deviations as our cutoff, we drop a lot of observations and lose statistical power. Instead, we relax the cutoffs for this dimension and accept any match within 1 standard deviation. In our data, this procedure results in 2,508 matched establishment pairs.

Matching on establishment-level characteristics attempts to minimize the differences between those establishments. Comparing within city, category, and price point is natural. Matching on year added to the Yelp database allows us to compare establishments that have been on Yelp for a similar amount of time, and therefore have similar levels of digital marketing. We match coarsely on Yelp rating and number of reviews to ensure that we are looking at establishments that are similar in quality and popularity. We do not match on zip code characteristics, since our measure of gentrification is a function of zip code characteristics. Appendix table B.3 shows that the final samples of matched establishments are balanced on establishment characteristics.

To examine the impact of being located in a gentrifying or non-gentrifying area on restaurant closure, we estimate the following regression equation, where *i* indexes establishments, *p* indexes pairs, and *z(i)* denotes the zip code in which *i* is located:

$$Closed_{ip}^{2013-2017} = \alpha_0 + \alpha_1 Gentrify_{z(i)}^{2012-2017} + \lambda_p + \epsilon_{ip}$$

The outcome variable is a dummy for whether establishment *i* closes between 2013 and 2017. $Gentrify_{z(i)}^{2012-2017}$ is a dummy variable indicating whether the zip code in which *i* is located gentrifies between 2012 and 2017. The pair-level fixed effect λ_p absorbs any initial differences between pairs. The coefficient β_1 tells us how the probability of closure changes when the establishment is located in a gentrifying area. In other words, it is a "within" estimator: it tells us, within a pair of observably similar restaurants, whether gentrification increases its closure rate. We cluster standard errors at the pair level.

The results are presented in the third column of Table 8. Our estimate of α_1 is positive, small and not statistically significant. Given our previous results, it seems that the "between" effect is larger than the "within" effect: gentrification is likely to impact some types of establishments more than others. This matching estimator again largely supports our basic finding of a modest positive correlation between gentrification and closure rates.

VI. Gentrification and Changing Mix of Retail Businesses

We now turn to gentrification and the changing mix of retail businesses. Our model suggests that gentrification could yield inefficiencies and large welfare losses if community businesses are replaced by higher end stores that deliver less social welfare for nearby residents. We test for this possibility by looking at stores that close. We match closed stores with new stores that occupy the same location, and then look both at changes in the nature of retail and changes in the price level, as measured by the number of dollar signs in the Yelp rating.

Business Transitions across Categories

Tables 8a, 8b and 8c show the transition matrices across Yelp categories for all neighborhoods, gentrifying neighborhoods and neighborhoods that were poor and did not gentrify. Establishments that close are almost always replaced by another establishment, although there are small numbers of transitions into the "none" category, which means that no replacement establishment was entered into Yelp's database between a store's closure and the end of 2018. The share going to none is always less than five percent, except for convenience stores that close in gentrifying areas, which transition to "none" nine percent of the time. We first look at transition matrices for a fine categorization of sectors. We then turn to regressions that look at movement across broader categories of retail.

Table 9a shows the results for all transitions. When almost all types of retail establishments close, their modal replacement is a restaurant. This fact surely reflects the dominance of restaurants in Yelp listings and in our sample, but there also appears to be a general trend away from tradable goods-selling establishments towards non-tradable service establishments, which perhaps reflects the rise of e-commerce. The one prominent exception is hair salons, which are far more likely to remain hair salons than to transition to restaurants. We were somewhat surprised that other goods-selling stores, such as convenience stores, delis and groceries, were all somewhat more likely to become restaurants than to remain within their category. These transitions seem again more compatible with the power of e-commerce than with gentrification alone.

Tables 9b and 9c show the transition tables for gentrifying and non-gentrifying neighborhoods. The tables are extremely similar. The largest differences, to our eyes, are that convenience stores and hair salons were somewhat less likely to be replaced by stores in their category in non-gentrifying areas. Eleven percent of convenience store locations remained in that category when they closed in gentrifying areas and only four percent remained in that category when they closed in non-gentrifying areas. In non-gentrifying areas, convenience stores were more likely to become restaurants.   Fifty-three percent of hair salon locations reopened as hair salons after they closed in gentrifying areas, whereas only forty percent of hair salons continued in the hair business in non-gentrifying areas.   Grocery stores were more likely to remain as groceries in non-gentrifying areas.

In Table 10, we test in a regression framework whether there is any statistically significant difference between gentrifying and non-gentrifying areas in the transition between tradable goods selling establishments and non-tradable service selling establishments. The first column asks whether establishments are more likely to sell non-tradable services after a closure. The constant term shows that slightly over 80 percent of establishments sell non-tradable service after closure in non-gentrifying areas. The coefficient on an indicator that takes on a value of one if the neighborhood is gentrifying indicates that this probability is 0.1 percent lower in gentrifying areas, but this difference is not statistically significant.

In the second regression, we control for whether the establishment initially sold tradable goods and we find that this reduces the probability of transitioning to a non-tradable service-type establishment after closure by 19.8 percentage points.   This coefficient still means that over 65 percent of tradable goods-selling storefronts are replaced by non-tradable service-selling establishments after closure, which reflects the general urban shift from selling tradable goods to selling non-tradable services. In the last regression, we interact the gentrification variable with whether the establishment initially sold tradable goods. In none of the regressions is the gentrification variable or its interaction statistically significant or large in size.

Business Transitions across Price Points

We now look at whether the businesses that close are replaced by stores that are more expensive in gentrifying areas. Tables 11a, 11b and 11c show transition matrices using the dollar signs. Table 11a shows the transition matrix across all neighborhoods. The overwhelming number of establishments begin and end with one or two dollar signs. Stores that start with one dollar sign are about equally likely to be replaced by a one dollar sign store as by a two dollar sign store. Stores that start with two dollar signs have a seventy-one percent chance of remaining a two dollar sign store. Overall, there are about 500 fewer one dollar sign stores after closure than before closure, which suggests that prices are rising overall. Somewhat surprisingly, stores that start with three or four dollar signs have a greater than seventy percent probability of reopening as a one or two dollar sign store, so there is substantial downward price movement.

Tables 11b and 11c show that there are small differences between gentrifying and non-gentrifying poorer areas. The probability of transitioning to a one dollar sign store is higher in non-gentrifying areas no matter what the initial price point. The probability of adding dollar signs, however, is not appreciably higher in gentrifying areas. Even stores that start with one dollar sign are only four percent more likely to transition to two dollar signs in gentrifying areas.

Table 11d presents an ordered logit regression, which we use to assess whether the transition matrices for gentrifying and poor non-gentrifying areas are statistically significantly different. The coefficients on the gentrification dummy and its interaction terms are generally not statistically significant, with the exception of storefronts that initially contained $$$$-type establishments (which is a very small sample). The probability that a storefront's new tenant has a higher number of dollar signs on Yelp is not statistically significantly different between gentrifying and non-gentrifying poor neighborhoods.

Since the ordered logit is difficult to interpret, we turn to a simpler regression form to assess whether gentrification is associated with increases in the number of dollar signs within storefronts when there is establishment turnover. Table 12 uses a linear model where the outcome variable is an indicator for whether the number of dollar signs increases when a new store replaces an old store. We include only establishments in neighborhoods that either gentrified or started with a high poverty rate and did not gentrify. The first three regressions show that rent growth is not associated with increases in dollar signs overall, for tradable goods-selling stores or for non-tradable service-selling establishments. In the last three regressions, we include an indicator variable that takes on a value of one if the store initially had one-dollar sign. That indicator variable is significant, suggesting some mean reversion of prices, which is also suggested in the transition tables. When we interact this variable with rent growth, we also find that the coefficient is marginally significant for all establishments, driven mostly by non-tradable service-selling establishments. Again, there is little evidence that store transitions are substantially different in gentrifying areas overall, although there does seem to be a modest upgrading of the lowest cost stores in gentrifying areas, and that is significant for non-tradable service-type establishments.

Table 13 adds controls and zooms in on within-category transitions for cleaner comparisons. In this case, we do see a positive impact of gentrification on price increases. The regressions allow us to compare 2 identical stores in neighborhoods with similar poverty rates but different rates of median rent growth. Column 1 of the table indicates that a storefront located in a neighborhood with a one percentage point higher growth in rents has a 4.5 percentage-point higher chance of being replaced with a higher-price-point store. This coefficient is significant, but the magnitude is modest. The second regression shows results only for establishments that initially sold tradable goods. The coefficient is larger, but statistically insignificant. The third regression includes only establishments that initially sold non-tradable services. In this case, the coefficient is significant and slightly smaller than for all establishments.

The fourth regression looks only at establishments that remain within their own category (e.g., a storefront in which a restaurant is replaced by another restaurant, or a grocery store is replaced by another grocery store). This regression eliminates price point changes that are associated with switching retail categories. In this case, using the same thought experiment as in column one, the store in the faster-growing neighborhood has a 6.7 percentage-point higher chance of being replaced with a higher-price-point store. If the establishment started selling tradable goods and remained in its category then the impact of gentrification on price is actually negative, which is shown in the fifth regression. In the sixth regression, we find that if the establishment initially sold experiences, and the new store is in the same retail category as the old store, then the coefficient is positive and larger.

Gentrification is associated with slight price upgrading. This finding gives statistical weight to the complaints that some cheaper stores are lost through the gentrification process. Yet across all specifications, the impact of gentrification on price points is quite modest and the impact on shifting retail categories is even smaller. Gentrification is associated with slight increases in price points, not a massive shift in store type, at least in the short run.

VII. Conclusion

Gentrification creates winners and losers. When rents rise in poorer areas, landlords benefit and long-term tenants lose. In this paper, we consider the possibility that changing neighborhood character causes poorer incumbents to lose even more than their rent increases would suggest. Our model shows that gentrification can be particularly welfare decreasing if it causes a large shift in the nature of neighborhood retail, away from idiosyncratic non-tradable services to more generic luxury services, and if residents particularly value proximity to the stores that close.

We tested the model using Yelp data, by examining whether gentrification is associated with high closure rates and shifts in the nature of local retail. Closure rates are indeed higher in poorer areas that are experiencing price increases than in areas that are poor and static. In a regression framework with closures as the dependent variable, we find a statistically significant, albeit small, interaction between initial poverty and the rental price increase, meaning that rent increases cause more closures in initially poor places. The key fact is not that stores close so much more quickly in gentrifying areas, but they close much less in poorer areas that have stable prices, especially in our larger cities. This fact suggests that without gentrification stores in poor neighborhoods are just subject to much less competition than stores in rich neighborhoods.

There is little evidence for extreme changes in the character of the retail stores. Most stores that close are replaced by restaurants in all areas. Gentrifying areas experience a slight switch away from goods providing stores, like groceries, into experience providing stores, like cafes, but the difference with non-gentrifying areas is modest. When low price point stores close in gentrifying areas, they are more likely to be replaced by higher price point stores, but there is no overall price point growth associated with gentrification. Gentrification is also associated with an increase in the number of retail establishments.

While the evidence is sufficiently mixed that it can support a variety of interpretations, our conclusion is that gentrification’s impact on retail mix is modest, at least over our short five-year window. Perhaps the longer-term impacts will be much larger, but there is little so far in the evidence to suggest that changing retail mix is a primary cost of gentrification. Rising rents seem likely to generate a far larger loss for local residents. Those losses represent a redistribution from tenants to landlords. Politicians can reverse that redistribution by using cash transfers, without resorting to regulating the normal churning of retail establishments.

References

Austin, Peter C. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat*. 2011 Mar-Apr; 10(2): 150-61.

Baum-Snow, Nathaniel and Daniel Hartley. "Accounting for central neighborhood change, 2980-2010." *Journal of Urban Economics*, no. 117 (2020).

Betancur, John. "Gentrification and community fabric in Chicago." *Urban Studies* 48, no. 2 (2011): 383-406.

Dragan, Kacie, Ingrid Ellen, and Sherry A. Glied. *Does gentrification displace poor children? New evidence from New York City Medicaid data*. No. w25809. National Bureau of Economic Research, 2019.

Ellen, Ingrid Gould, and Katherine M. O'Regan. "How low income neighborhoods change: Entry, exit, and enhancement." *Regional Science and Urban Economics* 41, no. 2 (2011): 89-97.

Ellen, Ingrid Gould, and Katherine O'Regan. "Gentrification: Perspectives of economists and planners." In *The Oxford Handbook of Urban Economics and Planning*. Oxford University Press, 2012.

Ellen, Ingrid Gould, Keren Mertens Horn, and Davin Reed. "Has falling crime invited gentrification?." *Journal of Housing Economics* 46 (2019): 101636.

Freeman, Lance, and Frank Braconi. "Gentrification and displacement New York City in the 1990s." *Journal of the American Planning Association* 70, no. 1 (2004): 39-52.

Freeman, Lance. "Displacement or succession? Residential mobility in gentrifying neighborhoods." *Urban Affairs Review* 40, no. 4 (2005): 463-491.

Glaeser, Edward L., Jed Kolko, and Albert Saiz. "Consumer city." *Journal of economic geography* 1, no. 1 (2001): 27-50.

Glaeser, Edward, Hyunjin Kim and Michael Luca. "Measuring Gentrification: Using Yelp Data to Quantify Neighborhood Change." No w24952. National Bureau of Economic Research. 2018.

Glass, Ruth. "Aspects of change." *The gentrification debates: A reader* (1964): 19-30.

Hipsman, Nathan. "Housing Price Changes within the City." Harvard University. 2017.

McKinnish, Terra, Randall Walsh, and T. Kirk White. "Who gentrifies low-income neighborhoods?." *Journal of urban economics* 67, no. 2 (2010): 180-193.

Meltzer Rachel, and Sean Capperis. Neighbourhood differences in retail turnover: Evidence from New York City. *Urban Studies* 54, vol. 13 (2017): 3022-3057.

Meltzer R, and Jenny Schuetz. Bodegas or Bagel Shops? Neighborhood Differences in Retail and Household Services. Economic Development Quarterly 26, no. 1 (2012): 73-94.

Newman, Kathe, and Elvin K. Wyly. "The right to stay put, revisited: Gentrification and resistance to displacement in New York City." *Urban studies* 43, no. 1 (2006): 23-57.

O'Sullivan, Arthur. "Gentrification and crime." *Journal of Urban Economics* 57, no. 1 (2005): 73-85.

Sullivan, Daniel Monroe. "Reassessing gentrification: Measuring residents' opinions using survey data." *Urban Affairs Review* 42, no. 4 (2007): 583-592.

Urban Development Project. *Rising Housing Costs and Re-Segregation in the San Francisco Bay Area*. Report. 2019.

Vigdor, Jacob L., "Does gentrification harm the poor?." *Brookings-Wharton papers on urban affairs* (2002): 133-182.

Vigdor, Jacob L. "Is urban decay bad? Is urban revitalization bad too?." *Journal of Urban Economics* 68, no. 3 (2010): 277-289.

Appendix A: Model Proofs

As shown in a footnote, the price of the luxury service is $A-b\frac{\sigma_G S}{H}+r\frac{\omega}{2}$. Consequently, the willingness to pay for the poor and rich to live in the neighborhood equal $\theta_0 r+\sigma_L B_L S+\frac{b}{2}\left(\frac{\sigma_G S}{H}-r\frac{\omega}{2b}\right)^2-\underline{U}$ and $\sigma_L B_L S+\frac{b}{2}\left(\frac{\sigma_G S}{H}-r\frac{\omega}{2b}+\frac{\omega}{2b}\right)^2-\underline{U}-\Delta$ and respectively. These must be equal if both groups live in the neighborhood and that implies that $\theta_0 r+\Delta=\frac{b}{2}\left(\frac{\omega}{2b}\right)^2+\frac{\omega}{2}\left(\frac{\sigma_G S}{H}-r\frac{\omega}{2b}\right)$ or $r=\frac{\omega^2-8b\Delta+4\omega b\sigma_G(\frac{S}{H})}{2\omega^2+8b\theta_0}$.

Equilibrium in the market for commercial real estate implies that profits from all three activities must be equal, which means that $H\epsilon(1-\epsilon)^{\frac{1-\epsilon}{\epsilon}}\vartheta^{\frac{1}{\epsilon}}k^{\frac{\epsilon-1}{\epsilon}}=\pi_L=\tau-\varphi\sigma_T=A-b\frac{\sigma_G S}{H}+r\frac{\omega}{2}-k_G$, which implies that $\sigma_T=\frac{\tau-\pi_L}{\varphi}$, and $\sigma_G=\frac{H}{bS}\left(A+r\frac{\omega}{2}-k_G-\pi_L\right)$. This structure implies that the price for the luxury good ($A-b\frac{\sigma_G S}{H}+r\frac{\omega}{2}$) is independent of the level of gentrification and consequently, the welfare from the luxury good is also independent of the both the gentrification parameter and the internet commerce parameter (τ). Firm profits, and consequently commercial rents, are therefore independent of both τ and Δ, because they are pinned down by π_L.

Combining equations yields:

(A1) $$r=\frac{\omega^2-8b\Delta+4\omega\left(A+r\frac{\omega}{2}-k_G-\pi_L\right)}{2\omega^2+8b\theta_0}=\frac{\omega^2-8b\Delta+4\omega(A-k_G-\pi_L)}{8b\theta_0}.$$

This is declining with Δ and independent of τ. This in turn implies

(A2) $$\sigma_G=\frac{(H/S)}{4b^2\theta_0}\left((A-k_G-\pi_L)(4b\theta_0+\omega^2)+.25\omega^3-2b\Delta\omega\right)$$

This is declining with Δ and independent of τ. Then number of idiosyncratic services in the city will equal

(A3) $$\sigma_L=1-\frac{\tau-\pi_L}{\varphi}-\frac{(H/S)}{4b^2\theta_0}\left((A-k_G-\pi_L)(4b\theta_0+\omega^2)+.25\omega^3-2b\Delta\omega\right).$$

This quantity is declining with τ and increasing with Δ.

The willingness to pay for the poor (and rich) to live in the neighborhood equals

(A4) $Rent = (1 - \sigma_G - \frac{\tau - \pi_L}{\varphi})\frac{\pi_L}{1-\epsilon}S + \frac{1}{2b}(A - k_G - \pi_L + .5\omega)^2 - \underline{U} - \Delta$..

The derivative of rents with respect to Δ equals $-\frac{\pi_L}{1-\epsilon}S\frac{d\sigma_G}{d\Delta} - 1$. Using $\frac{d\sigma_G}{d\Delta} = -\frac{(H/S)\omega}{2b\theta_0}$ and $\frac{dr}{d\Delta} = \frac{-1}{\theta_0}$, this can be written as $B_L\frac{H\omega}{2b\theta_0} - 1$. Consequently, an increased demand for the neighborhood by the rich (a decrease in Δ) will lead to an increase in the (1) the share of the population in the neighborhood that are rich, (2) increase in the share of stores that sell generic luxury services and have an equal (3) decrease in the number of stores that sell idiosyncratic local services, (4) have no impact on the share of stores that sell goods, and (5) raise prices if and only if $2b\theta_0 > \omega HB_L$.

An improvement in telecommuting (a decrease in τ) will (1) have no impact on the share of the population that is rich, (2) have no impact on the number of luxury stores, and (3) increase the share of stores that sell idiosyncratic local services, (4) reduce the number of stores that sell goods and (5) always increase prices.

For welfare calculations, firm profits and commercial rents are both independent of both τ and Δ.

The welfare of the rich always equals $\underline{Y} + \underline{U} + \Delta$, which is obviously increasing with Δ and independent of τ.

The average welfare of the poor equals $\underline{Y} + \underline{U} + .5\theta_0(1-r)$. Consequently, if the rich enter then the poor will be worse off, and so consequently their welfare is increasing with Δ and independent of τ.

If we consider the incremental welfare of the inframarginal poor and the wealth of property owners, this will equal H times $.5\theta_0(1-r)(1-r) + \sigma_L B_L S + \frac{b}{2}\left(\frac{\sigma_G S}{H} - r\frac{\omega}{2b} + \frac{\omega}{2b}\right)^2 - \underline{U} - \Delta$.

The derivative of this with respect to τ is strictly negative, although this works only through the price effect.

The derivative of this with respect to Δ is $B_L\frac{H\omega}{2b\theta_0} - 1 + (1-r)$, or $B_L\frac{H\omega}{2b\theta_0} - r$ which will be positive if and only if $2rb\theta_0 < \omega HB_L$.

Appendix B: Robustness Checks

In this appendix we provide robustness checks on our analysis in section V.

Appendix C: Establishment Growth and Continuous Gentrification Measures

In this appendix, we turn to our results on overall business activity using County Business Patterns. As a reminder, we cannot exactly replicate the closure analysis or the within-storefront establishment turnover analysis with the CBP data because it does not allow us to track establishments over time. However, the CBP data allows us to check that the two datasets are not telling radically different stories about the nature of retail change.

In Table C.1, we look at the change in the number of establishments in these categories in a regression format. We define the dependent variable as the change in establishments between 2012 and 2017 divided by the number of establishments in 2012. We show results for all establishments in the first two regressions. In the next two regressions, we look only at tradables-selling establishments. In the last two regressions, we focus on nontradables-selling establishments.

The only meaningfully significant coefficients are associated with rent increases in the fifth regression and the share of young people in the sixth regression. Both of these variables predict increases in the number of nontradables-selling establishments. The interaction between rent growth and initial poverty levels, which is meant to capture gentrification, is everywhere insignificant and small in magnitude.

Table C.2 uses distance to the city center as an instrument for price increases as before. The coefficients become slightly larger in magnitude, but they remain insignificant for all establishments and for experience selling establishments. There is a slightly negative impact of gentrification on the number of goods selling establishments. In general, however, our gentrification measures again show little correlation with the number of establishments in our data. Perhaps our time period is just too short. Perhaps we are missing more granular changes that are missed with the County Business Patterns data. Yet overall, the weight of the data suggests that urban change is a constant, but that gentrification is primarily about the cost of living, not about the retail landscape.

Figure 1: Discrete definitions of rich, gentrifying, and non-gentrifying zip codes

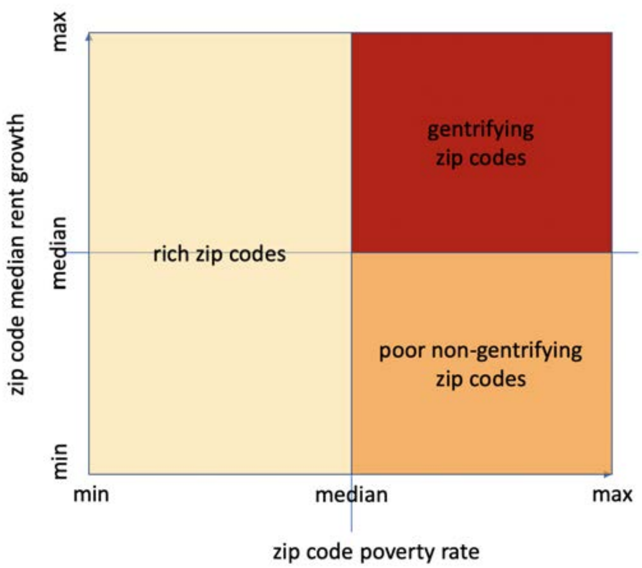

Figure 2: Maps of gentrification as implied by rent growth.

Top: Chicago, Los Angeles, and New York City. Bottom: Boston and San Francisco.

For each city, the maps on the left and right show areas that are gentrifying according to the rent-based and education-based definitions, respectively. Dark red indicates gentrifying areas, orange denotes poor non-gentrifying areas, and light beige denotes rich areas. The center of the black circle denotes the location of city hall.

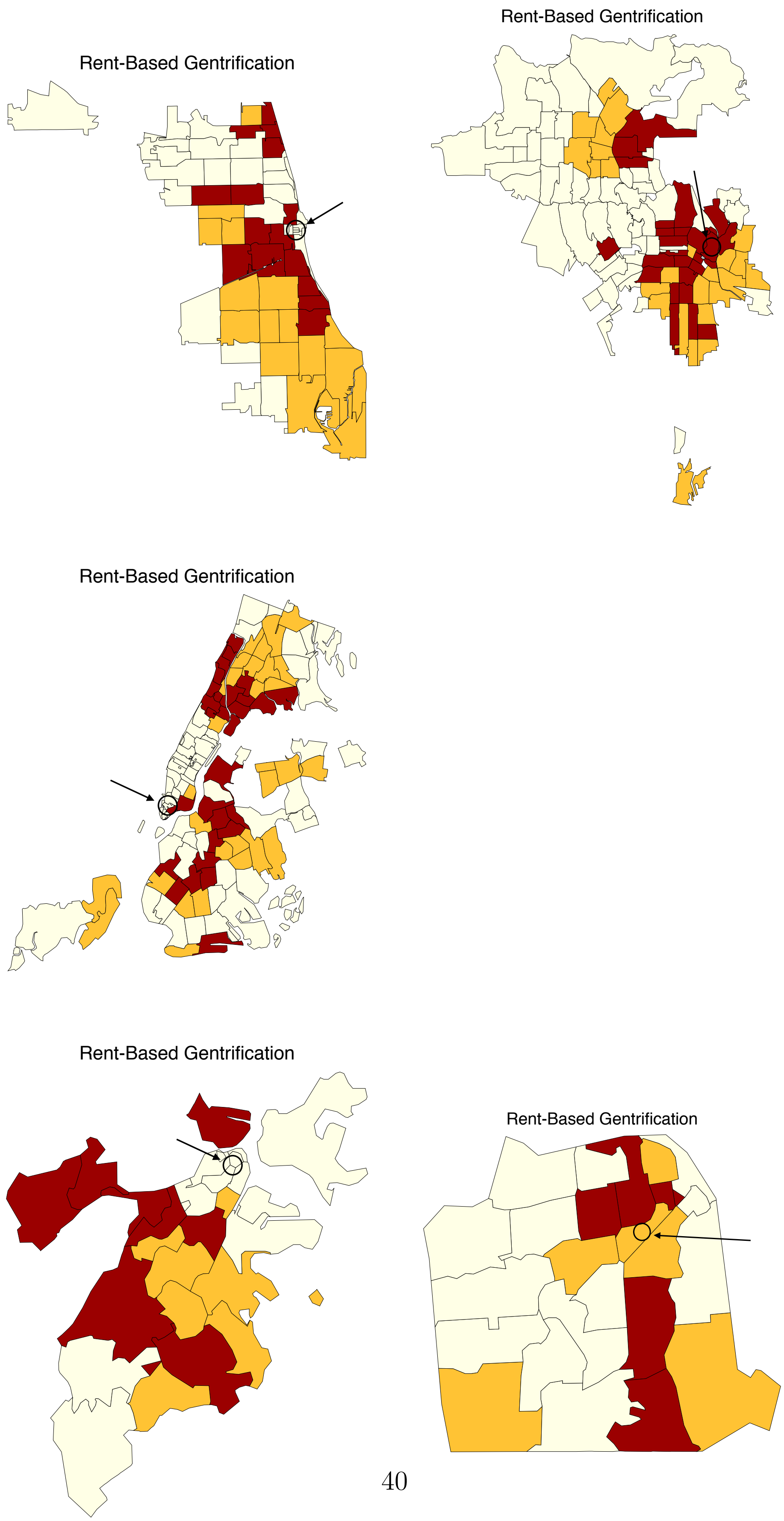

Figure 3: Correlation between Yelp and CBP establishment counts

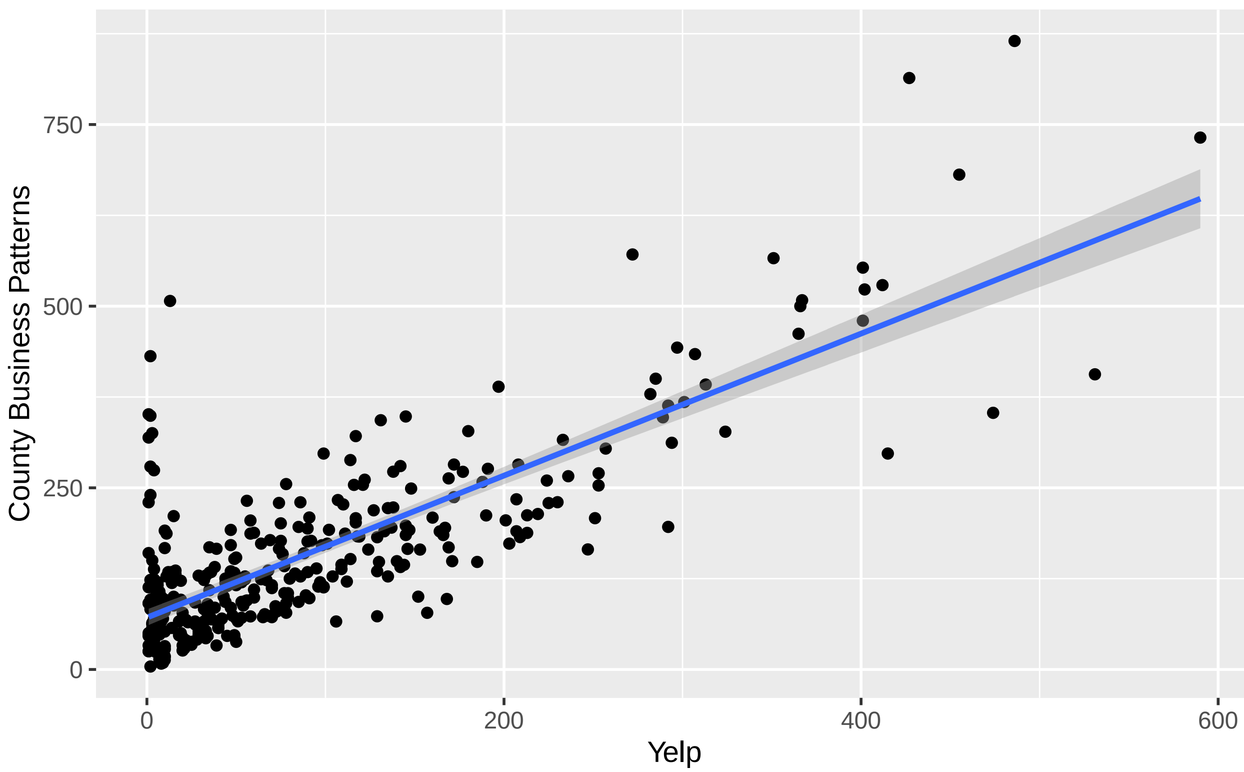

Table 1: Summary Statistics for Rent-Based Gentrification

	Chicago		Los Angeles		New York		Boston		San Francisco	
	gentrify	poor	gentrify	poor	gentrify	poor	gentrify	poor	gentrify	poor
Demographic information (2012 level and percentage change 2012-2017)										
share college	21.72	9.51 ***	15.83	10.09 ***	16.39	13.86	29.61	15.21 ***	30.19	27.83
change (p.p.)	1.82	0.73 *	2.89	1.02 ***	2.82	1.29 **	0.6	2.93	0.75	3.07
share 25-34	20.87	12.97 ***	18.72	16 ***	18.57	15.71 ***	24.36	16.09 **	18.93	22.18
change (p.p.)	0.66	1.07	0.85	-0.14 *	1.52	0.95	2.85	3	5.68	0.87
share white	43.94	21.73 **	39.73	45.93	34.07	30.32	62.83	28.17 ***	50.86	43.92
change (p.p.)	1.78	1.03	1.02	-1.8	-0.67	-1.98	-2.37	-1.34	-6.36	-1.85
median income	43180.13	33831 **	36064.77	38064.2	36858.3	37269.67	50507.62	38889	55353.33	53196
% change	11.06	2.11 ***	15.15	6.61 ***	20.38	6.65 ***	29.45	6.01 **	46.88	29.38
Housing market variables										
miles to city hall	5.14	8.77 ***	5.51	9.36 **	6.66	8.42 **	3.34	3.51	1.77	2.25
median rent	922.27	870.53	1005.27	992.52	997.9	1055.6	1264.12	1105.14	1112.5	1206
% change	13.13	1.32 ***	17.73	8.01 ***	21.02	11.98 ***	17.17	6.31 ***	30.33	13.99 **
share single-family	11.24	35.33 ***	28.4	43.72 ***	2.2	5.54 **	6.59	9.44	11.43	13.2
share 0-2 bedrooms	68.67	50.33 ***	77.85	66.38 ***	76.81	72.83 *	75.52	59.6 *	78.98	75.45
share public transit	29.56	26.88	18.37	13.72 *	64.96	61.93	33.67	35.62	31.18	33.32

Table 2: Mapping of Yelp categories to NAICS codes

Yelp category	NAICS codes	NAICS descriptions
Bars	722410, 312120, 312140, 312130, 4453, 44531, 7224, 72241	Drinking Places (Alcoholic Beverages), Breweries, Distilleries, Wineries, Beer, Wine, and Liquor Stores, Beer, Wine, and Liquor Stores, Drinking Places (Alcoholic Beverages), Drinking Places (Alcoholic Beverages)
Cafe	311811, 722515	Retail Bakeries, Snack and Nonalcoholic Beverage Bars
Convenience	445120	Convenience Stores
Fast food	722330, 722513, 722514	Mobile Food Services, Limited-Service Restaurants, Cafeterias, Grill Buffets, and Buffets
Grocery	445291, 445292, 445110, 445230, 445210, 445220, 4451, 44511, 44512, 4452, 44521, 44522, 44523, 44529, 445299	Baked Goods Stores, Confectionery and Nut Stores, Supermarkets and Other Grocery (except Convenience) Stores, Fruit and Vegetable Markets, Meat Markets, Fish and Seafood Markets, Grocery Stores, Supermarkets and Other Grocery (except Convenience) Stores, Convenience Stores, Specialty Food Stores, Meat Markets, Fish and Seafood Markets, Fruit and Vegetable Markets, Other Specialty Food Stores, All Other Specialty Food Stores
Hair	812111, 812112	Barber Shops, Beauty Salons
Restaurant	7225, 72251, 722511	Restaurants and Other Eating Places, Restaurants and Other Eating Places, Full-Service Restaurants

Table 3a: Correlation between Yelp and CBP establishment counts for gentrifying zip codes

	2013			2017		
category	avg # Yelp estabs	corr	mean Yelp/CBP ratio	avg # Yelp estabs	corr	mean Yelp/CBP ratio
all	106.45	0.82	0.63	143.66	0.83	0.80
alcohol	2.88	0.66	0.74	4.03	0.69	0.53
cafe	9.03	0.80	0.71	12.96	0.76	0.75
convenience	2.24	0.06	0.96	2.71	0.33	0.49
fastfood	12.41	0.72	0.30	18.28	0.82	0.38
grocery	11.04	0.72	0.38	12.99	0.74	0.54
hair	9.00	0.48	1.04	11.42	0.39	1.03
restaurant	48.60	0.91	1.33	63.95	0.90	1.42

Table 3b: Correlation between Yelp and CBP establishment counts for poor non-gentrifying zip codes

	2013			2017		
category	avg # Yelp estabs	corr	mean Yelp/CBP ratio	avg # Yelp estabs	corr	mean Yelp/CBP ratio
all	63.52	0.59	0.47	88.78	0.56	0.66
alcohol	2.34	0.60	0.69	3.61	0.63	0.61
cafe	6.05	0.69	0.70	8.70	0.67	0.75
convenience	2.30	0.03	1.16	2.75	0.20	0.54
fastfood	9.32	0.52	0.26	14.10	0.54	0.36
grocery	11.55	0.56	0.38	13.37	0.55	0.61
hair	5.12	0.13	1.01	6.59	0.20	0.67
restaurant	27.35	0.87	1.42	37.91	0.81	1.53

Table 3c: Correlation between Yelp and CBP establishment counts for rich zip codes

	2013			2017		
category	avg # Yelp estabs	corr	mean Yelp/CBP ratio	avg # Yelp estabs	corr	mean Yelp/CBP ratio
all	100.69	0.85	0.52	137.51	0.84	0.70
alcohol	3.38	0.49	0.44	4.84	0.54	0.46
cafe	11.98	0.77	0.74	17.51	0.77	0.77
convenience	2.37	0.44	0.64	2.87	0.46	0.51
fastfood	13.80	0.83	0.35	19.48	0.84	0.44
grocery	6.74	0.73	0.37	8.36	0.70	0.56
hair	9.59	0.56	0.66	12.85	0.63	0.69
restaurant	51.21	0.92	0.87	66.39	0.89	1.01

Table 4a: Percentage change in number of establishments, big cities

This table presents the percentage change in the number of establishments in different retail categories in gentrifying, poor non-gentrifying, and non-poor areas . Stars in the poor non-gentrifying and rich columns indicate that the average outcome in gentrifying neighborhoods is statistically significantly different from the average outcome in poor non-gentrifying or rich neighborhoods, at the 10% (1 star), 5% (2 stars), or 1% (3 stars) level.

	Chicago			Los Angeles			New York		
	gentrify	poor	rich	gentrify	poor	rich	gentrify	poor	rich
all	6.94	-11.15 ***	6.18 *	5.35	3.68	7.8	15.64	11.19	1.88 ***
bar	4.4	-28.39 **	0.35	-0.51	2.46 **	-6.68	7.99	-1.63 **	6.94 **
cafe	28.66	11.81 **	20.67	19.75	13.48	16.12	26.2	12.29 *	21.9 ***
convenience	-8.86	-14.29	-0.82	-12.59	-6.12	21.71	23.88	21.05	16.3 **
fastfood	18.9	7.03	10.7	11.4	12.52	13.88	18.16	12.63	8.8 **
grocery	-15.89	-24.7 *	-23.97 *	-25.62	-19.59	-23.46	-2.77	2.69 **	-16.2 ***
hair	23.11	-1.67 *	-0.24	8.44	40	7.04 *	17.96	22.71	5.13
restaurant	33.71	13.36	16.24 ***	19.38	18.15	10.16	38.08	21.87	16.44 ***

Table 4b: Percentage change in number of establishments, small cities

	Boston			San Francisco		
	gentrify	poor	rich	gentrify	poor	rich
all	1.54	6.46	3.38	6.16	4.98	7.65
bar	-21.69	-50	-20.16	16.57	-1.44	27.93
cafe	0.95	48.02	9.46	31.55	29.2	21.29
convenience	14.46	58.52	15.97	80	NA	NA
fastfood	5.43	21.02 *	2.24	9.8	24.79	23.37
grocery	-19.59	-26.5	-24.19	-10.74	-24.76	-16.62
hair	8.12	-30.32 *	-2.15	35.36	9.06	9.29 *
restaurant	18.2	18.73	26.82	3.22	-1.46	5.36

Table 4c: Percentage change in number of establishments 2006-2011, big cities

This table presents the percentage change in the number of establishments in different retail categories in gentrifying, poor non-gentrifying, and non-poor areas . Stars in the poor non-gentrifying and rich columns indicate that the average outcome in gentrifying neighborhoods is statistically significantly different from the average outcome in poor non-gentrifying or rich neighborhoods, at the 10% (1 star), 5% (2 stars), or 1% (3 stars) level.

	Chicago			Los Angeles			New York		
	gentrify	poor	rich	gentrify	poor	rich	gentrify	poor	rich
all	6.95	4.09	27.93	6.54	-2.4	7.55	37.28	37.57	82.96
bar	-5.63	-24.49 **	2.34	-0.14	-13.73	-16.35	6.04	-5.61	-2.69
cafe	9.33	-14.2	7.17 *	13.65	9.24	20.01	19.6	12.21	31.72
convenience	-11.68	1.11	-8.63	1.88	-16.67	-20.87 **	14.97	23.07	3.69
fastfood	1.46	2.58	11.95	6.28	2.42	-0.96	39.55	43.67	26.67
grocery	-2.93	11.54 *	1.92	-7.35	-16.23	-18.92	27.44	30.88	2.55
hair	19.18	1.67 **	11.39	6.09	-12.67	-3.55	5.59	14.54	10.26
restaurant	-0.32	-2.23	3.38	6.77	-11.01 ***	3.93	31.43	42.73	22.06

Table 4d: Percentage change in number of establishments 2006-2011, small cities

	Boston			San Francisco		
	gentrify	poor	rich	gentrify	poor	rich
all	10.71	11.64	15.77	15.04	13.68	11.68
bar	-37.09	-30	-11.09	-17.24	26.26 **	0.59
cafe	41.28	-6.84 *	14.31	34.68	18.38	23.56
convenience	-3.24	54.17 *	10.27	NA	-60	0
fastfood	11.06	-9.98 *	23.05	12.74	8.42	-1.17
grocery	7.59	1.03	-3.13	-1.96	-9.88	-0.31
hair	15.59	35.54 **	-30.8	10.94	7.65	-6.95 **
restaurant	20.37	3.68	14.37	21.33	20.85	12.31

Table 5a: Closure Rates by Retail Category (%), big cities

This table presents the closure rate of Yelp establishments in different retail categories in gentrifying, poor non-gentrifying, and non-poor areas . Stars in the poor non-gentrifying and rich columns indicate that the average outcome in gentrifying neighborhoods is statistically significantly different from the average outcome in poor non-gentrifying or rich neighborhoods, at the 10% (1 star), 5% (2 stars), or 1% (3 stars) level.

	Chicago			Los Angeles			New York		
	gentrify	poor	rich	gentrify	poor	rich	gentrify	poor	rich
all	20	8 ***	22 ***	19	10 ***	21 ***	18	15 ***	24 ***
bar	7	0	14	10	0	8	10	0 **	12
cafe	23	4 **	21	18	7 *	22	28	24	24
convenience	4	0	8	23	0	7	3	2	7
deli	26	8 **	28	31	24	30	16	8 **	21 **
dessert	22	5 *	26	17	11	24 *	19	17	29 ***
fastfood	28	9 ***	22 *	23	13 **	22	24	23	27
grocery	10	3 **	9	9	2 ***	17 *	3	2	9 ***
hair	16	3 **	19	13	2 **	10	10	5 *	17 **
restaurant	21	13 ***	23	20	14 ***	23	24	20 *	25

Table 5b: Closure Rates by Retail Category (%), small cities

	Boston			San Francisco		
	gentrify	poor	rich	gentrify	poor	rich
all	22	20	20	20	19	20
bar	0	NA	8	6	0 *	9
cafe	22	22	11	24	21	19
convenience	0	0	5	9	11	18
deli	28	30	27	22	28	28
dessert	31	0	26	29	19 *	22
fastfood	16	44	26	22	20	15
grocery	16	14	9	8	10	13
hair	13	22	22	16	13	14
restaurant	25	18	20	22	21	21

Table 6: Probit model of store closure

A probit model estimating the effect of gentrification on Yelp store closures. Standard errors are clustered at the Zip Code level.

	Dependent variable:		
	Indicator for closure		
	All	tradables	nontradables
	(1)	(2)	(3)
poverty rate	0.003	0.009	0.002
	(0.002)	(0.008)	(0.003)
rent growth (%)	−0.077	−1.289**	0.030
	(0.199)	(0.625)	(0.195)
poverty rate × rent growth	0.019*	0.070*	0.015
	(0.010)	(0.042)	(0.013)
log(category density)	−0.013	0.059	−0.016
	(0.014)	(0.052)	(0.027)
log(establishment density)	0.011	−0.030	0.012
	(0.020)	(0.061)	(0.033)
$$	0.041	−0.026	0.048*
	(0.025)	(0.072)	(0.025)
$$$	0.036	0.200	0.036
	(0.052)	(0.126)	(0.055)
$$$$	−0.071	−0.393	−0.050
	(0.046)	(0.316)	(0.079)
log(share college)	0.008***	0.012*	0.008***
	(0.002)	(0.006)	(0.003)
log(share 25 to 34)	0.001	0.001	0.0004
	(0.002)	(0.007)	(0.003)
log(share white)	0.0004	0.001	0.0003
	(0.001)	(0.002)	(0.001)
log(median income)	−0.012	0.173	−0.056
	(0.051)	(0.169)	(0.071)
constant	−1.257***	−0.803**	−0.762***
	(0.110)	(0.374)	(0.137)
City FEs?	Yes	Yes	Yes
Category FEs?	Yes	Yes	Yes
Observations	19,567	2,270	16,635

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

Table 7: Linear and Linear IV models of store closure
ootnotesizeAdditional specifications to test the robustness of store closure patterns in Table 4. We include the same controls, but suppress the estimated coefficients here for brevity. Standard errors are again clustered at the Zip Code level.

	Dependent variable:					
	Indicator for closure					
	OLS			*instrumental variable*		
	all	tradables	nontradables	all	tradables	nontradables
	(1)	(2)	(3)	(4)	(5)	(6)
poverty rate	0.001	0.003*	0.001	−0.0004	−0.0001	−0.0002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)
rent growth (%)	−0.030	−0.405***	0.007	0.089	−0.166	0.093
	(0.068)	(0.089)	(0.090)	(0.146)	(0.393)	(0.147)
poverty rate × rent growth	0.006*	0.023*	0.005	0.019**	0.045**	0.013
	(0.003)	(0.012)	(0.004)	(0.009)	(0.019)	(0.009)
constant	0.083**	0.230	0.224***	0.095*	0.269**	0.231***
	(0.034)	(0.153)	(0.038)	(0.053)	(0.112)	(0.048)
Observations	19,567	2,270	16,635	19,567	2,270	16,635
R^2	0.013	0.035	0.008	0.012	0.031	0.007
Adjusted R^2	0.011	0.027	0.007	0.010	0.023	0.006
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 8: Store closure model with discrete measure of gentrification

This table regresses an indicator for store closure on a gentrification dummy, instead of our continuous measure of gentrification, and uses as a sample only the establishments in zip codes that are eligbile to gentrify according to our definition. In the OLS and IV columns, we control for the variables specified in the notes. For the matching estimator, we match exactly on the categorical controls and fuzzy-match on the continuous controls.

	Indicator for store closure		
	OLS	*instrumental variable*	*matching estimator*
	(1)	(2)	(3)
gentrify	0.023**	0.067**	0.031
	(0.010)	(0.030)	(0.018)
Constant	0.023	−0.044	0.297***
	(0.045)	(0.047)	(0.009)
Observations	9,144	9,144	2,508
R^2	0.027	0.023	0.002
Adjusted R^2	0.025	0.021	

Notes:
Categorical controls: city, category, price point
Continuous controls: poverty rate (2012), distance to city hall, rating, review count

Table 9a: Category transition matrix, all zip codes

This table examines storefronts where we observe one store close and another store open. The second column shows the number of establishments that started out as bars, cafes, convenience stores, etc. Columns 3-12 show the probability that a given storefront transitions from being occupied by one type of establishment to another type of establishment. Column 13 ("none") captures storefronts where establishments close but no new establishment enters the Yelp database before the end of 2018.

		to										
from	establishments	bar	cafe	convenience	deli	dessert	fastfood	grocery	hair	other	restaurant	none
bar	277	0.23	0.08	0.02	0.07	0.06	0.11	0.03	0.03	0.06	0.27	0.03
cafe	1694	0.01	0.21	0.01	0.08	0.12	0.11	0.02	0.02	0.05	0.33	0.04
convenience	136	0.02	0.15	0.09	0.15	0.06	0.09	0.08	0.05	0.01	0.26	0.03
deli	1154	0.01	0.10	0.00	0.16	0.06	0.14	0.03	0.02	0.03	0.39	0.04
dessert	1462	0.01	0.11	0.01	0.04	0.23	0.13	0.03	0.03	0.09	0.29	0.03
fastfood	2664	0.01	0.07	0.00	0.06	0.07	0.31	0.02	0.02	0.05	0.37	0.03
grocery	823	0.02	0.10	0.01	0.09	0.07	0.10	0.21	0.03	0.06	0.27	0.02
hair	1050	0.01	0.08	0.00	0.03	0.04	0.07	0.02	0.52	0.04	0.18	0.02
other	1049	0.02	0.09	0.00	0.05	0.12	0.14	0.03	0.03	0.23	0.28	0.01
restaurant	8618	0.01	0.06	0.00	0.05	0.04	0.12	0.01	0.01	0.03	0.62	0.04

Table 9b: Category transition matrix, gentrifying zip codes

		to										
from	establishments	bar	cafe	convenience	deli	dessert	fastfood	grocery	hair	other	restaurant	none
bar	74	0.19	0.11	0.00	0.07	0.00	0.12	0.03	0.05	0.08	0.31	0.04
cafe	464	0.02	0.23	0.01	0.09	0.10	0.12	0.02	0.02	0.05	0.32	0.03
convenience	35	0.00	0.11	0.11	0.17	0.06	0.09	0.06	0.06	0.00	0.26	0.09
deli	291	0.01	0.11	0.00	0.15	0.05	0.14	0.05	0.03	0.03	0.37	0.05
dessert	356	0.02	0.11	0.00	0.03	0.25	0.11	0.04	0.03	0.10	0.27	0.05
fastfood	754	0.01	0.06	0.00	0.05	0.06	0.32	0.02	0.02	0.04	0.38	0.04
grocery	303	0.02	0.10	0.02	0.11	0.07	0.10	0.20	0.04	0.06	0.28	0.01
hair	298	0.00	0.07	0.00	0.03	0.06	0.05	0.03	0.53	0.02	0.20	0.01
other	301	0.02	0.10	0.00	0.05	0.11	0.14	0.02	0.02	0.29	0.24	0.00
restaurant	2438	0.01	0.06	0.00	0.05	0.03	0.12	0.01	0.01	0.03	0.64	0.04

Table 9c: Category transition matrix, poor non-gentrifying zip codes

		to										
from	establishments	bar	cafe	convenience	deli	dessert	fastfood	grocery	hair	other	restaurant	none
bar	47	0.17	0.04	0.02	0.04	0.04	0.19	0.04	0.04	0.06	0.34	0.00
cafe	218	0.00	0.22	0.00	0.04	0.16	0.12	0.02	0.00	0.05	0.35	0.04
convenience	24	0.00	0.17	0.04	0.17	0.04	0.00	0.04	0.12	0.00	0.38	0.04
deli	148	0.01	0.07	0.00	0.16	0.05	0.21	0.02	0.01	0.01	0.43	0.03
dessert	250	0.00	0.10	0.00	0.03	0.26	0.12	0.03	0.03	0.09	0.30	0.03
fastfood	460	0.01	0.06	0.01	0.04	0.07	0.34	0.02	0.01	0.06	0.35	0.03
grocery	181	0.02	0.08	0.02	0.08	0.06	0.10	0.29	0.03	0.04	0.26	0.02
hair	111	0.00	0.08	0.00	0.02	0.07	0.11	0.04	0.40	0.05	0.21	0.03
other	165	0.02	0.06	0.01	0.05	0.16	0.15	0.06	0.02	0.19	0.26	0.01
restaurant	1244	0.01	0.05	0.00	0.03	0.05	0.12	0.02	0.01	0.04	0.63	0.04

Table 10: Within-storefront retail type change

This table estimates the impact of being located in a gentrifying area on the probability that a given storefront transitions to an nontradables provider (conditional on its previous tenant closing). In the third column, we examine whether tradables producers in gentrifying areas are more likely to become nontradables providers. Standard errors are clustered at the Zip Code level.

	Dependent variable:		
	Transition to nontradables provision		
	(1)	(2)	(3)
gentrifying	−0.001	−0.002	−0.002
	(0.012)	(0.012)	(0.013)
tradables producer		−0.198***	−0.200***
		(0.018)	(0.028)
gentrifying × tradables producer			0.004
			(0.036)
Constant	0.838***	0.863***	0.863***
	(0.008)	(0.008)	(0.009)
Observations	8,162	8,162	8,162
R^2	0.00000	0.030	0.030
Adjusted R^2	−0.0001	0.030	0.030
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 11a: Price point transition matrix, all zip codes

This table examines storefronts where we observe one store close and another store open. The second column shows the number of storefronts that started out as 1 dollar sign, 2 dollar sign, etc. establishments. Columns 3-6 show the probability that a storefront occupied by an establishment in one price bucket transitions to an establishment in another price bucket.

		to			
from	establishments	1	2	3	4
1	2367	0.48	0.49	0.02	0.01
2	3447	0.21	0.71	0.06	0.01
3	456	0.13	0.60	0.21	0.05
4	79	0.18	0.54	0.18	0.10

Table 11b: Price point transition matrix, gentrifying zip codes

		to			
from	establishments	1	2	3	4
1	737	0.52	0.45	0.02	0.01
2	901	0.22	0.71	0.05	0.01
3	106	0.17	0.60	0.19	0.04
4	17	0.18	0.35	0.29	0.18

Table 11c: Price point transition matrix, poor non-gentrifying zip codes

		to			
from	establishments	1	2	3	4
1	308	0.56	0.41	0.02	0.01
2	321	0.28	0.64	0.06	0.01
3	32	0.19	0.56	0.16	0.09
4	9	0.33	0.67	0.00	0.00

Table 11d: Testing whether transition matrices for gentrifying and poor non-gentrifying neighborhoods are different

This table estimates an ordered logit regression, where each observation is a storefront that changes establishments and the outcome variable is price point of the new establishment. Standard errors are clustered at the Zip Code level.

	Dependent variable:		
	New establishment price point		
	(1)	(2)	(3)
gentrify	0.200	0.341**	0.268
	(0.159)	(0.166)	(0.205)
initially $$	1.222***	1.213***	1.138***
	(0.091)	(0.091)	(0.171)
initially $$$	2.141***	2.124***	2.274***
	(0.244)	(0.236)	(0.561)
initially $$$$	2.327***	2.280***	0.819
	(0.778)	(0.760)	(0.738)
gentrify × initially $$			0.108
			(0.207)
gentrify × initially $$$			−0.172
			(0.627)
gentrify × initially $$$$			2.348**
			(1.037)
City FEs?	No	Yes	Yes
Observations	2,431	2,431	2,431
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 12: Do storefronts with low pricepoint establishments "trade up" when neighborhoods gentrify?
This table estimates the impact of being in an area where rent is rising on the probability that a given storefront transitions to a higher price bracket (conditional on its previous tenant closing). In columns 4-6, we examine whether establishments in the lowest price bracket in gentrifying areas are more likely to move into a higher price bracket. Standard errors are clustered at the Zip Code level.

	Dependent variable:					
	Indicator for increase in number of dollar signs					
	all	tradables	nontradables	all	tradables	nontradables
	(1)	(2)	(3)	(4)	(5)	(6)
rent growth (%)	0.006	0.305	0.001	−0.017	0.092	−0.043
	(0.091)	(0.423)	(0.082)	(0.067)	(0.258)	(0.062)
low pricepoint				0.345***	0.354***	0.326***
				(0.045)	(0.125)	(0.048)
rent growth × low pricepoint				0.358*	0.408	0.435**
				(0.188)	(0.513)	(0.207)
Constant	0.235***	0.211***	0.234***	0.066***	0.053	0.071***
	(0.020)	(0.081)	(0.019)	(0.015)	(0.052)	(0.016)
Observations	2,513	239	2,161	2,513	239	2,161
R^2	0.00000	0.003	0.00000	0.225	0.239	0.219
Adjusted R^2	−0.0004	−0.001	−0.0005	0.224	0.229	0.218
Note:				*p<0.1; **p<0.05; ***p<0.01		

Table 13: Likelihood of transitioning to a higher-priced establishment

This table estimates a probit model of the impact of gentrification on the probability that a given storefront transitions to a higher price bracket (conditional on its previous tenant closing), using the same continuous measure of gentrification as in Table 5. Standard errors are clustered at the Zip Code level.

	All	initially tradables	initially nontradables	Within-category	tradables to tradables	nontradables to nontradables
	(1)	(2)	(3)	(4)	(5)	(6)
poverty rate	-0.009^{***}	-0.004	-0.010^{***}	-0.013^{***}	0.030	-0.013^{***}
	(0.001)	(0.013)	(0.002)	(0.001)	(0.020)	(0.001)
rent growth (%)	-0.056	-0.111	-0.050	-0.403	1.491	-0.419
	(0.216)	(1.015)	(0.287)	(0.501)	(4.726)	(0.546)
poverty rate × rent growth	0.045^{***}	0.059	0.043^{***}	0.067^{***}	-0.025	0.072^{***}
	(0.005)	(0.068)	(0.009)	(0.015)	(0.158)	(0.019)
log(category density)	-0.017	-0.044	-0.011	0.075^{***}	-0.310	0.085^{***}
	(0.036)	(0.221)	(0.040)	(0.028)	(0.391)	(0.028)
log(establishment density)	0.098^{***}	0.144	0.084^{**}	-0.013	1.122^{**}	-0.041
	(0.034)	(0.217)	(0.038)	(0.042)	(0.531)	(0.037)
$$	-1.573^{***}	-1.713^{***}	-1.546^{***}	-1.720^{***}	-2.362^{*}	-1.709^{***}
	(0.060)	(0.224)	(0.078)	(0.038)	(1.211)	(0.042)
$$$	-1.902^{***}	-5.432^{***}	-1.855^{***}	-1.901^{***}	-6.277^{***}	-1.885^{***}
	(0.096)	(0.233)	(0.068)	(0.091)	(0.235)	(0.088)
$$$$	-5.393^{***}	-5.498^{***}	-5.395^{***}	-5.623^{***}		-5.607^{***}
	(0.178)	(0.197)	(0.225)	(0.190)		(0.217)
constant	-0.750^{***}	-1.660	-0.489^{***}	-0.039	-9.933^{***}	-0.725^{***}
	(0.181)	(1.152)	(0.052)	(0.199)	(1.964)	(0.055)
City FEs?	Yes	Yes	Yes	Yes	Yes	Yes
Category FEs?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,506	511	4,823	2,645	81	2,546

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

B Robustness Checks

Table B.1: Probit model of store closure, large cities only
A probit model estimating the effect of gentrification on Yelp store closures, for our large cities only. We control for the density of all retail in the zip code from the County Business Patterns, and the density of establishments in the same category as the store in question on Yelp. In the even-numbered columns, we add additional demographic controls. We include city and category fixed effects, and standard errors are clustered at the Zip Code level.

	Dependent variable:		
	Indicator for store closure		
	all	tradables	nontradables
	(1)	(2)	(3)
poverty rate	0.004	0.016***	0.003
	(0.003)	(0.005)	(0.003)
rent growth (%)	−0.074	−1.510***	0.030
	(0.239)	(0.437)	(0.314)
poverty rate × rent growth	0.029***	0.120***	0.020
	(0.011)	(0.028)	(0.014)
log(category density)	−0.022	0.031	−0.014
	(0.017)	(0.066)	(0.022)
log(establishment density)	0.020	−0.012	0.011
	(0.022)	(0.120)	(0.024)
$$	0.055**	−0.043	0.070**
	(0.027)	(0.055)	(0.032)
$$$	0.053	0.257*	0.054
	(0.056)	(0.145)	(0.058)
$$$$	−0.101	−0.559	−0.072**
	(0.061)	(0.380)	(0.036)
log(share college)	0.013***	0.016***	0.012***
	(0.002)	(0.005)	(0.002)
log(share 25 to 34)	−0.003	−0.005	−0.003
	(0.003)	(0.003)	(0.004)
log(share white)	0.001	0.001	0.001
	(0.001)	(0.004)	(0.001)
log(median income)	−0.074	0.312**	−0.133*
	(0.079)	(0.129)	(0.068)
constant	−1.101***	−0.874*	−0.677***
	(0.115)	(0.504)	(0.089)
City FEs?	Yes	Yes	Yes
Category FEs?	Yes	Yes	Yes
Observations	14,929	1,661	12,782
Note:			*p<0.1; **p<0.05; ***p<0.01

Table B.2: Linear and Linear IV models of store closure, large cities only

An instrumental variables model estimating the effect of gentrification on Yelp store closures, for large cities only. We control for the density of all retail in the zip code from the County Business Patterns, and the density of establishments in the same category as the store in question on Yelp. In the even-numbered columns, we add additional demographic controls. We include city and category fixed effects, and standard errors are clustered at the Zip Code level.

	Dependent variable:		
	Indicator for store closure		
	all	tradables	nontradables
	(1)	(2)	(3)
poverty rate	0.001	0.003	0.001
	(0.001)	(0.003)	(0.002)
rent growth (%)	−0.115	−0.340	−0.153
	(0.198)	(0.488)	(0.237)
poverty rate × rent growth	0.025**	0.060***	0.020*
	(0.012)	(0.022)	(0.012)
log(category density)	−0.007	0.017	−0.005
	(0.008)	(0.019)	(0.010)
log(establishment density)	0.006	−0.011	0.004
	(0.011)	(0.023)	(0.013)
$$	0.019**	−0.013	0.024**
	(0.009)	(0.027)	(0.010)
$$$	0.018	0.082*	0.019
	(0.021)	(0.049)	(0.022)
$$$$	−0.033	−0.127*	−0.022
	(0.028)	(0.067)	(0.031)
log(share college)	0.004***	0.005*	0.004***
	(0.001)	(0.002)	(0.001)
log(share 25 to 34)	−0.0002	−0.002	−0.0001
	(0.001)	(0.003)	(0.002)
log(share white)	0.0002	−0.00005	0.0003
	(0.0003)	(0.001)	(0.0003)
log(median income)	−0.017	0.108	−0.038
	(0.032)	(0.073)	(0.036)
constant	0.122**	0.234**	0.246***
	(0.057)	(0.103)	(0.047)
City FEs?	Yes	Yes	Yes
Category FEs?	Yes	Yes	Yes
Observations	14,929	1,661	12,782
R^2	0.013	0.043	0.009
Adjusted R^2	0.011	0.034	0.007
Note:			*p<0.1; **p<0.05; ***p<0.01

Table B.3: Balance table for matching estimator

	(1)		(2)		(3)	
	mean	sd	mean	sd	b	t
Yelp rating	3.60	0.73	3.59	0.74	-0.00	(-0.16)
Number of Yelp reviews	118.77	238.72	117.19	236.11	-1.58	(-0.17)
Distance to city hall	3.11	2.67	3.18	2.67	0.07	(0.65)
Delivers	0.17	0.37	0.15	0.36	-0.02	(-1.37)
Avg order revenue	23.21	7.18	24.14	9.04	0.93	(1.21)
Avg menu item price	8.98	3.17	9.25	3.09	0.27	(0.94)
Avg items per order	2.68	1.19	2.74	1.60	0.06	(0.48)
Median zip income 2012	45950.90	16869.00	45068.71	16075.13	-882.19	(-1.34)
Share college 2012	23.98	9.00	21.99	11.75	-1.99***	(-4.77)
Share age 25-34 2012	21.67	4.08	21.20	5.69	-0.47*	(-2.38)
Share public transit 2012	41.09	16.32	39.26	17.33	-1.83**	(-2.73)
Yelp estabs in zip 2012	264.66	163.88	229.38	150.43	-35.28***	(-5.62)
Observations	1254		1254		2508	

C Establishment Growth and Continuous Gentrification Measures

Table C.1: Effect of gentrification on net change in tradables/nontradables stores

A linear model estimating the effect of gentrification on the percentage change in the number of establishments in the County Business Patterns. We control for the density of all retail in the zip code and the initial density of stores in each category from the County Business Patterns. In the even-numbered columns, we add additional demographic controls. We include city and category fixed effects, and standard errors are clustered at the Zip Code level.

	Dependent variable: percent change in establishments					
	All		tradables		nontradables	
	(1)	(2)	(3)	(4)	(5)	(6)
poverty rate	0.135	0.047	0.001	−0.001	0.002***	0.002*
	(0.122)	(0.107)	(0.002)	(0.003)	(0.001)	(0.001)
rent growth (%)	30.633*	19.868**	−0.247	−0.231	0.349*	0.206*
	(17.451)	(9.082)	(0.381)	(0.368)	(0.197)	(0.117)
poverty rate × rent growth	−0.724	−0.395	−0.001	−0.0001	−0.008	−0.004
	(0.889)	(1.151)	(0.008)	(0.007)	(0.013)	(0.015)
log(category density)	2.272	0.196	−0.034	0.009	0.163*	0.109
	(1.403)	(1.186)	(0.022)	(0.036)	(0.090)	(0.097)
log(establishment density)			0.027	−0.030	−0.147	−0.121
			(0.022)	(0.034)	(0.092)	(0.099)
log(share college)		−0.063		−0.005***		0.00003
		(0.222)		(0.002)		(0.002)
log(share 25 to 34)		0.748***		0.009***		0.008***
		(0.164)		(0.001)		(0.003)
log(share white)		0.005		−0.001***		0.0002
		(0.052)		(0.0004)		(0.0003)
log(median income)		−3.650		0.018		−0.030
		(6.562)		(0.045)		(0.053)
constant	−6.820	0.315	0.062	0.024	−0.033	0.071
	(6.008)	(5.052)	(0.058)	(0.078)	(0.065)	(0.068)
City FEs?	Yes	Yes	Yes	Yes	Yes	Yes
Category FEs?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	308	308	308	308	308	308
R^2	0.140	0.201	0.170	0.207	0.111	0.165
Adjusted R^2	0.117	0.168	0.145	0.171	0.084	0.128

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

Table C.2: Linear IV, effect of gentrification on change in number of establishments

An instrumental variables model estimating the effect of gentrification on the percentage change in the number of establishments in the County Business Patterns. We control for the density of all retail in the zip code and the initial density of stores in each category from the County Business Patterns. In the even-numbered columns, we add additional demographic controls. We include city and category fixed effects, and standard errors are clustered at the Zip Code level.

	Dependent variable: percent change in establishments					
	All		tradables		nontradables	
	(1)	(2)	(3)	(4)	(5)	(6)
poverty rate	0.133	0.082	0.001	−0.003	0.002**	0.003**
	(0.144)	(0.094)	(0.003)	(0.004)	(0.001)	(0.001)
rent growth (%)	33.928	−3.185	0.374	−0.442	0.456	0.017
	(34.952)	(42.163)	(0.407)	(0.881)	(0.323)	(0.321)
poverty rate × rent growth	−1.765	0.100	−0.076***	−0.067***	−0.011	0.014
	(2.277)	(2.230)	(0.005)	(0.023)	(0.023)	(0.020)
log(establishment density)	2.300*	0.420	−0.027	0.051	0.153**	0.143
	(1.353)	(0.902)	(0.043)	(0.047)	(0.074)	(0.090)
log(category density)			0.006	−0.062	−0.141*	−0.154*
			(0.037)	(0.053)	(0.074)	(0.091)
log(share college)		−0.036		−0.005		0.0003
		(0.163)		(0.003)		(0.002)
log(share 25 to 34)		0.793***		0.008**		0.008**
		(0.265)		(0.003)		(0.003)
log(share white)		0.007		−0.001		0.0001
		(0.045)		(0.001)		(0.0002)
log(median income)		−2.879		−0.063*		−0.007
		(6.265)		(0.038)		(0.050)
constant	−7.401	−0.685	0.048	−0.098	−0.020	0.060
	(4.579)	(3.081)	(0.107)	(0.082)	(0.044)	(0.047)
City FEs?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	308	308	308	308	308	308
R^2	0.133	0.192	0.032	0.083	0.109	0.151
Adjusted R^2	0.109	0.159	0.003	0.043	0.082	0.113

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