Separating Selection From Spillover Effects: Using the Mode to Estimate the Return to City Size

Hugo Jales
Department of Economics and Center for Policy Research
Syracuse University Syracuse, New York, 13244-1020
Phone: (315) 412-8764
hbjales@syr.edu

Boqian Jiang
Amazon Inc., Seattle
and Department of Economics, Syracuse University
bjiang03@syr.edu

Stuart S. Rosenthal
Maxwell Advisory Board Professor of Economics
Department of Economics and Center for Policy Research
Syracuse University Syracuse, New York, 13244-1020
Phone: (315) 443-3809
ssrosent@maxwell.syr.edu

December 31, 2018

Helpful comments are gratefully acknowledged from seminar and conference participants at Camp Econometrics, the Institute for Economics at Barcelona, Jinan University, the NBER urban workshop, and the University of Colorado. We also thank Nate Baum-Snow, Pierre Philippe Combes, Laurent Gobillon, Carlos Martins-Filho and Dan McMillen for helpful comments. All errors are our own.

Abstract

We develop a new method to identify and control for selection when estimating the productivity effects of city size. Selecting out low-performing agents has no effect on modal productivity but reduces the CDF evaluated at the mode. Agglomeration economies have the reverse effect. Estimates based on these principles confirm that selection contributes to productivity among full-time skilled workers but is limited for low-skilled workers. Doubling city size causes skilled and low-skilled worker productivity to increase by roughly 2.25 and 3.75 percent, respectively. Our approach can be applied to other settings provided necessary conditions formalized in the paper are satisfied.

JEL Codes: R00 (General Urban, Rural, and Real Estate Economics)
Key Words: Agglomeration economies, selection, city size, mode versus mean.

1. Introduction

A challenge for all studies that seek to estimate the productivity effects of agglomeration and city size is the need to separate out selection from spillover effects (see Rosenthal and Strange (2004) and Combes and Gobillon (2015) for reviews). This arises because cities are expensive places in which to live, work and operate a business (e.g. Rosenthal and Strange (2012), Combes et al. (2012), Black et al (2014)) so that only the most productive workers and companies participate – a threshold effect. It also arises because cities may attract unusually talented individuals who thrive on the intensity of urban life – a migration effect (e.g. Glaeser and Mare (2001), Rosenthal and Strange (2008), Combes et al. (2008), de La Roca (2017)). Both forms of selection contribute to higher levels of productivity in cities, confounding efforts to identify the causal impact of agglomeration on productivity. Building off of recent work by Combes et al. (2012), this paper develops a simple method that identifies the presence and nature of selection while yielding estimates of the causal effect of city size on productivity.[1]

Combes et al. (2012) argue that the presence or absence of selection effects can be identified by examining the shape of the observed factor return distribution. They note that if companies drop out when factor productivity is below a common threshold, selection left-truncates the observed distribution of returns. Assuming further that productivity thresholds increase with city size, they examine whether truncation is more prevalent among larger cities, using data on manufacturing plants in France. They fail to find evidence of such patterns and conclude that higher manufacturing productivity in larger French cities arises primarily from spillover effects and not from selection.

This paper extends the Combes et al. (2012) model in ways that yield a more general approach to controlling for selection effects. Our model applies to settings in which the latent factor return distribution is single peaked with an interior and well-defined mode. This is characteristic of wage and earnings patterns, for example, where it is common to model the underlying distribution of returns as log

[1] Common approaches to deal with the endogenous selection of workers and companies into different sized cities include the use of pseudo-random experiments (e.g. Ahlfeldt et al (2015)) and instrumental variables (e.g. Rosenthal and Strange (2008)). Nevertheless, the confounding effects of selection remain challenging and pseudo-experiments and instrumental variable approaches often offer solutions that do not extend beyond the immediate study.

normal (e.g. Chotikapanich et al (1997), Clementi and Gallegati (2005), Lopez et al (2006), and Sala-i-Martin and Pinkovsky (2009)). In such instances, if selection primarily culls out lower performing units to the left of the mode, then the CDF evaluated at the mode will be reduced. For sufficiently single-peaked factor return distributions – in a manner to be clarified – the modal level of productivity is also highly robust to selection whereas selecting out lower performing agents pushes up the observed mean return. Implementing these ideas points to three complementary regressions. The first regresses the CDF of the modal factor return in the city on log city size. The second and third regressions replace the dependent variable with the log modal factor return in the city and the log mean factor return in the city, respectively. Evidence of a negative city size effect in the first regression indicates that selection occurs disproportionately to the left of the mode. In such instances, the estimated modal return to city size should be below that of the estimated mean return, a prediction that can be confirmed using results from the second and third regressions. When evidence of selection is present, the return to city size based on the mode is highly robust to selection provided the factor return distribution is sufficiently single peaked.[2]

The model above can be used to evaluate the presence, nature and impact of selection both when selection arises from threshold effects and also when selection arises from migration (sorting). In the threshold model, we focus on the decision to participate, as when a firm survives and remains in business or when an individual chooses to work, while treating location choice as exogenous. In this model, selection contributes to higher productivity in larger metropolitan areas if city size increases operating costs and/or the intensity of competition (as in Combes et al (2012)). In the migration model, we focus on the possibility that talented individuals may sort into larger cities; in this model location is a choice but the decision to participate is treated as exogenous. However, while the two sources of selection are

[2] Selecting out low-performing agents will also push up the median return. However, throughout the paper we emphasize comparisons between modal and mean values rather than the median. In part, that is because the first regression above does not extend to median values since, by definition, the CDF evaluated at the median is always 50 percent. Also, the vast majority of studies in the agglomeration literature have focused on mean returns.

different, both point to the same qualitative patterns that motivate the regressions described above as will become clear.[3]

Later in the paper we formalize conditions under which the mode is an effective device to detect and control for selection. As emphasized above, a key condition is that the factor return density function must have a well-defined single-peaked shape with an interior mode. In the simplest case, if no selection occurs at the mode, as with the threshold example above, then the mode of the conditional density function is unaffected by selection and the mode does not shift. In a more general setting, if selection is random to the right of the mode, then regardless of the shape of the selection function to the left of the mode, once again the mode will not shift. If instead the selection probability increases beyond the mode, we show later in the paper that as the density function becomes increasingly flat, the mode will no longer be a useful device to detect and control for selection. However, provided the underlying density function has a well-defined single-peaked shape, the mode provides a robust opportunity to address selection effects as will become apparent.

This paper is the first we are aware of to use modal productivity when measuring the return to city size. The vast literature on agglomeration economies has instead focused almost exclusively on mean returns (see Rosenthal and Strange (2004) and Combes and Gobillon (2015) for reviews). Whether our estimates based on the modal return are informative for a given question depends in part on the degree to which the mode is of intrinsic interest for the outcome measure being considered. In instances where factor return distributions are symmetric and single-peaked, this is straightforward as the mode, mean and median are all alike. Where factor return distributions are single-peaked but skewed, the mode is still often as informative a summary measure of the central tendency of the return distribution as the median and mean, but for all three measures context matters.[4]

---

[3] Later in the paper we show that this is not the case if one instead attempts to use the density of the latent distribution evaluated at the mode to discern the presence of selection, as compared to the CDF.

[4] Our emphasis on using the mode as a measure of the central tendency of a distribution in the presence of selection effects has antecedents in earlier work by Lee (1989). Lee showed that under certain conditions, the mode from a truncated distribution is a consistent estimate of the conditional mean from the original, non-truncated distribution. As with Combes et al (2012), Lee (1989) focused on the case where the point of truncation is known and common

We use three data sets to illustrate our approach and to provide new estimates of the nature of selection and return to city size. To highlight threshold effects, we use law firm productivity for all law firms across the United States, drawing on establishment level data from Dun and Bradstreet. A key modeling assumption for this example is that entry and exist costs for law firms are low but annual operating costs are high. Moreover, we assume that lawyers initially have imperfect knowledge of their own ability and discover through experience whether the rewards from running their own firm dominate working for someone else's company. Law firms that are not sufficiently productive eventually close. Under these conditions, threshold effects should be more pronounced among older law firms since only the most productive companies survive. Assuming further that operating costs are higher in larger cities and/or that competition is more intensive, these patterns and related threshold effects are expected to increase with city size.

We also test our model using individual wage rates for married full-time working women age 25-55 who are white, non-Hispanic and native born, drawing on data from the 5 percent file of the 2000 U.S. census (obtained from IPUMS). It is well-established that married female labor supply is highly elastic so the decision to work full time is relevant to threshold effects (e.g. Heim (2007); Blau and Kahn (2007)).[5] Contributing to this view, Black et al. (2014) argue that higher commuting costs in large cities discourage married women from working. Married female labor supply decisions may also affect choice of metropolitan area as in Costa and Kahn (2000). For these reasons, wage patterns for married women are likely to be driven by a combination of threshold and migration effects. In the analysis to follow, our

---

across agents. Our work is also broadly related to the modal regression literature in statistics that construct statistical models by exploiting different properties of the mode (Huang et al. 2013; Yao and Li, 2014; Chen et al., 2016). That literature, however, does not consider the robust nature of the mode in the presence of selection. In economics, focus of modal values is rare. Cardoso and Portugal (2005) show that modal wage is a better measure of the central tendency of the underlying wage distribution when there is collective bargaining. Bound and Krueger (1991) and Hu and Schennach (2008) discuss how to use mode to account for certain forms of reporting errors when measuring the distribution income. Our approach is also related to the "identification at infinity" models in Chamberlain (1986), Lewbel (2007), and D'Haultfoeuille and Maurel (2013). These models assume that selection effects shrink to zero as certain key control variables approach "infinity". In these models, however, selection is based on one or more control variables whereas selection in our paper is based on the dependent (outcome) variable.

[5] Based on 1999-2001 CPS data, Blau and Kahn (2007) find that the elasticity of annual working hours with respect to *own* log wage is 0.357 for the married women and 0.046 for the married men. Moreover, the elasticity of annual working hours with respect to *spouse's* log wage is -0.192 for the married women and -0.006 for the married men.

models based on female wage rates are estimated stratifying women into skilled (college degree or more) and low-skilled (high school degree or less) individuals. That is because labor supply elasticity likely differs for high- and low-skilled married women and for that reason selection effects associated with city size may differ as well for reasons that are not directly modeled. In this context, our model has potential to reveal whether selection effects related to city size are more pronounced for skilled versus low-skilled married women.

To illustrate migration effects, we focus on male full-time workers age 25-55 who are white, non-Hispanic and native born, also drawing on the 2000, 5 percent U.S. census. Consistent with extensive work in the labor literature, for this sample, we treat the decision to work as inelastic and exogenous (e.g. Heim (2007); Blau and Kahn (2007)). Under the assumption that labor supply is exogenous, selection effects for this group are likely driven primarily by migration and location decisions.

For all three exercises, plots of the output measures (sale per worker and wage) are strongly single-peaked which suggests that the mode provides an effective opportunity to detect and control for selection. Indeed, for all three samples, estimates based on the CDF evaluated at the mode yield compelling evidence that selection contributes to productivity in larger cities. That evidence is reinforced by related estimates that compare the return to MSA size based on the mode and the mean. For older law firms that have survived the initial weeding out effects of competition, the modal return from doubling city size is approximately 1.16% as compared to 1.63% at the mean. For the wage applications, for both married women and full-time working men, the return to MSA size among skilled workers (college or more) is roughly 2.25% when evaluated at the mode. When evaluated at the mean, the corresponding estimates are roughly 2 percentage points higher, indicating that selection nearly doubles the perceived returns to city size when estimating based on the mean.

Interestingly, evidence of selection among low-skilled workers (high school or less) for both married women and men is much weaker in the CDF regressions, and consistent with that, the returns to city size are very similar when based on both the mode and mean. Moreover, for low-skilled workers the wage elasticity with respect to MSA size is roughly 3.9% for women and 3.5% for men somewhat higher than for skilled individuals.

We proceed as follows. The next section develops our model. Section 3 describes the data and summary statistics. Section 4 discusses how to measure the mode. Section 5 presents the results and Section 6 concludes.

2. Model

This section presents our modeling framework. We begin with the influence of agglomeration economies on the distribution of worker productivity in large versus small cities in the absence of selection effects. The model is then extended to allow for threshold-based selection and selection arising from migration. The focus in these initial subsections is qualitative and intuitive. For a broad set of common distributions and selection processes, the final subsection formalizes conditions that govern the degree to which the mode of a single-peaked density function can be used as a robust way to identify and control for selection.

2.1 Productivity spillovers from city size

Suppose initially that there are no selection effects that influence the distribution of productivity in large versus small cities. Instead, the only force that causes productivity distributions to differ across metropolitan areas are spillovers arising from city size. To simplify, we assume two different size cities, denoted as 0 for small cities and 1 for large cities. Productivity spillovers from agglomeration increase productivity in larger cities.

Let individual worker productivity be denoted by y which, to simplify discussion below, we interpret in logs. Also let $f_0(y)$ and $f_1(y)$ represent the distribution of productivity among individuals in small and large cities, respectively. Cities are assumed to be large enough that $f_0(y)$ and $f_1(y)$ are approximately continuous on y, and worker productivity in a size-0 city depends only on a worker's intrinsic level of skill. If agglomeration economies increase productivity by a common percentage for all workers, $f_1(y)$ shifts to the right relative to $f_0(y)$. If instead, the returns to city size increase with skill, possibly because more talented workers are better able to take advantage of large city opportunities, then this would create a "dilation effect" (Combes et al., 2012) causing $f_1(y)$ to become right skewed with an elongated right tail. Allowing for both effects, for a given individual, productivity in a larger city is given by,

$$y_1 = \beta_0 + \beta_1 y_0 \tag{2.1}$$

In expression (2.1), β_0 measures the common productivity boost for all workers in a larger city while $\beta_1 > 1$ would imply that the returns to city size increase with worker skill. As in Combes et. al (2012), expression (2.1) specifies spillover effects in a linear form for which shift and dilation effects preserve an individual's productivity rank within a given city. Under these conditions, the cumulative distribution function (CDF) for productivity up to a given skill level, y_0, is the same in each city, denoted as F_0 and F_1,

$$F_0(y_0) = F_1(y_1(y_0)) \tag{2.2}$$

Substituting for y_1 from expression (2.1) and taking derivatives, the relationship between large and small city productivity densities is given by,

$$f_1(y) = \frac{1}{\beta_1} f_0\left(\frac{y-\beta_0}{\beta_1}\right) \tag{2.3}$$

Our most important modeling assumption in the empirical work to follow, referred to as Assumption 1, is given by:

Assumption 1: *$f_0(y)$ is single peaked with a well defined mode at an interior location.*

In conjunction with spillover effects as modeled in (2.3), this assumption has important implications for the shape of productivity density functions in large versus small cities. To illustrate, we took 10,000 random draws of $\log(y_0)$ from a normal distribution, mirroring assumptions in the labor and agglomeration literatures that typically treat wage and earnings distributions as log-normal. We set the standard deviation of the distribution to 0.4, a value small enough to conveniently plot the distribution.

We also set $\beta_0 = 0.5$ and $\beta_1 = 1.3$, implying that larger cities boost productivity by 50 percent regardless of skill ($\beta_0 = 0.5$) and an additional 30 percent with a doubling of worker ability (since $\beta_1 = 1.3$). These values are purely for illustrative purposes and have been chosen to make plotting of the distributions to follow convenient while also allowing us to higlight key general principles.

Figure 1 traces out the simulated productivity density functions for large cities (the dashed red line) and small cities (the solid black line) based on the specified parameters above. Notice that for large cities, the density function is right shifted with an elongated right tail (right skewed) relative to the density function for small cities. The large city density is also flatter, with a lower density for any given level of productivity, and a right shifted mode. A positive value for β_0 shifts the large city distribution along the x-axis by β_0 units while preserving its shape. This is apparent from (2.1) and (2.3). In (2.1), the derivative of y_1 with respect to β_0 is 1 while from (2.3) the large city density with $\beta_1 = 1$ is $f_0(y - \beta_0)$. Observe, however, that even though β_0 is just 0.5, the mode in Figure 1 shifts by a larger amount, from 0.85 in small cities to 1.55 in large cities. The additional rightward shift in the mode is because of the dilation effect arising from $\beta_1 > 1$ which draws the mode further to the right, although not immediately apparent from a casual viewing of (2.3). As is evident in the figure, the mode in the large city density is also not as pronounced relative to a smaller city. This also is a consequence of $\beta_1 > 1$, which flattens the density function by shifting mass from the center of the distribution into the elongated right tail, and bearing in mind that the density function must always integrate to 1.

2.2 Threshold effects

Consider now the influence of threshold effects that contribute to selection and which differ across agents within a given city. We assume that the latent productivity distributions are identical in small and large cities but threshold effects are more pronounced in larger metropolitan areas. For simplicity, small city residents are described below as participating in the labor market with probability 1 regardless of skill, or $\pi_0(y) = 1$, where $\pi_0(y)$ is the probability of participating. If in the large city $\pi_1(y)$ is also constant with $\pi_1(y) = p < 1$, then the selection process is random and $f_1(y) = f_0(y)$. More relevant for our context, is the possibility that in large cities participation increases with skill, which we formalize as our second core modeling assumption:

> **Assumption 2:** *In large cities, the probability of participating in the labor market increases monotonically and linearly with skill, $\partial\pi_1(y)/\partial y > 0$, up to $\pi_1(y) = p \leq 1$ and remains at p thereafter.*

Assumption 2 captures the tendency for operating costs to be higher in larger cities and/or the environment more competitive (as in Combes et al (2012)). For that reason, weaker companies are more likely to drop out in larger cities relative to outcomes in smaller metropolitan areas. Analaogously, because commuting costs tend to be higher in larger cities, Assumption 2 captures the sense that lower productivity workers are more likely to drop out of the labor force in larger cities relative to smaller metropolitan areas (see Black et al (2014) for related discussion).

Allowing for heterogeneous threshold effects as above, expression (2.3) becomes,

$$f_1(y) = \frac{\pi_1(y)}{\beta_1 c} f_0\left(\frac{y-\beta_0}{\beta_1}\right) \tag{2.4}$$

where $c = \int \pi_1(u) f_0(u) du$ captures the mass lost to selection and ensures that the conditional density integrates to 1. Note also that $\pi_1(y) < 1$ reduces the density for a given level of y in the larger city.

We illustrate the qualitative effects of threshold-based selection in Figures 2 and 3 using the same simulated data as for Figure 1, first without and then with spillovers. In Figure 2, we set $\beta_0 = 0$ and $\beta_1 = 1$, consistent with the absence of agglomeration economies. The $\pi_1(y)$ function is specified such that $\pi_1(y)$ increases up to a value of 1 at the mode of the latent distribution (at $y = 0.85$) and remains at 1 thereafter.[6] Imposing these features, ten percent of the simulated work force is selected out of the large city labor market, all of whom have skill levels to the left of the mode. The important point to recognize in Figure 2 is that even though all selection occurs to the left of the mode, selection steepens the slope of the large

[6] More precisely, we set $\pi(y) = -0.27 + 1.5y$ for $y \leq 0.85$ and $\pi(y) = 1$ for $y \geq 0.85$. Specified in this manner, $\pi_1(y) = 0$ for the lowest level of y in the simulated sample and approaches 1 asymptotically from below at $y = 0.85$.

city density function on both sides of the mode while also increasing the height of the mode. Together, these effects cause the modal level of productivity in the density function to become more pronounced.

Figure 3 illustrates the combined influence of threshold and spillover effects. In this instance we set β_0 and β_1 to the values used in Figure 1 and specify $\pi_1(y)$ as in Figure 2. In Panel A, notice that the influence of threshold effects is difficult to discern relative to the pattern in Figure 2. That is because dilation associated with $\beta_1 > 1$ flattens and right-skews the distribution causing the mode to become less pronounced. This offsets the tendency for threshold effects to accenuate the mode. On the other hand, because in this example all selection is to the left of the mode in the large city population, the CDF evaluated at the mode must be reduced relative to the CDF at the mode in the small city distribution. This is readily apparent in Panel B which shows that the corresponding CDFs evaluated at the respective small and large city modes are 0.34 and 0.27.

The patterns in Figures 2 and 3 motivate our first regression described in the Introduction and point to a simple way to identify whether selection occurs more to the left or to the right of the mode of a latent productivity distribution. Moreover, in the special case where selection occurs only to the left (or right) of the mode, the difference in CDF evaluated at the mode for large versus small cities is an exact measure of the extent of selection.

2.3 Migration effects

Consider next the influence of migration as the source of selection effects. In this instance, we assume a common aggregate single-peaked (latent) productivity distribution from which individual workers sort into two types of cities, small (size 0) and large (size 1). In this setting, $\pi_1(y)$ represents the probability that a worker with skill level y chooses to locate in the larger city. As with the threshold model, if $\pi_1(y)$ equals a constant p, the selection process is random and $f_1(y) = f_0(y)$. In this instance, selection would not affect the CDF evaluated at the modes in small and large cities. A more realistic scenario, however, is that $\pi_1(y)$ increases in a smooth, monotonic fashion with y, analogous to Assumption 2 above, and consistent with the view that higher skilled individuals are more likely to select into larger cities. This would also simultaneously reduce skill levels in smaller urban areas. Nevertheless, the core patterns outlined above that motivate our key estimating equations still hold.

To clarify, consider first an extreme but illustrative selection process. We set $\pi_1(y) = 0$ for $y \leq y^*$ and $\pi_1(y) = 1$ for $y > y^*$, where y^* is an interior point in the aggregate distribution. Specified in this manner, all workers below y^* sort into the small city while all of those above y^* sort into the large city. Figures 4a and 4b highlight implications of these conditions using the same simulated data as above. The key difference between the figures is whether y^* is below or above the modal level of skill in the aggregate distribution, denoted by y_m and equal to 0.85 as before.

In Figure 4a we set y^* equal to 0.65 so that $y^* < y_m$. This causes the small city density function (in the top portion of Panel A) to increase monotonically with y with a mode equal to $y^* = 0.65$. The large city density, in contrast (in the top portion of Panel B), declines monotonically from a modal value equal to $y_m = 0.85$. In the lower portions of each panel, notice also that the CDF evaluated at the mode in the small city equals 1 since all workers have productivity below y^*, while the CDF for the large city must be less than 1 since the mode is at an interior location. In Figure 4b we instead set y^* equal to 1.0 so that $y^* > y_m$. This causes the small city mode to equal y_m while the large city mode becomes y^*. Under these conditions, the CDF evaluated at the large city mode collapses to 0 and the corresponding CDF for the small city is positive but less than 1. The important point to emphasize from these patterns is that regardless of whether y^* is above or below y_m, the CDF evaluated at the mode declines with city size. This is the same pattern as obtained for the threshold model.

Consider now a more realistic characterization of migration for which $\pi_1(y)$ increases with y in a smooth, gentle and monotonic fashion. To illustrate the influence of such a process, in Figures 5 and 6 we again display large and small city productivity density functions using the same simulated data as before. In both figures, we also specify $\pi_1(y)$ so that the likelihood of locating in a large city increases linearly with y at rate $0.1y$ and with $\pi_1(y)$ set equal to 0.5 for the least skilled individual in the sample.[7] In Figure

[7] This also ensures that $\pi_1(y) = 1$ for the most skilled individual in the sample. Specified in this manner, 60 percent of workers in the simulated sample sort into the large city.

5a, spillover effects are set to zero with $\beta_0 = 0$ and $\beta_1 = 1$ in expression (2.4), while in Figure 5b we allow for spillover effects using the same specification as for Figure 1.

Focusing first on Figure 5a, it is evident that the specified migration process has little effect on modal productivity values, similar to the pattern in Figure 2 for threshold effects. Migration does, however, have noteworthy effects in Figure 5a. Relative to large cities, migration increases the height of the density function evaluated at the small city mode and steepens the slope of the density function on either side of the small city mode. This is opposite from the influence of threshold effects in Figure 2, and reinforces the principle that the height of the density function evaluated at the mode and the slopes of the density function on either side of the mode are not necessarily reliable indicators of selection effects even when the modeling assumptions 1 and 2 hold. This conclusion is made even stronger when the influence of productivity spillovers is taken into account. In the upper panel of Figure 5b, dilation arising from $\beta_1 > 1$ flattens and right-skews the productivity density function in large cities relative to small cities, further masking the influence of migration (as in Figure 3). In the lower panel of Figure 5b, however, which plots the CDFs for the small and large city productivity distributions, the respective CDFs evaluated at the modes are 0.38 and 0.31. Once again, the CDF evaluated at the mode declines with city size.

Returning to the upper panel of Figure 5b, observe also that the modal productivity values for small and large cities are 0.85 and 1.55, respectively. Because the underlying latent distribution is single peaked and the selection process is not too extreme, the difference in modal productivity between large and small cities is largely unaffected by selection and reflects primarily the effect of city size on productivity. More generally, because migration shifts mass to the right in the large city productivity density function relative to the small city, that will tend to increase the spread between large and small city means (and medians). This once again suggests that the mode is less sensitive to selection relative to the mean and median of the underlying productivity density functions.

The results from the threshold and migration models above indicate that for a single peaked factor return density function, the mode can be used to infer whether selection occurs disproportionately to the left or the right of the mode, and can also be used to mitigate the effect of selection. The following section formalizes these results.

2.4 By how much does the mode shift in response to selection?

This section extends the qualitative analysis above by establishing conditions that govern the extent to which selection will cause the mode of the outcome density function to shift. This will help to formalize conditions under which one can use the CDF evaluated at the mode to infer evidence of selection since a shift in the mode will also affect the value of the CDF at the mode, ceteris paribus. It is also necessary to assess the extent to which the mode offers a robust way to control for selection when estimating the returns to city size.

2.4.1 Elasticity condition

Suppose that selection effects are present but agglomeration economies are not. Then $\beta_0 = 0$, $\beta_1 = 1$, and the conditional density in (2.4) becomes,

$$f_1(y) = \frac{\pi_1(y)}{c} f_0(y) \; . \qquad (2.5)$$

The question we seek to answer is by how much selection may shift the mode of the conditional density $f_1(y)$ relative to the unconditional density $f_0(y)$. Since $f_0(y)$ is assumed to be differentiable and single peaked, its slope at the mode is zero. Differentiating (2.5) with respect to y and setting the derivative to zero, the modal value for y (denoted by y_m) in the conditional density $f_1(y)$ must satisfy,

$$\frac{\pi_1'(y)}{\pi_1(y)} = -\frac{f_0'(y)}{f_0(y)} \qquad (2.6)$$

Expression (2.6) indicates that at the mode of the conditional distribution, a small change in y yields equal magnitude but opposite signed percentage changes in the selection probability and the density of y. Multiplying both sides of (2.6) by y this can be expressed as an elasticity condition,

$$\xi_{\pi,y} = -\,\xi_{f_0,y} \tag{2.7}$$

where $\xi_{\pi,y} \approx \frac{\%\Delta\pi(y)}{\%\Delta y}$ and $\xi_{f_0,y} \approx \frac{\%\Delta f_0(y)}{\%\Delta y}$.[8]

Expression (2.7) says that at the modal value of the conditional density function, the elasticity of the selection probabilty is equal to minus the elasticity of the latent density. Provided that both the density and selection functions are log-concave, the conditional density $f_1(y)$ will also be single-peaked and the elasticity condition above will be satisfied at a unique value for y. This uniqueness property follows from arguments in An (1996) and Saumard and Wellner (2014).[9] The assumptions that $f_1(y)$ is single-peaked and that $\pi_1(y)$ increases monotonically over the relevant range of y satisfies these conditions and ensures a unique interior solution for (2.7).

Figure 6 illustrates these principles. The upper panel displays a differentiable single peaked density function and a linear monotonically increasing selection function with a vertical intercept at the origin. The lower panel plots the corresponding values for $-\,\xi_{f_0,y}$ and $\xi_{\pi,y}$. In the case where y is normally distributed, it is straightforward to show that $-\,\xi_{f_0,y} = y(y - y_m)/\sigma^2$ with a slope of $(2y - y_m)/\sigma^2$ that increases at a rate of $2/\sigma^2$. In this instance, $-\,\xi_{f_0,y}$ initially declines from zero at the origin to a minimum at $y = y_m/2$, and increases monotonically thereafter, taking on a value of 0 at the mode and positive values thereafter. As drawn in the upper panel, the selection function has a constant unit elasticity up to the point where $\pi(y) = 1$, after which $\xi_{\pi,y} = 0$. The elasticity plots in the lower panel must therefore intersect to the right of y_m, indicating that selection shifts the mode of the conditional density function to the right. The precise amount by which the mode shifts is clarified next.

---

[8] The elasticities above express the percent change along the vertical axis in response to a percent change along the horizontal axis. This is the inverse of familiar demand and supply elasticities. The elasticities in (2.6) are specified as above because y is the exogenous determinant of f and π.

[9] Proposition 2 in An (1996) indicates that a random variable y is distributed in a log-concave fashion if and only if its density function is strongly unimodal. Proposition 3.2 in Saumard and Wellner (2014) indicates that the product of two log-concave functions is log-concave. Together, these principles imply that $f_1(y)$ is unimodal given the assumed shapes of the latent density and selection functions.

2.4.2 Magnitude of mode shift

Below we formalize three propositions that govern the amount by which the mode may shift in response to selection effects. This also implicitly describes conditions under which the mode provides a robust way to discern and control for selection.

The first case is relevant to when all selection occurs on just one side of the mode. In this instance, it is not necessary to specify any structure for the selection function.

> **Proposition 1:** *Given Assumption 1, if selection does not remove any mass from the latent density mode y_m then ...*
>
> (i) *The mode of the conditional density function does not shift.*
> (ii) *The CDF evaluated at the mode will decline if selection occurs more to the left of the mode and will increase if selection occurs more to the right of the mode.*
> (iii) *These conditions hold regardless of the shape of the selection function.*

<u>Proof</u>: By definition, the mode of a single-peaked density function has the highest density. Removing mass from all other points in the distribution increases the density at the original mode relative to other points in the distribution and ensures that the mode does not shift.

The second case includes situations in which over a relevant range of y the selection probability $\pi_1(y)$ is a constant less than 1, indicating that selection is random. In this instance, the selection function will be flat and $\xi_{\pi,y} = 0$.

> **Proposition 2:** *Given Assumptions 1 and 2, if the selection probability is constant for $y > y_p$ where y_p is to the left of y_m, then the mode will not shift and points (i) and (ii) from Proposition 1 will hold even when mass is withdrawn from the mode.*

<u>Proof:</u> If $\pi_1(y) = p < 1$ from a point y_p to the left of the mode and beyond, then selection is random for all $y > y_p$. With sufficiently large sample, the rank order of the density at the mode will be preserved and the mode will not shift.

Consider now a third and more general case in which the selection function increases with y and reaches an asymptote p to the right of the mode at $y_p > y_m$. In this instance, additional structure must be imposed on both the selection and the density functions in order to characterize the degree to which the mode will shift in response to selection. For that reason, we replace assumptions 1 and 2 by assuming the following shapes for the latent density and selection function.

Assumption 3: *$f_0(y)$ is drawn from a distribution that belongs to the family of generalized error distributions for which …*

$$f_0(y) = \frac{1}{2^{\kappa+1}\sigma\Gamma(\kappa+1)} e^{-\kappa\left|\frac{y-\mu}{\sigma}\right|^{\frac{1}{\kappa}}} \tag{2.8}$$

where μ is the mean, σ is the scale parameter or dispersion of the distribution, κ is the shape parameter that governs the degree to which the mode is sharply defined (with range from 0 to ∞), and Γ denotes the gamma function.

The generalized error distribution is a symmetric distribution with the mode, median and mean all equal to μ. For the density function above, κ = 1/2 corresponds to the normal distribution. For κ = 1, the resulting distribution is a double exponential or Laplace distribution which has a more sharply defined mode than the normal, and for κ < 1/2 the distribution has a flatter mode than the normal. In the limit, as $\kappa \to 0$, $f_0(y)$ converges to a uniform U(μ - σ, μ + σ) and at the other extreme, as $\kappa \to \infty$, $f_0(y)$ becomes degenerate with all mass concentrated at a single value for *y*.

Assumption 4: The probability of selecting into and participating in a large city labor market is given by,

$$\pi_1(y) = \begin{cases} a + by, & \text{for } y < y_p \\ p \leq 1, & \text{for } y \geq y_p \end{cases} \tag{2.9}$$

with $0 \leq \pi_1(y) \leq 1$, $b > 0$, and $y_p > y_m$.

Assumption 4 makes explicit that $\pi_1(y)$ reaches an asymptote $p \leq 1$ at y_p to the right of the latent density mode y_m. The selection elasticity in expression (2.7) is then given by,

$$\xi_{\pi,y} = \begin{cases} b\left[\frac{y}{a+by}\right], & \text{for } y < y_p \\ 0, & \text{for } y \geq y_p \end{cases} \quad . \tag{2.10}$$

A subtle property of (2.10) that is relevant for the discussion to follow is that for $y \geq 0$ the elasticity $\xi_{\pi,y}$ must always be in the inelastic range between 0 and 1. To see this, recall that $\pi_1(y)$ is bounded between 0 and 1 since $\pi_1(y)$ is a probability and also that *b* is assumed to be positive. If *y* is interpreted in levels with the lowest skill level at $y = 0$, $\pi_1(0) = a$ which requires $a \geq 0$ to ensure that

$\pi_1(0)$ is nonnegative. If instead y is interpreted in logs with a range over the real number line, for a negative value of y, $by < 0$ and a must be strictly positive ($a > 0$). Thus, $\pi_1(y)$ between 0 and 1 requires that $0 \leq a \leq 1$. Imposing that restriction on (2.10), for $y \geq 0$ the elasticity $\xi_{\pi,y}$ equals a maximum of 1 when a equals zero and shrinks towards zero at rate $-by(a + by)^{-2}$ as a increases towards 1. For these reasons, $\xi_{\pi,y}$ is always in the inelastic range for $y \geq 0$. We will return to this point shortly.

For the density and selection functions assumed above, the extent to which the mode shifts in response to selection is governed by the parameters a, b, σ and κ, subject to the elasticity condition in (2.7). These and related principles are formalized in Proposition 3.

Proposition 3: *If the selection function reaches an asymptote p, at* $y_p > y_m$, *and if assumptions 3 and 4 hold, then …*

(i) *The elasticity condition in (2.7) is satisfied at a value for y equal to:*

$$y_{m,1} = y_m + \sigma\xi_{\pi,y}^k$$

(ii) *The mode of the conditional density, denoted by* $y_{m,1}$, *is given by:*

$$y_{m,1} = min\left[y^*, \;\; y_m + \sigma\xi_{\pi,y}^k\right] \tag{2.11}$$

Proof: See appendix A.

Proposition 3 has several subtle features. If the selection function hits an asymptote before the elasticity condition from (2.7) is satisfied, then the mode changes by $y_p - y_m$. In this instance, if y_p is close to y_m then selection will have little effect on the mode. If instead y_p is sufficiently large so that the elasticity condition determines the conditional mode, then the mode shifts to the right by the product of the scale parameter of the density function multiplied by the selection elasticity raised to the k power, $\sigma\xi_{\pi,y}^k$. If selection is random, as with the case associated with Proposition 2, then $\xi_{\pi,y} = 0$ and (2.11) confirms that the mode does not shift. If instead $\xi_{\pi,y} > 0$, then as the scale parameter of the latent density increases, the distribution becomes more spread out and this amplifies modal shift.

Recall now that $\xi_{\pi,y}$ equals 1 when the vertical intercept of the selection function in (2.9) is zero. However, the far more realistic case is that $\xi_{\pi,y}$ will be positive and less than 1 since a will typically be positive, implying some positive probability of selecting into a larger city labor market even for very low-skilled individuals. In that instance, as k becomes increasingly large, the generalized error distribution collapses to a point and the mode should not shift. Expression (2.11) confirms this since $\xi_{\pi,y}$ < 1 implies that $\xi_{\pi,y}^{k}$ goes to zero as k becomes large.

For the normal distribution, k = ½. In this instance, (2.11) indicates that a gently sloped linear selection process (for which y^* is not binding) will cause the mode to shift by the scale parameter multiplied by the square root of the selection elasticity, $\sigma\xi_{\pi,y}^{0.5}$. Given how often normal distributions are relevant to economic data, this case is of particular interest. But whether $\sigma\xi_{\pi,y}^{0.5}$ is a small or a large amount will depend on the context in question and the related plausible values for σ and $\xi_{\pi,y}$.

The special case of the uniform distribution is also instructive. In the limit as k goes to 0, the generalized error distribution converges to a uniform distribution, U(y_m - σ, y_m + σ). In this extreme case $f_0(y)$ becomes a constant equal to q for the relevant range of y. From (2.5) it follows that $f_1(y) = \pi_1(y)\frac{q}{c}$. The conditional density function therefore takes on the shape of the selection function scaled by q/c. With b > 0, the mode of $f_1(y)$ must shift all the way to the right edge of the distribution. This is also confirmed by (2.11), where the shift in the mode is given by σ when k = 0. Consider also that with positive mass to the left of y_m + σ, the mean of the conditional density shifts by less than the mode. Thus, for a sufficiently flat density function, it is possible for selection to cause the mode to shift *further* to the right than the mean. This reinforces a central point emphasized at the start of the paper: the mode is an effective way to discern and control for selection when the density function has a well defined interior mode. If that condition is not met, then the mode is not an effective tool to address selection.

Two final comments remain when considering the viability of using the mode to test and control for selection effects. First, sample size must be large enough to yield sufficiently reliable estimates of the mode for purposes of evaluating the CDF at the mode and the impact of city size on modal productivity. This point is considered further in the sections to follow. Second, the mode needs to be of intrinsic interest for the problem being considered. While these conditions will not always hold, they are met in many problems regularly considered in economics.

3. Data and Summary Statistics

3.1 Three datasets

This section describes the three datasets used to estimate the returns to city size based on the model above. In the first instance, we use sale per worker at all law firms in the United States. As described in the Introduction, entry and exit costs are low for lawyers operating their own firms. Suppose also that lawyers only learn whether they can profitably operate their own firm from experience, and the returns from operating a law firm are high if the venture is successful. Under these conditions, a wide range of lawyers may attempt to establish their own companies, including many who are less adept but do not realize their firms are likely to fail. This would reduce tendencies for threshold-based selection at the point of entry. Over time, however, lawyers discover their type and weaker companies drop out so that threshold effects should be especially apparent among older companies. Moreover, with higher operating costs and a more competitive environment in larger cities, evidence of threshold-related selection and related differences between new and older law firms should increase with city size. These ideas point to testable hypotheses.

As also described in the Introduction, among full-time working married white, non-Hispanc women, it is plausible that both threshold and migration effects would contribute to selection and higher observed wages in larger cities. Threshold effects, for example, could arise if longer commute times in larger cities discourage women from working (e.g. Black et al (2014), while migration effects could be associated with job market co-location challenges that draw skilled couples to larger cities (e.g. Costa and Kahn (2000)). In contrast, for full-time working white men, labor supply is highly inelastic. For this group, migration effects seem likely to be the dominant source of selection. The data used for each of these applications is describe below.

3.2 Law firm establishment data from Dun & Bradstreet

We collected 2016 establishment-level data for all law firms in the United States (excluding Alaska and Hawaii) from the Dun & Bradstreet Million Dollar Database. The data provides information on establishment location, level of employment, sales, industry (SIC 8-digit code), year established, and other information. Compared to the Census data, an advantage of Dun & Bradstreet database is that it provides comprehensive coverage of small businesses including those with just one or two-workers.[10]

The data were collected in December 2016 and provide a snapshot of all law firms operating in the U.S. at that time. We use establishment-level sales per worker as a proxy for productivity, trimming out the top and bottom 0.1% of the data to reduce outliers.[11] Certain types of law offices may be more prevalent in large cities (e.g. corporate law). Because concerns about selection stem from unobserved factors embedded in the error term, we pre-cleaned the data to difference out the average return for the primary classifications of law firms identified in the data.[12] This was done by regressing individual establishment sale per worker on dummy variables for each type of 8-digit law office reported by Dun and Bradstreet. We then added back to the residual from this regression the average sale per worker for general law offices/attorneys which account for 90% of the sample. The adjusted cleaned residual has the same sample mean as in the raw data and is used as the dependent variable.

A key part of our empirical strategy is to measure the mode of the adjusted sales per worker distribution in each MSA. To ensure sufficient sample, we retain only MSAs for which all of the following conditions are satisfied: (i) more than 30 law firms age five or younger are present, (ii) more than 30 law firms over five years in age are present, and (iii) MSA total population is over 100,000. For all of the law firm models, MSA size was estimated using the 2015 American Community Survey which

[10] In our sample, there are 545,873 law establishments. Of these, 8.5% have one worker, 62.8% have two employees, 15.0% have three workers, and 13.5% have four or more workers. In comparison, in the 2012 Economic Census, there are 186,831 law establishments in the U.S. The main reason for the difference is that Census indicates that it does not "survey very small businesses". For details see the Census website: https://www.census.gov/programs-surveys/economic-census/about/faq.html.

[11] Similar trimming procedure is also used in Combes et al. (2008), Combes et al. (2012) and Gaubert (forthcoming).

[12] Based on SIC 8-digit codes, approximately 90% of the sample is coded as general law offices/attorneys. The remaining 10% of the sample is coded into more specialized classifications, including corporate law, family law, etc.

is approximately the same period as the 2016 law firm sample.[13] Cleaning as above, we are left with 239 MSAs. The total count of law firms in the sample is 545,873 firms. Of these, 74,079 firms are young, defined as five years or less in age, and 471,794 firms are old, defined as over five years in age.[14]

Table 1 Panel A presents summary statistics of sales per worker for all law firms sample and also separately by age group (young and old). Based on the 25th and 75th quantile, the majority of the sales per worker measures fall within the range of $60,000 and $85,000. Measured at the mean and different quantiles, old firms have higher sales per worker than the young firms, indicating that older law firms are more productive than younger companies.

Figure 7a provides kernel density plots of sales per worker for the all firms sample as well as for the different age groups. In each panel, the sales per worker distribution is strongly singled peaked.[15] In Figure 7b, kernel density plots are provided again, stratifying each sample into small (population < 1 million) and large (population > 2.5 million) MSAs. For each sample (all firms, young and old), the large-city distribution of sales per worker is clearly right-shifted as compared to small cities.

It is also worth noting that for each panel in Figures 7a and 7b there are several spikes in the density estimation. This likely reflects the presence of rounding errors in the sales data since firms tend to report sales rounded by thousands of dollars. Additional rounding errors in reported counts of workers may also contribute. However, because modal values are used as dependent variables in the regressions to follow measurement error of this sort should increase the variance of our estimates but have limited effect on our point estimates.

---

[13] The 2013 Office of Management and Budget metropolitan area delineations are used to define MSAs.

[14] Among the 545,873 establishments, age related information was missing in the D&B data for 67,358 establishments (12% of the sample). For roughly 200 of these firms, we searched the companies on the web by establishment name (which is also reported by D&B). In each instance, the establishments was over 5 years in age. For that reason, we classified all law firms in D&B with missing age information as over 5 years in age (i.e. as old establishments).

[15] There are also several spikes in the density estimation, indicating rounding errors in the sales per worker data. The rounding errors are likely to be caused by the fact that firms tend to report sales rounded by thousands of dollars. We discuss how we deal with such rounding errors in Section 4.

3.3 Married white female full-time workers in the 2000 Census

The sample of female married white non-Hispanic native-born full-time workers (age 25-54) was obtained from the 2000 decennial census 5% public use micro sample (PUMS) from IPUMS.[16] Full-time workers were coded as those who report working at least 35 hours per week and 40 weeks per year.[17] Hourly wage was used as a proxy for productivity and was computed by dividing annual earnings by annual hours worked. As above, we trim the top and bottom 1% of the sample based on hourly wages to remove outliers.

Also analogous to above, the data were pre-cleaned to remove the influence of observables. This was done by regressing individual wage on age fixed effects, education fixed effects, occupation fixed effects and industry fixed effects.[18] We retain the wage residual from each worker and calculate the adjusted "cleaned" wage by adding back a constant that sets the mean of the adjusted wage series equal to that of the raw data sample mean. Wage data were cleaned separately for skilled (college degree or more) and low-skilled (high school degree or less) workers separately.[19]

MSA population size was estimated using the 2000 census, the same year as the wage data were drawn from.[20] We retain only those MSAs for which all of the following conditions are satisfied: (i) more than 100 married female non-Hispanic white native-born workers with a college degree or more present, (ii) more than 100 married female non-Hispanic white native-born workers with a high school degree or less are present, and (iii) MSA total population is over 100,000. The data cleaning procedure leaves us a

[16] See Steven et al., 2015 and www.ipums.org. Observations from Alaska and Hawaii were excluded.
[17] We focus on full-time workers in part to reduce measurement error when calculating hourly wages which is more pronounced among part-time workers. See Baum-Snow and Neal (2009) for related discussion.
[18] To be specific, there are 15 age fixed effects, 359 occupation fixed effects and 94 industry fixed effects. In the census, the most detailed version of occupation classification is at 6 digits, which is too refined that certain occupations do not have enough sample size to yield precise estimates of fixed-effects. Therefore, we choose to control for occupation fixed effects using 5-digit classification. As a robustness check, we find that controlling for occupation fixed effects at 4-digit or 6-digit level also yield similar results.
[19] Using this approach, a small number of observations had negative adjusted wage. Dropping these obsrevations did not affect our results.
[20] The population estimate is obtained through the IPUMS website. Link: https://usa.ipums.org/usa-action/variables/MET2013#description_section

sample composed of 152,704 skilled married female workers and 153,168 low-skilled married female workers from 216 MSAs in the United States.

Table 1, Panel B provides summary statistics of adjusted hourly wage for the married female workers. Measured at the mean and each quantile, the adjusted hourly wage is higher among the skilled workers. Figure 8 Panel A and B present kernel density plots of the adjusted hourly wage for high- and low-skilled workers. The first thing to note is that the both distributions are strongly single-peaked. The density plot for the skilled workers (Panel A) also has a longer right tail and a larger variance as compared to the plot for low-skilled workers (Panel B). Splitting the samples into small and large MSAs, we reproduce the density plots in Panels C and D. For both groups of workers, the wage density plots for large cities is right-shifted and dilated as compared to the density plot for small cities. It is also worth noting that the density plots in Figure 8 are smooth, indicating less rounding error relative to the law sample in Figure 7.

3.4 Male full-time white workers in the 2000 Census

Male non-Hispanic white full-time workers (age 25-54) data is also drawn from the 5% PUMS of the 2000 decennial Census. These data are cleaned in the same way as for the married female workers. This leaves us with 383,728 workers with a college degree or more and 393,598 have a high school degree or less. These workers are spread across 262 MSAs in the United States. Table 1, Panel C summarizes the adjusted hourly wage for the skilled (college degree or more) male workers and low-skilled male workers (high school degree or less).

Not surprisingly, skilled workers have higher adjusted hourly wage than the low-skilled group, both at various quantiles and also at the mean. Figure 9, Panels A and B display kernel density plots of the adjusted hourly wage for the two groups of male workers. In both panels, the aggregate adjusted wage distributions are strongly single-peaked. Splitting the samples into small and large cities (Panels C and D, respectively), it is also evident that the wage density for large cities is right-shifted and dilated as compared to the density plot for smaller cities, similar to the patterns for the married female sample. Also similar to the female worker sample, the density plots for men in Figure 9 are quite smooth, suggesting little rounding error in the data.

4. Measuring the mode

4.1 Measurement procedure

Our estimation procedure requires that we measure the modal value of the outcome variables (e.g. sale/worker, wage) in each MSA. To do so we must select a bandwidth with which to group the data. It is worth noting that if the bandwidth is too narrow, the estimated distribution will converge towards a uniform distribution with a poorly defined mode. If the bandwidth is too wide, variation across data cell groups will be so reduced that it will not be possible to discern meaningful patterns. Choice of bandwidth must balance these concerns.

To illustrate, Figures 10-12 present histograms of the data using different bandwidths. Figure 10a displays histograms of the aggregate sales per worker data using a fixed $5,000 bandwidth across all MSAs. It is evident that there is a well-defined mode in the distributions for all three groups of firms (all law firms, young law firms, and older companies). Figure 10b, Panels B and C, provide analogous histograms using alternative bandwidths. When we decrease the bandwidth to $2,500 in Panel B, three nearby cells in the center have similar height and the mode is not well defined. When we increase the bandwidth to $7,500 in Panel C, the mode is well-defined but the histograms are thick and we lose considerable variation. Figure 11a provides analogous histograms for the adjusted hourly wage for women. In this instance we impose a common fixed bandwidth of $3. Notice that there is a well-defined mode for both low- and high-skilled workers. Distribution plots in Figure 11b based on bandwidths of $1 and $5 also provide well-defined modes. Very similar patterns are obtained for men in Figures 12 and 12b using the same set of bandwidths.

Based on the plots in Figures 10-12, our ability to identify the mode for sale/worker in law firms appears to be somewhat sensitive to choice of bandwidth. This may reflect again the influence of rounding error in the data. The plots for the female and male wages, however, appear to be robust and suggest that the modal wage estimates are likely not so sensitive to the choice of bandwidth.

In choosing the bandwidths, summary measures in Table 1 provide additional guidance. In Panel A, notice that the inter-quartile range for the law firm sales per worker is roughly \$20,000 to \$25,000 for all three main samples, including all law firms, young and old. This suggests that any bandwidth larger than \$10,000 would likely not preserve enough variation in the data to yield reliable results. In Panel B, observe that for female wages, the interquartile range of adjusted wage is \$10 for the skilled married female workers and \$5 for the low-skilled workers. For men the analogous measures in Panel C are \$16 and \$7. The range of these data suggests that a bandwidth as small as just \$2 or \$3 would likely be effective for the wage measures.

For the estimation to follow, our preferred approach to specifying the bandwidths is to separately estimate modal values for each MSA for each sample using kernel density estimation with optimally chosen bandwidths based on the default bandwidth choice in Stata. The optimization criteria balances the loss of precision associated with too narrow a bandwidth against bias by minimizing the mean integrated squared error using the Epanechnikov kernel. As a robustness check, later in the paper we also report estimates for all of the models to follow based on alternate fixed bandwidths that are imposed across all MSAs. To anticipate, most of the core estimates for the law firm sale per worker models are robust but in some instances the estimates are sensitive to choice of bandwidth. In all instances the wage models are quite robust to reasonable alternative choices of bandwidth.

4.2 Summary statistics for the modal values

Based on the MSA by MSA kernel density estimates, Table 2 displays summary measures for the model values of the sale per worker and wage measures for each of our samples. Thus, each observation is a separate MSA and observation counts differ across the samples because of the sample cleaning procedures described above.

Panel A reports values for law firm sale per worker. Regardless of age of the company, the median modal value across 239 MSAs is roughly $57,000 with a standard deviation of roughly $4,000 and range of $25,000. For female wages, the median modal wage among college educated workers is $21.30 with a standard deviation of $1.77 and a range of roughly $9. For college educated men (Panel C) wages are higher by about $5 at all points in the distribution and with a larger standard deviation of $2.1. Among workers with high school or less, differences in the distribution of modal wages for women and men are reduced and the overall distribution is shifted down. The median is close to $14 for both groups although there is more variation for men with a standard deviation of $1.5 and a range of $3.75.

5. Estimates

5.1 Young and old law firms: Threshold effects

Table 3 presents MSA level regressions based on law firm sales per worker. Panel A reports results for the all firms sample, Panel B for young law firms (5 years or less in age), and Panel C for older firms (older than 5 years). For each panel, column (1) displays estimates from the first stage regression of the CDF of sale per worker evaluated at the mode on log population of the MSA. Column (2) reports estimates from the second stage regression of log sale per worker at the mode on log population of the MSA. For comparison, Column (3) reports estimates from a regression of log sale per worker at the mean for the MSA. Column (4) reports a test of the difference between the two elasticity estimates in columns 3 and 4.

Recall from the data plots presented earlier that law firm sale per worker densities are strongly single peaked. Under such conditions, Proposition 1-3 in Section 2 suggest that disproportionately selecting out establishments to the left of the mode should cause the CDF evaluated at the mode to shrink. Moreover, if threshold effects increase with MSA population, evidence of selection should increase with MSA size. Column (1) estimates in Panel A for all law firms combined confirm that prior. The coefficient on log population is -0.0162 with a t-ratio of 4.14. This indicates that doubling city size reduces the CDF evaluated at the mode by 1.6 percentage points. This is consistent with weaker firms disproportionately selecting out in larger cities.

Stratifying law firms into young and old companies yields a more nuanced pattern. In Panel B for young firms, the column (1) coefficient based on the CDF evaluated at the mode is positive 0.0036 with a t-ratio of 0.82. In addition, at the mode, doubling MSA size is associated with an increase in young law firm productivity of 3.10 percent (with a t-ratio of 11.02), while at the mean the corresponding estimate is smaller, just 1.74 percent (with a t-ratio of 6.44). That difference is also highly significant, as confirmed in column 4 where the t-ratio on the difference is -5.22. Together, these estimates are suggestive that for newly created law firms, selection increases the presence of weaker companies in larger cities relative to smaller metropolitan areas. In the Introduction we described a set of assumptions that could support this result. Specifically, if entry and exit costs are low but the rewards from operating a successful law firm increase with MSA size, weaker managers may be disproportionately tempted to open new law firms in larger metropolitan areas in the hopes of earning a large return. We recognize, however, that other processes could also be contributing to the patterns in Panel B, including for example, access to business loans which might be easier to obtain in larger cities.[21]

On the other hand, it seems unequivocal that weaker law firms will disproportionately drop out over time and that threshold costs will be higher in larger cities. For these reasons, we have a strong prior that among older law firms, selection will disproportionately cause weaker law firms to fail as MSA size increases. Panel C for older firms supports that prior. In column (1), the coefficient on the CDF evaluated at the modal sale per worker is -0.0185 with a t-ratio of -5.01. Observe also that the elasticity of sale per worker with respect to MSA size is 1.16 percent (with a t-ratio of 2.78) when evaluated based on modal returns in column (2). This is smaller than the corresponding estimate based on mean returns in column (3), 1.63 percent (with a t-ratio of 6.41). That difference is also significant at the 5 percent level based on

[21] From Proposition 3 in Section 2, it is also possible that the new law firm density function could be sufficiently flat so that the mode shifts further to the right than the mean in response to selection effects. However, the sharply single-peaked sale per worker distribution described earlier likely rules out that possibility as a primary driver of the results in Table 3.

a 1-tailed test in column (4), with a t-ratio of 1.72. Together, these estimates confirm that among older companies, weaker firms disproportionately drop out in larger cities.[22]

5.2 Married full-time working women: Threshold and migration effects

Table 4 reports results based on the married female wage data. Panel A displays estimates for skilled (college degree or more) workers while Panel B displays results for low-skilled workers. Columns 1-4 are organized in the same manner as in Table 3.

In Panel A (for skilled workers), observe that the column (1) coefficient on log MSA population is -0.012 with a t-ratio of -3.82. This indicates that as MSA size increases, selection effects disproportionally drive less productive skilled married women out of the full-time labor market. This, along with the earlier discussion of sharply single-peaked wage distributions, suggests that the return to city size measured at the mean should be upward biased and higher than the return measured at the mode. This is confirmed in columns (3) and (4). Notice that the wage elasticity with respect to city size is 2.4 percent based on modal workers and 4.3 percent when evaluated at MSA means. Both estimates are highly significant and also significantly different from each other (the t-ratio in column (4) is 5.83). These patterns indicate that selection upward biases estimates of the mean return to city size, almost doubling the estimated elasticity.

Panel B presents corresponding estimates for low-skilled (high school degree or less) married women. These estimates are different. In column (1), the coefficient on log population is -0.007 with a t-ratio of -2.08. This indicates that selection also discourages the lower end of the low-skill population form selecting into a large metro labor market, but the effect is smaller than for men. This pattern is reinforced once again by the estimates on the return to MSA size in columns 2-4. Based on the mode, the

[22] It is worth emphasizing that the point estimate in column (2) of Panel C indicates that doubling MSA size increases productivity among older law firms by 1.16 percent. This is on the lower side of many estimates reported in the literature, where recent reviews suggest that most estimates are between 2 and 5 percent (Rosenthal and Strange (2004); Combes and Gobillon (2015)). However, it is also worth emphasizing that previous studies of agglomeration economies have focused mostly on manufacturing and/or wage rates for an entire city's population (e.g. Rosenthal and Strange (2008)). We are not aware of prior estimates based on law firms.

estimated elasticity with respect to MSA size is 3.26 percent, while based on the mean, the corresponding estimate is 3.88 percent. This is a much smaller difference than for skilled workers, although still significant as indicated by the t-ratio of 2.21 in column (4).

5.3 Male full-time working men: Migration effects

As discussed earlier, because male labor supply is very inelastic, migration is likely the dominant mechanism by which selection affects wage distributions. Table 5 reports estimates for this sample. The table is organized in the same fashion as for the discussion of female workers in the previous table.

Focus first on Panel A which reports estimates for college educated men. In column (1), the impact of log population on the CDF evaluated at the mode is -0.007 with a t-ratio of -2.60. Consistent with that estimate, the wage elasticity with respect to city size is 2.1% when measured at the mode (column 2) and 4.5% when measured at the mean column (3) and highly significant in both cases. The difference in these estimates is also significant with a t-ratio of 3.94 in column (4). Together, these estimates confirm that among college educated men, unusually productive individuals tend to sort into larger MSA labor markets.

Results for low-skilled men are presented in Panel B. Findings here mirror those for married women. In column (1), there is little evidence of selection based on the CDF evaluated at the mode; the coefficient is positive -0.0020 with a t-ratio of -0.87. In the absence of systematic sorting, the return to city size measured at the mode and the mean should be similar. That prior is supported by estimates in columns 2 and 3. Notice that the estimated elasticities based on the mode and mean are 3.5% and 3.6%, respectively while the t-ratio on that difference in column (4) is 0.0010.

Summarizing, estimates from Tables 3-5 all support the view that selection contributes to productivity in larger metropolitan areas. Selection effects are especially prevalent among older law firms and college educated women and men. For these latter samples, and for both women and men, doubling MSA size increases wage at the mode by about 2.25%, roughly 2 percentage points lower than estimates based on the mean return. Thus, selection nearly doubles the perceived return to MSA size. For low-skilled workers, in contrast, selection is mostly absent and for these workers, doubling MSA size increases wage by roughly 3.25% regardless of gender.

5.4 Robustness checks: Alternate fixed bandwidths when measuring modal values

This section presents a series of robustness checks that consider the sensitivity of our results to alternative measures of the bandwidths used to measure modal values. Two sets of checks are conducted. The first is quickly summarized. We re-ran all of the models in Tables 3-5 using one-half the optimal bandwidth from the kernel density routine. Results were quite close to estimates presented in Tables 3-5 and are not reported for that reason.

Our second set of robustness checks uses a different, arguably less reliable, procedure for selecting the bandwidths used to measure modal values. In this instance, for each sample in Tables 3-5, we imposed a common fixed bandwidth across all MSAs and then re-estimated the corresponding models. We did this many times for different fixed bandwidths and then plotted the resulting estimates in Figures 13-15 (for law establishments, female wages and male wages, respectively) with the alternative bandwidths reported along the horizontal axes. For the wage models for both women and men, plots are based on fixed common bandwidths that range from $2 to $4 in $0.20 increments.

We begin with the law firm sale per worker estimates in Figures 13a and 13b. In all cases, coefficient plots are reported based on fixed common bandwidths that range from $4,000 to $6,000 in $200 increments. Figure 13a reports estimates for the regression of the CDF at the mode while Figure 13b reports estimates of the sale per worker elasticity with respect to MSA size based on modal values. In both cases the point estimates are displayed as heavy solid lines while the 95% confidence bands are light dashed lines. To facilitate viewing, in Figure 13a a heavy dashed horizontal line marks the zero axis while in Figure 13b an analogous line marks the estimated elasticity based on mean values as reported in Table 3. Both figures include three panels for all, young and older law firms, respectively.

In Figure 13a, in all three panels, the coefficient estimates remain largely stable when we vary the bandwidth from $4,000 to $5,500. Within that range, the coefficient estimates for the all- and old-firm samples are less than zero, while the coefficient estimates for young firms are always positive. The 95% confidence bands (indicated by the dashed lines in the figures) also indicate that the point estimates are significantly different from zero. For bandwidths beyond $5,500, however, estimates become unstable in all three panels. This latter result is likely because the increasingly thick bandwidths eliminate too much of the variation in the sample, making identification difficult.

Figure 13b presents analogous plots for the elasticity of the return to city size based on the mode. For these estimates, and for each of the samples (all, young and older law firms), estimates based on the modal law firm are sensitive to bandwidth. Notice, for example, that the elasticity differs by up to three percentage points as we vary the bandwidth.

Figure 14a plots coefficient estimates for female wage rates based on the CDF evaluated at the mode with bandwidths ranging from $2 to $4 in $0.20 increments. In Panel A, for skilled workers, the coefficient estimate on log population is negative and varies little with the different bandwidths. In Panel B, for low-skilled workers, the pattern is also consistent with the results in Table 3. In this instance, the coefficient estimates are mostly not significantly different from zero while also displaying more sensitivity to choice of bandwidth.[23] Figure 14b presents coefficient plots for the second stage regression estimates of the return to MSA size. For both skilled (Panel A) and low-skilled (Panel B) workers, and different from the law firm sample, the estimated elasticities display little variation with the different bandwidths and are close to the values reported in Table 3.

Figure 15a plots estimates of the MSA size coefficient for the CDF regression (column (1) in Table 5) using bandwidths from $2 to $4. For both skilled and low-skilled workers the patterns are very robust. For skilled workers (Panel A), the coefficient is always significantly negative and similar in value for the different bandwidths. For low-skilled workers, the coefficient estimates also vary little with the bandwidths and are generally very small slightly below zero. These patterns are reinforced in Figure 15b

[23] Although there is a visible drop in the coefficient estimate using around the bandwidth of $3.5, it is likely to be caused by limited variation in the modal wage estimates. Recall from Table 1 Panel B, the majority of the adjusted hourly wage for the low-skilled workers fall within the range of $12 to $18 so that larger bandwidths greatly reduce variation in the data.

which plots alternate estimates of the wage elasticity with respect to MSA size. For both the skilled and low-skilled workers, estimates based on the mode are robust to bandwidth choice, displaying little variation. For low-skilled workers, the return to MSA size at the mode is also always very close to the corresponding estimate based on the mean, sometimes slightly above and sometimes slightly below.

Summarizing, the estimates reported in this section suggest that our core estimates are quite robust to reasonable alternative choices of the bandwidth. The primary exception is when estimating the influence of MSA size on sale per worker among law firms.

6. Conclusion

This paper develops a new method to identify and control for selection when estimating the productivity effects of city size. For single peaked factor return distributions, selecting out low-performing agents will often have little or no effect on modal productivity while reducing the CDF evaluated at the mode. Spillovers from agglomeration have the reverse effect. We show that these patterns hold regardless of whether selection arises from the decision to participate or location choice. Formal conditions under which our arguments hold are developed and motivate two core regressions. The first regresses the CDF evaluated at the mode of the factor return distribution on city size. This reveals whether selection occurs disproportionately to the left or right of the mode. The second regresses the log factor return (e.g. wage) on city size. This along with estimates based on the mean provides corroborating evidence on the nature of selection while yielding estimates of the return to city size that are largely robust to selection.

We estimate our model using three different data sets, each of which highlights different features of the approach. The first includes establishment-level data for newly formed and older law firms using sale per work as a measure of productivity. The second uses wages for full-time working married women who are age 25-55, white, non-Hispanic and native born. The third uses wages for full-time working men who are also age 25-55, white, non-Hispanic, and native born. Results from all three exercises yield compelling evidence that selection contributes to urban productivity. Also evident is that selection is especially apparent among skilled individuals (college plus workers and law firms) but largely absent among low-skilled workers (high school or less).

Based on the mode, we estimate that doubling MSA size increases male labor productivity by roughly 2.1% for skilled workers and 3.5% for low-skilled. Similar estimates are obtained for married female wages. For both samples, ignoring selection by estimating at the mean nearly doubles the perceived return to MSA size for skilled workers. Among law firms, doubling MSA size increases productivity of the modal law firm by roughly 1.5% with larger effects among newly originated firms and compelling evidence that selection among older companies causes weaker firms to drop out over time.

Our approach based on the mode can be applied to other contexts where selection is important and the underlying latent density being modeled has a well-defined single-peaked shape. This could include instances where the goal is to assess whether discrimination is present, a source of selection. It could also include settings in which the selection process is largely known and the focus is on identifying causal effects as with the influence of larger or more elite schools on student performance.

Appendix A: Proof of Proposition 3

This appendix shows how the result in Proposition 3 is obtained. Recall from (2.8) that the density function for the generalized error distribution is given by

$$f_0(y) = \frac{1}{2^{\kappa+1}\sigma\Gamma(\kappa+1)} e^{-\kappa\left|\frac{y_{m,1}-y_m}{\sigma}\right|^{\frac{1}{\kappa}}} \tag{A.1}$$

where the notation is defined as in the text. Notice also that the density function is evaluated at the normalized value for the conditional mode, $y_{m,1}$, having subtracted off the unconditional mode y_m and divided by the scale parameter.

Taking the derivative of f_0 with respect to the normalized value for y,

$$f_0{}'(y) = -\, f_0(y)\left|\frac{y_{m,1}-y_m}{\sigma}\right|^{\frac{1}{\kappa}-1} \tag{A.2}$$

Recall also from (2.6) that the mode of the conditional density function must satisfy

$$\frac{\pi_1{}'\left(\frac{y_{m,1}-y_m}{\sigma}\right)}{\pi_1\left(\frac{y_{m,1}-y_m}{\sigma}\right)} = -\frac{f_0{}'\left(\frac{y_{m,1}-y_m}{\sigma}\right)}{f_0\left(\frac{y_{m,1}-y_m}{\sigma}\right)} \tag{A.3}$$

Substituting into (A.3),

$$\frac{\pi_1{}'\left(\frac{y_{m,1}-y_m}{\sigma}\right)}{\pi_1\left(\frac{y_{m,1}-y_m}{\sigma}\right)} = \left|\frac{y_{m,1}-y_m}{\sigma}\right|^{\frac{1}{\kappa}-1} \tag{A.4}$$

Next we multiply both sides of (A.4) by the normalized conditional mode which yields an expression for the elasticity of the selection probability,

$$\left[\frac{y_{m,1}-y_m}{\sigma}\right]^{1/k} = \left[\frac{\pi_1{}'\left(\frac{y_{m,1}-y_m}{\sigma}\right)}{\pi_1\left(\frac{y_{m,1}-y_m}{\sigma}\right)}\right]\frac{y_{m,1}-y_m}{\sigma} \equiv \xi_{\pi,y} \tag{A.5}$$

From (A.5), the change in the mode in response to selection is given by,

$$y_{m,1} - y_m = \sigma\xi_{\pi,y}^{k} \;. \tag{A.6}$$

References:

Ahlfeldt, G., S.J. Redding, D.M. Sturm, and N. Wolf (2015). The Economics of Density: Evidence from the Berlin Wall. *Econometrica* 83(6), 2127-2189.

An, Mark Yuying (1996). Log-Concave Probability Distributions: Theory and Statistical Testing. Duke University Department of Economics, Working Paper No. 95-03: 9-10

Behrens, K., Duranton, G., & Robert-Nicoud, F. (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, 122(3), 507-553.

Behrens, K., and F. Robert-Nicoud (2015). Agglomeration Theory with Heterogeneous Agents. G. Duranton, J. V. Henderson and W. Strange (eds), *Handbook in Regional and Urban Economics*, Amsterdam (Holland), Elsevier Press.

Black, D. A., Kolesnikova, N., & Taylor, L. J. (2014). Why do so few women work in New York (and so many in Minneapolis)? Labor supply of married women across US cities. *Journal of Urban Economics*, 79, 59-71.

Blau, F. D., & Kahn, L. M. (2007). Changes in the labor supply behavior of married women: 1980–2000. *Journal of Labor Economics*, 25(3), 393-438.

Baum-Snow, N., & Pavan, R. (2013). Inequality and city size. *Review of Economics and Statistics*, 95(5), 1535-1548.

Baum-Snow, N., & Neal, D. (2009). Mismeasurement of usual hours worked in the census and ACS. *Economics Letters*, 102(1), 39-41.

Bosquet, Clement and Pierre-Philippe Combes (2017). Sorting and agglomeration economies in French economics departments. *Journal of Urban Economics*, 101, 27-44.

Cardoso, A. R., & Portugal, P. (2005). Contractual wages and the wage cushion under different bargaining settings. *Journal of Labor economics*, 23(4), 875-902.

Chamberlain, G. (1986). Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics*, 32(2), 189-218.

Chen, Y. C., Genovese, C. R., Tibshirani, R. J., & Wasserman, L. (2016). Nonparametric modal regression. *The Annals of Statistics*, 44(2), 489-514.

Chotikapanich, Duangkamon, Rebecca Valenzuela, and DS Prasada Rao (1997). "Global and regional inequality in the distribution of income: estimation with limited and incomplete data." *Empirical Economics* 22.4: 533-546.

Clementi, Fabio, and Mauro Gallegati (2005). "Pareto's law of income distribution: Evidence for Germany, the United Kingdom, and the United States." *Econophysics of wealth distributions*. Springer, Milano, 3-14.

Combes, P.P., G. Duranton, and L.Gobillon (2008). Spatial Wage Disparities: Sorting Matters!. *Journal of Urban Economics* 63(2), 723-742.

Combes, Pierre-Philippe, et al. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica* 80.6: 2543-2594.

Combes, P.P., and L. Gobillon (2015). The Empirics of Agglomeration Economies. *in* G. Duranton, J. V. Henderson and W. Strange (eds), *Handbook in Regional and Urban Economics*, Volume 5, Amsterdam (Holland), Elsevier Press.

Costa, Dora L. and Matthew E. Kahn (2000), "Power Couples: Changes in the Locational Choice of the College Educated, 1940-1990," *Quarterly Journal of Economics*, Volume CXV, 1287-1315.

De la Roca, Jorge (2017). Selection in initial and return migration: Evidence from moves across Spanish cities. *Journal of Urban Economics*, 100, 33-53.

De la Roca, Jorge, and Diego Puga (2017). Learning by Working in Big Cities. *The Review of Economic Studies,* 84.1: 106-142.

Drennan, M. P., and H.F. Kelly (2012). Measuring Urban Agglomeration Economies with Office Rents. *Journal of Economic Geography* 12(3),481-507.

Duranton, Gilles, and Diego Puga (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*: 1454-1477.

Gaubert, Cecile. (forthcoming) Firm sorting and agglomeration. *American Economic Review*.

Glaeser, Edward L., and David C. Mare. (2001) Cities and skills. *Journal of labor economics* 19.2: 316-342.

Heim, B. T. (2007). The incredible shrinking elasticities married female labor supply, 1978–2002. *Journal of Human resources*, 42(4), 881-918.

Huang, M., Li, R., & Wang, S. (2013). Nonparametric mixture of regression models. *Journal of the American Statistical Association*, 108(503), 929-941.

Kemp, G. C., & Silva, J. S. (2012). Regression towards the mode. *Journal of Econometrics*, 170(1), 92-101.

Lee, Myoung-jae (1989). Mode regression. *Journal of Econometrics* 42.1-3: 337-349.

Lee, M. J. (1993). Quadratic mode regression. *Journal of Econometrics*, 57(1-3), 1-19.

Lewbel, A. (2007). Endogenous selection or treatment model estimation. *Journal of Econometrics*, 141(2), 777-806.

Lopez, J. Humberto, and Luis Servén (2006). *A Normal Relationship?: Poverty, Growth, and Inequality*. Vol. 3814. World Bank Publications.

Ruggles, Steve, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. *Integrated Public Use Microdata Series: Version 6.0* [dataset]. Minneapolis: University of Minnesota, 2015.

Roy, Andrew Donald (1951). Some thoughts on the distribution of earnings. Oxford economic papers 3.2: 135-146.

Rosenthal, S.S and W.C. Strange (2004). Evidence on the Nature and Sources of Agglomeration Economies. in Henderson, J.V. and Thisse, J.-F. (Eds.), *Handbook of Urban and Regional Economics, Volume 4*, Amsterdam: Elsevier, 2129-2172.

Rosenthal, S. S., and W. C. Strange (2008). The Attenuation of Human Capital Spillovers. *Journal of Urban Economics* 64(2), 373-389.

Rosenthal, Stuart S., and William C. Strange (2012). Female Entrepreneurship, Agglomeration, and a New Spatial Mismatch. *Review of Economics and Statistics*, 94(3), 764-788.

Saumard, A., & Wellner, J. A. (2014). Log-concavity and strong log-concavity: a review. *Statistics Surveys*, *8*, 45: 58-59.

Sala-i-Martin, Xavier, and Maxim Pinkovsky (2009). "Parametric estimations of the world distribution of income." *NBER Working Paper* number 15433.

Yao, W., & Li, L. (2014). A new regression model: modal linear regression. *Scandinavian Journal of Statistics*, 41(3), 656-67

Table 1: Summary statistics of the individual-level data [a]

Panel A
Adjusted sale per worker for law establishments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	5th quantile	25th quantile	50th quantile	75th quantile	95th quantile	mean	Observation
All Firms	46,666	59,000	67,609	82,222	121,446	72,639	545,873
Young Firms (<= 5 years)	38,422	52,287	60,000	70,000	96,531	62,735	74,079
Old Firms (> 5 years)	48,255	60,000	70,000	84,261	123,541	74,194	471,794

Panel B
Adjusted wage for married non-Hispanic native-born white female full-time workers, age 25-54

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	5th quantile	25th quantile	50th quantile	75th quantile	95th quantile	mean	Observation
College degree or more	10.27	17.85	22.66	28.09	41.95	24.08	152,704
High school degree or less	9.01	12.51	15.01	18.10	25.17	15.75	153,168

Panel C
Adjusted wage summary statistics for male non-Hispanic native-born white full-time workers, age 25-54

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	5th quantile	25th quantile	50th quantile	75th quantile	95th quantile	mean	Observation
College degree or more	6.59	21.17	29.08	37.52	81.18	32.49	383,728
High school degree or less	6.83	12.74	15.38	19.75	29.58	16.40	393,598

[a] Law firm data are from Dun and Bradstreet for December 2016. The sample is restricted to single-site firms which excludes roughly 2 percent of establishments. MSAs are restricted to those with 100,000 or more population that have at least 30 or more law firms present for both young and old classifications of law firms.

[b] Married female individual-level data are obtained from the 2000 Census. Hourly wage is adjusted by controlling for age, education, occupation and industry fixed effects. The sample is restricted to cities with at least 100,000 or more population that have at least 100 or more observation in each education category.

Table 2: Summary statistics of the mode estimates across MSAs[a]

Panel A: Modal sales per worker for law establishments[b]

	(1)	(2)	(3)	(4)	(5)	(6)
	Min	Max	Median	Mean	Std.	Observations
All Firms	48,000	69,772	57,500	58,640	4,080	239
Young Firms (<= 5 years)	45,262	64,351	56,090	55,580	4,059	239
Old Firms (> 5 years)	47,202	74,645	57,667	59,287	4,665	239

Panel B: Modal adjusted female wage rates[c,d]

	(1)	(2)	(3)	(4)	(5)	(6)
	Min	Max	Median	Mean	Std.	Observations
College degree or more	17.46	26.74	21.30	21.40	1.77	216
High school degree or less	11.39	17.79	13.92	13.98	1.15	216

Panel C: Modal adjusted male wage rates[c,e]

	(1)	(2)	(3)	(4)	(5)	(6)
	Min	Max	Median	Mean	Std.	Observation
College degree or more	21.54	33.78	27.66	27.63	2.13	262
High school degree or less	10.08	18.35	13.83	13.85	1.53	262

[a] All modal estimates are obtained using the default optimal bandwidth in Stata.

[b] Law firm data are from Dun and Bradstreet for December 2016. The sample is restricted to single-site firms which excludes roughly 2 percent of establishments. MSAs are restricted to those with 100,000 or more population that have at least 30 or more law firms present for both young and old classifications of law firms.

[c] Individual-level data are obtained from the 2000 Census. Wage is adjusted by controlling for occupation, industry, age and education fixed effects. The sample is restricted to cities with at least 100,000 or more population that have at least 100 or more observation in each education category.

[d] Observations are married white non-Hispanic native-born full-time female workers, aged 25-54.

[e] Observations are white non-Hispanic native-born full-time male workers, age 25-54.

Table 3: Law establishments

Panel A: All Firms	(1) CDF of Sale/Worker evaluated at the Mode	(2) Log(Sale/Worker) at the Mode	(3) Log(Sale/Worker) at the Mean	(4) Coefficient difference (3) - (2)
Log population in MSA	-0.0162	0.0151	0.0169	0.0017
	(-4.99)	(4.14)	(7.23)	(0.75)
R-squared	0.115	0.070	0.164	0.002
Observations	239	239	239	239
Panel B: Young Firms (<= 5 years)				
Log population in MSA	0.0036	0.0310	0.0174	-0.0136
	(0.82)	(11.02)	(6.44)	(-5.22)
R-squared	0.002	0.304	0.003	0.076
Observations	239	239	239	239
Panel C: Old Firms (> 5 years)				
Log population in MSA	-0.0185	0.0116	0.0163	0.0047
	(-5.01)	(2.78)	(6.41)	(1.72)
R-squared	0.110	0.033	0.132	0.013
Observations	239	239	239	239

[a] T-ratios based on robust standard errors in parentheses.
[b] Data are from Dun and Bradstreet for December 2016. The sample is restricted to single-site firms which excludes roughly 2 percent of establishments. MSAs are restricted to those with 100,000 or more population that have at least 30 or more law firms present for both young and old classifications of law firms.
[c] The modes are estimated using default optimal bandwidth in Stata.

Table 4: Female married white non-Hispanic native-born full-time workers, age 25-54 [a]

Panel A: College degree or more	(1) CDF of wage evaluated at the Mode	(2) Log(wage) at the Mode	(3) Log(wage) at the Mean	(4) Coefficient difference (3) - (2)
Log population in MSA	-0.0121	0.0241	0.0428	0.0187
	(-3.82)	(5.09)	(9.98)	(5.83)
R-squared	0.044	0.091	0.312	0.099
Observations	216	216	216	216
Panel B: High school degree or less				
Log population in MSA	-0.0071	0.0326	0.0388	0.0062
	(-2.08)	(6.89)	(10.03)	(2.21)
R-squared	0.019	0.291	0.171	0.021
Observations	216	216	216	216

[a] T-ratios based on robust standard errors in parentheses.
[b] Married female worker data are obtained from the 2000 Census. Wage is adjusted by controlling for occupation, industry, age and education fixed effects. The sample is restricted to cities with at least 100,000 or more population that have at least 100 or more observation in each education category.
[c] The modes are estimated using default optimal bandwidth in Stata.

Table 5: Male white non-Hispanic native-born full-time workers, age 25-5 [a]

Panel A: College degree or more	(1) CDF of wage evaluated at the Mode	(2) Log(wage) at the Mode	(3) Log(wage) at the Mean	(4) Coefficient difference (3) - (2)
Log population in MSA	-0.0069	0.0211	0.0452	0.0203
	(-2.60)	(5.26)	(11.58)	(3.94)
R-squared	0.021	0.082	0.352	0.048
Observations	262	262	262	262
Panel B: High school degree or less				
Log population in MSA	-0.0020	0.0352	0.0361	0.0010
	(-0.87)	(5.94)	(7.46)	(0.32)
R-squared	0.002	0.112	0.173	0.000
Observations	262	262	262	262

[a] T-ratios based on robust standard errors in parentheses.
[b] Male worker data are obtained from the 2000 Census. Wage is adjusted by controlling for occupation, industry, age and education fixed effects. The sample is restricted to cities with at least 100,000 or more population that have at least 100 or more observation in each education category.
[c] The modes are estimated using default optimal bandwidth in Stata.

Figure 1: Productivity Distributions with Agglomeration Economies

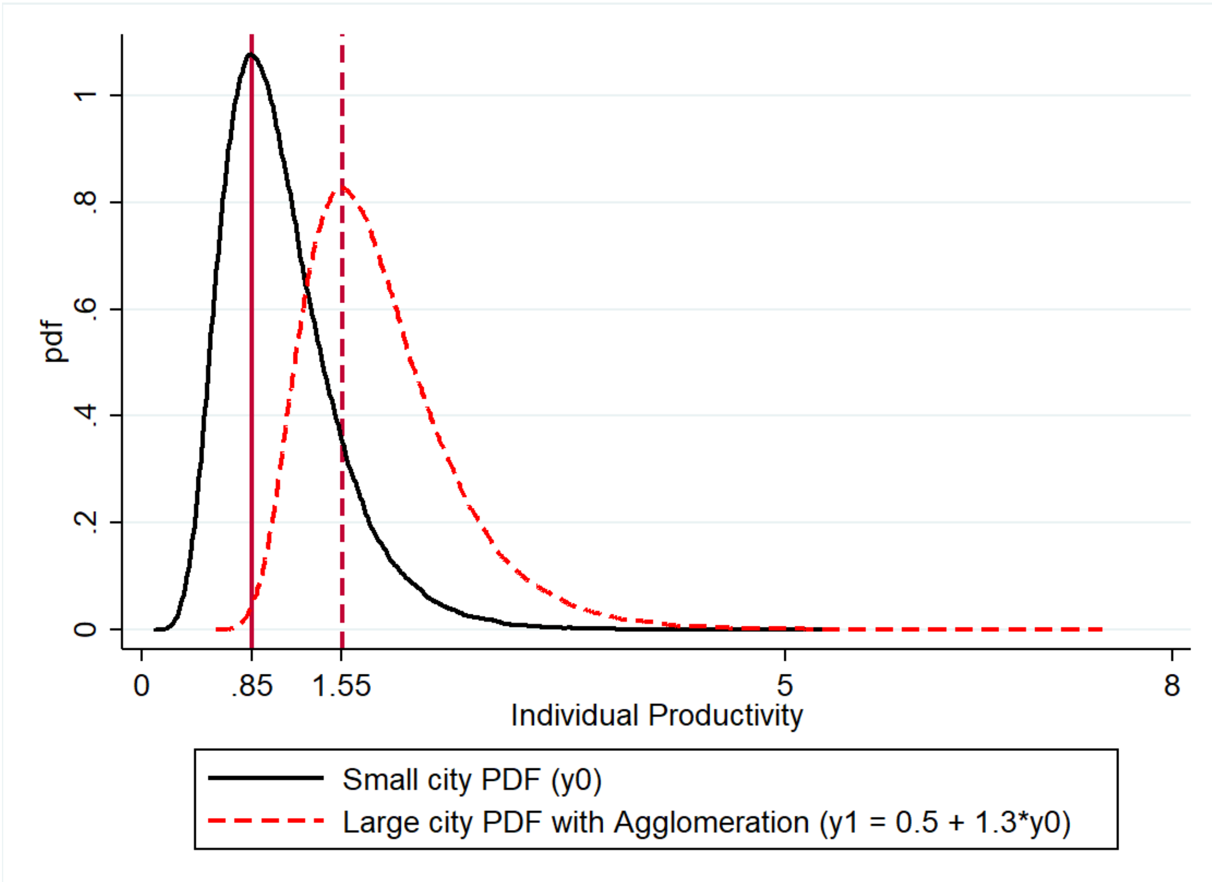
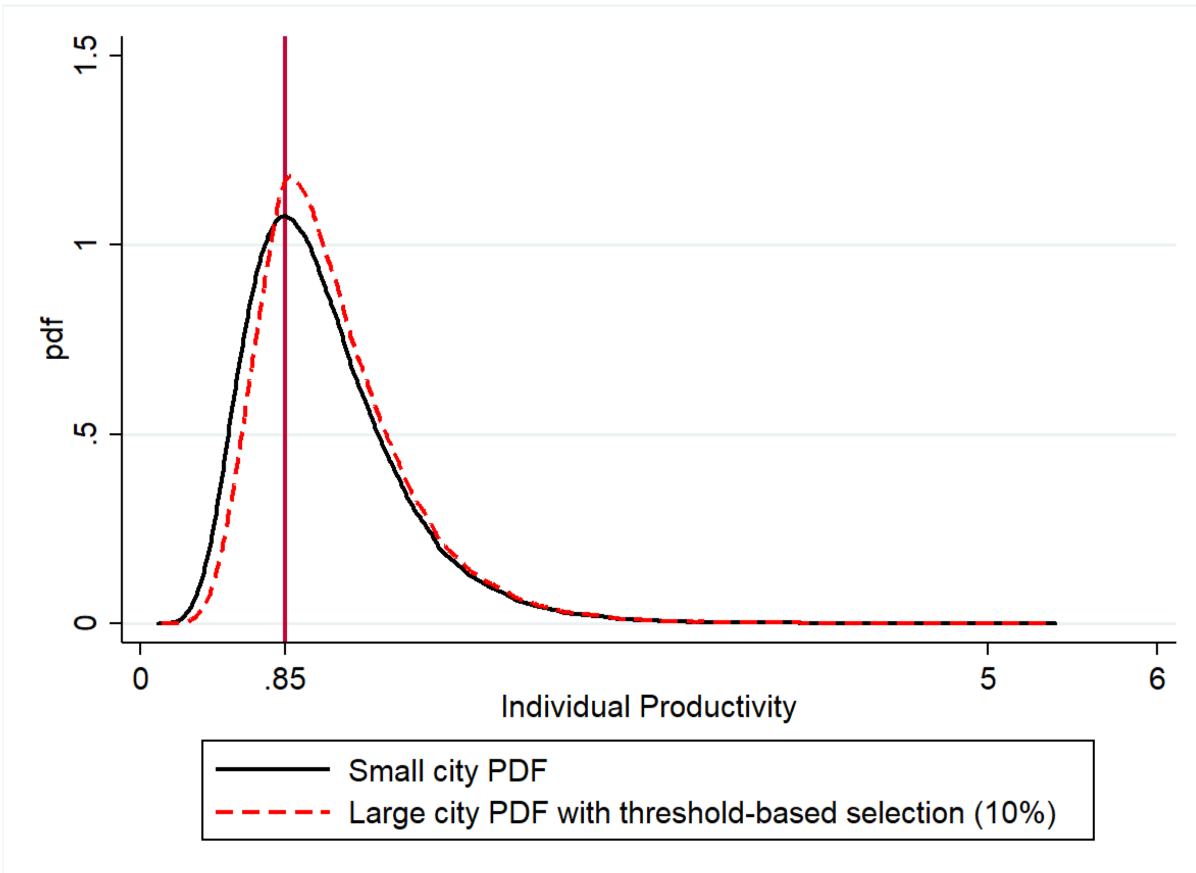


Figure 2: Productivity Distributions with Threshold Effects



Note: This figure illustrates a case where 10% of the workers are selected out of the large city's labor market.

Figure 3: Productivity Distributions with Agglomeration Economies and Threshold Effects

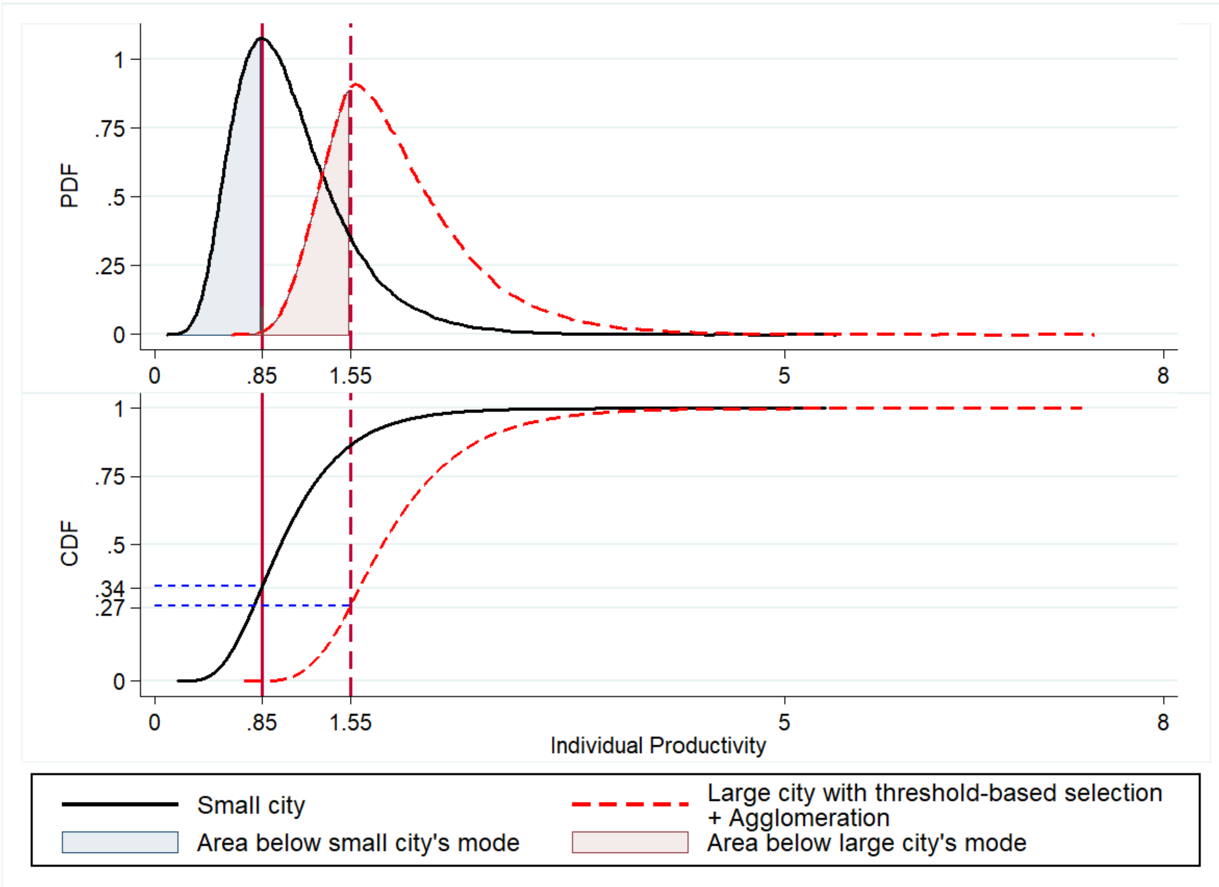
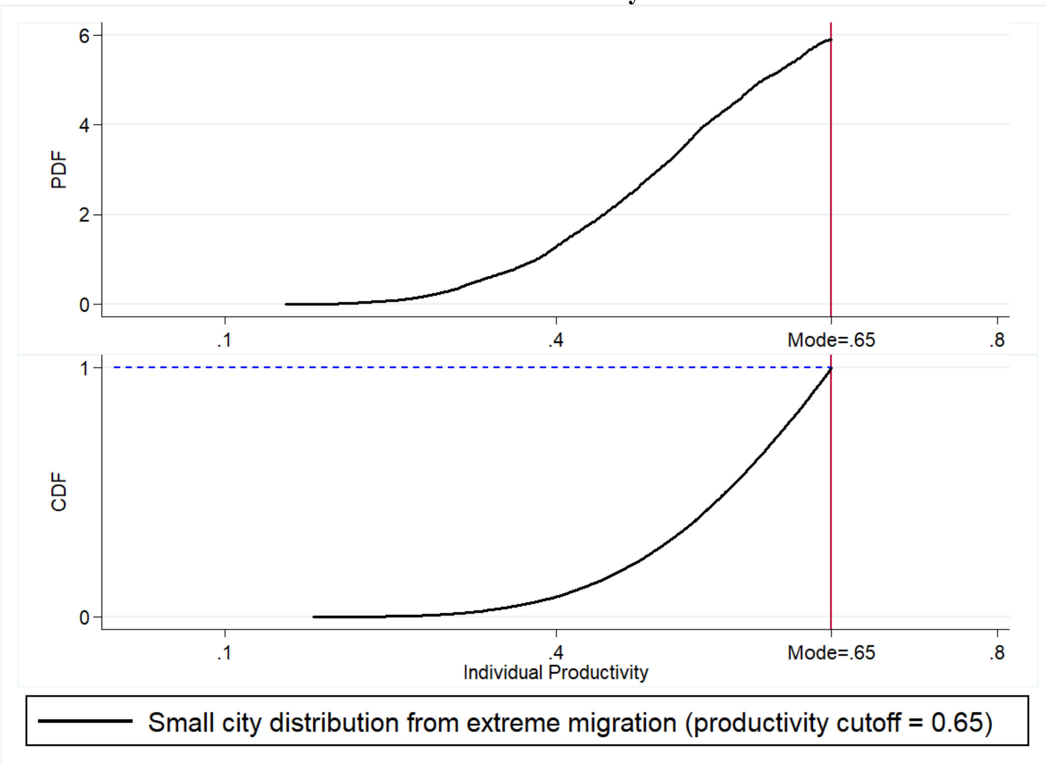


Figure 4a: Small and Large City Productivity Distributions from Extreme Migration Effects (Original Mode=0.85 and Migration Productivity Cutoff = 0.65)

Panel A: Small City



Panel B: Large City

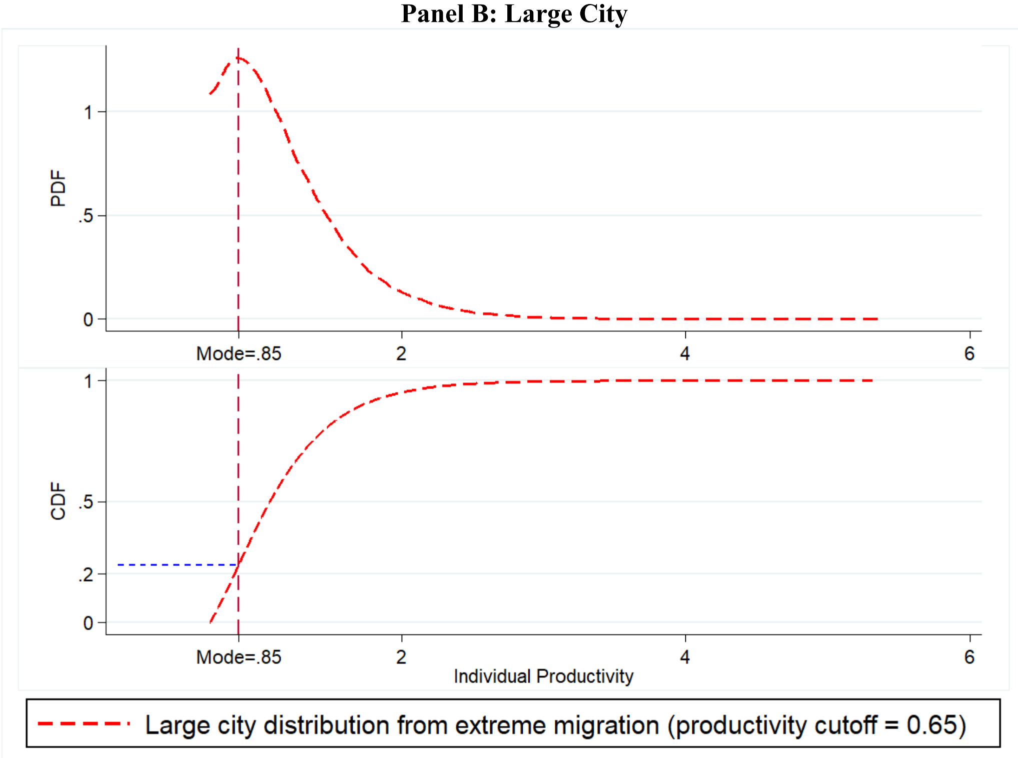
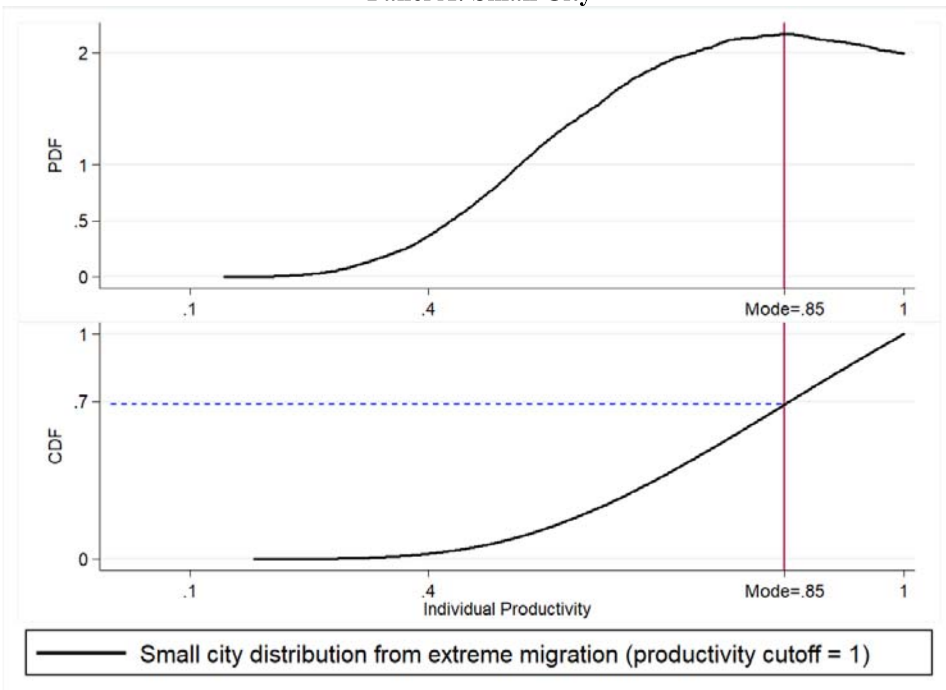


Figure 4b: Small and Large City Productivity Distributions from Extreme Migration Effects (Original Mode=0.85 and Migration Productivity Cutoff = 1)

Panel A: Small City



Panel B: Large City

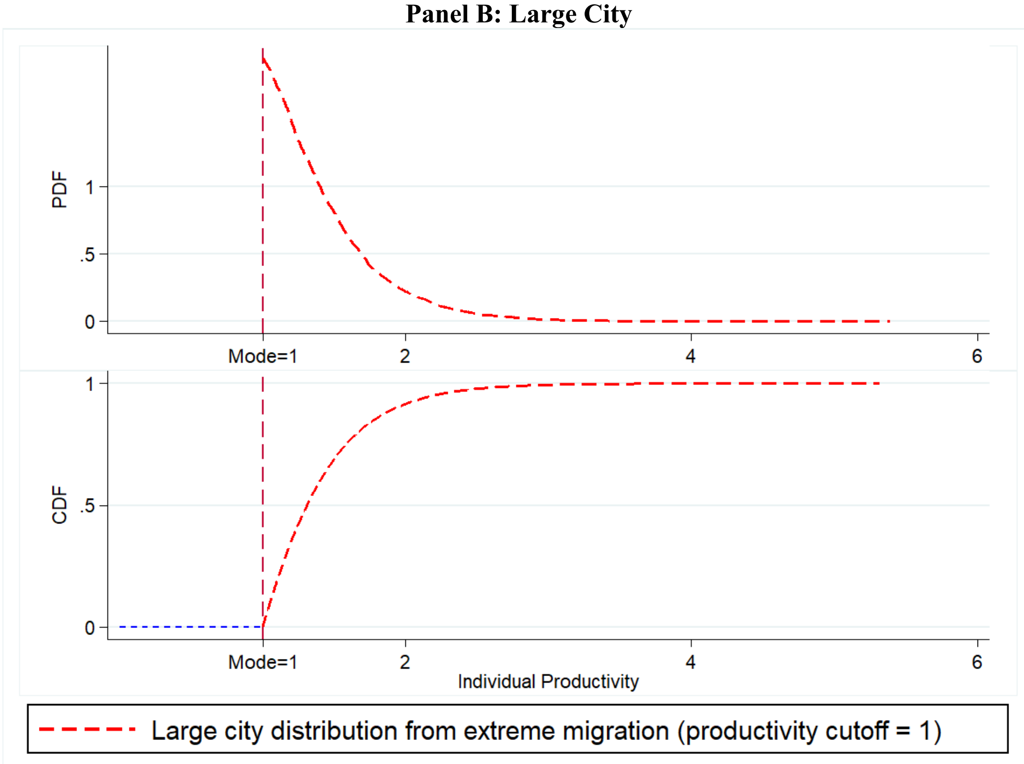


Figure 5a: Productivity Distributions with Migration Effects

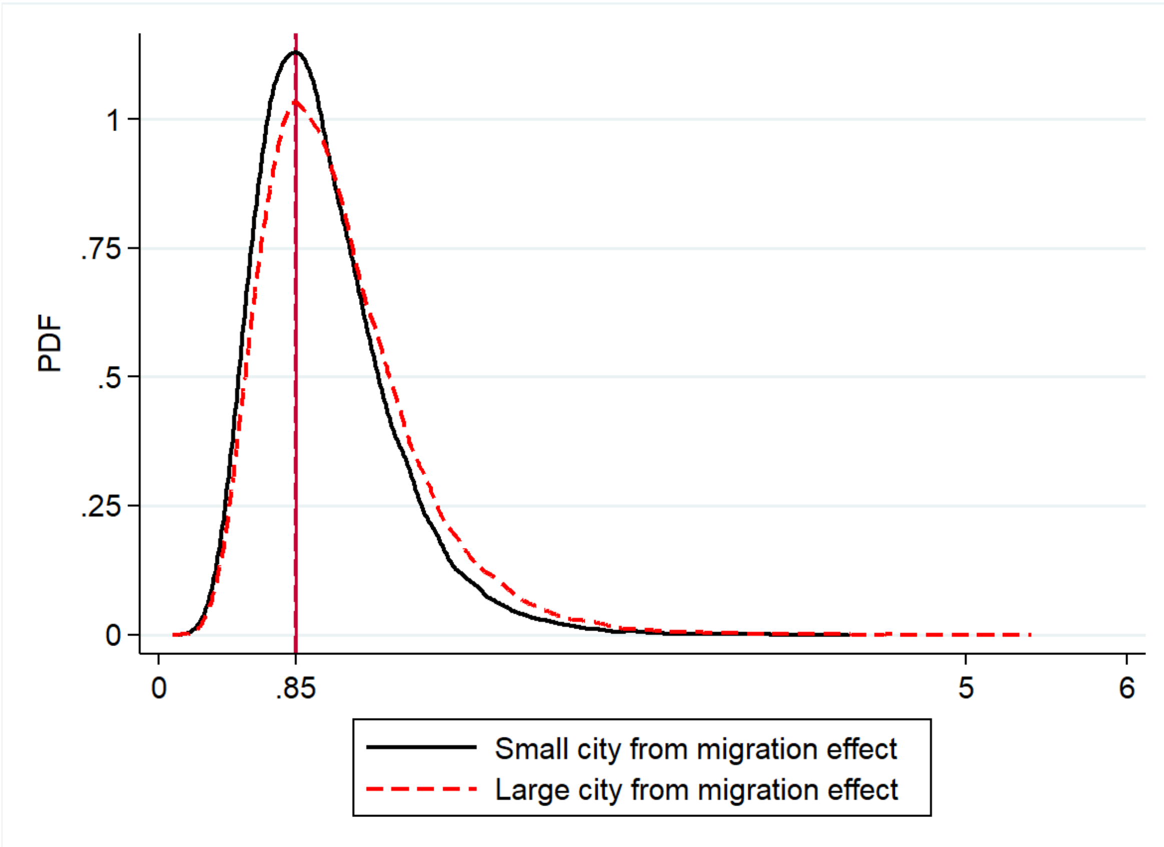


Figure 5b: Productivity Distributions with Agglomeration Economies and Migration Effects

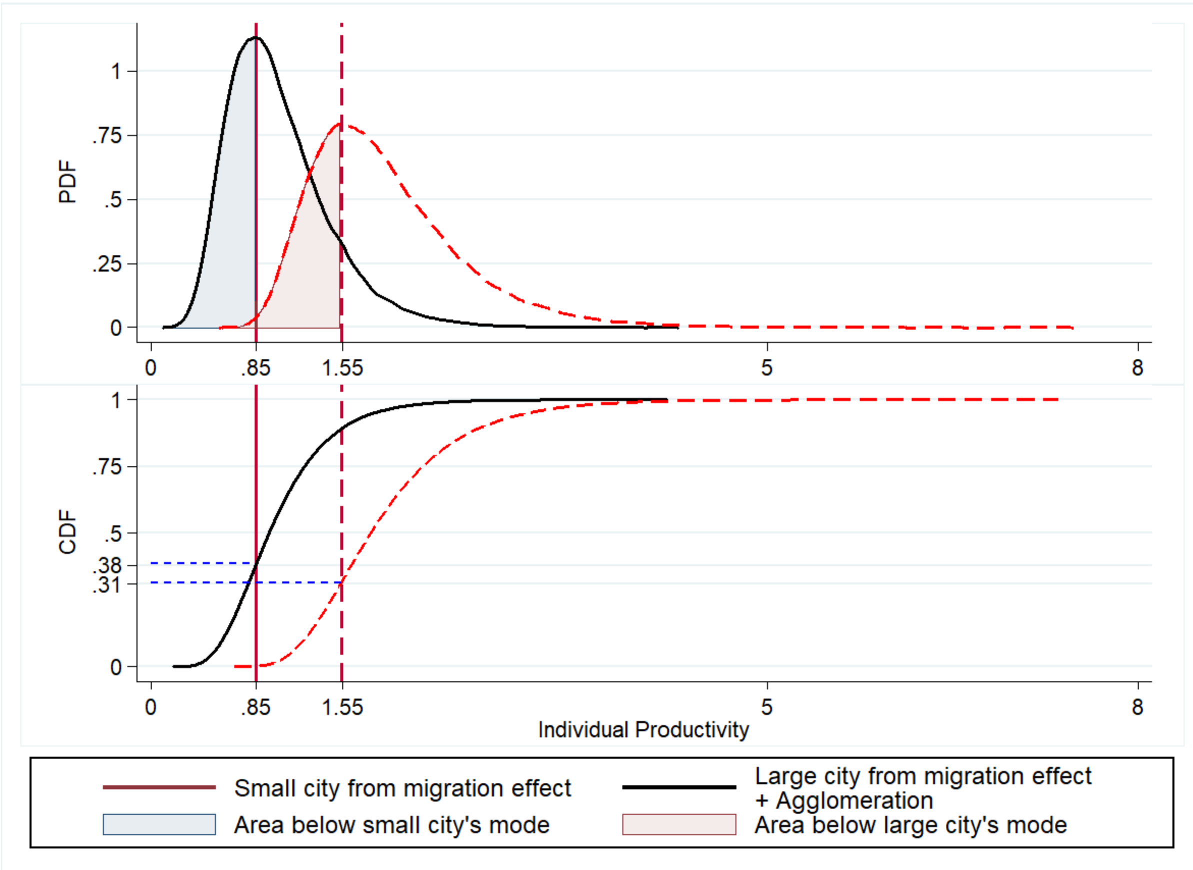


Figure 6: Slope Conditions and Shifts in the Mode

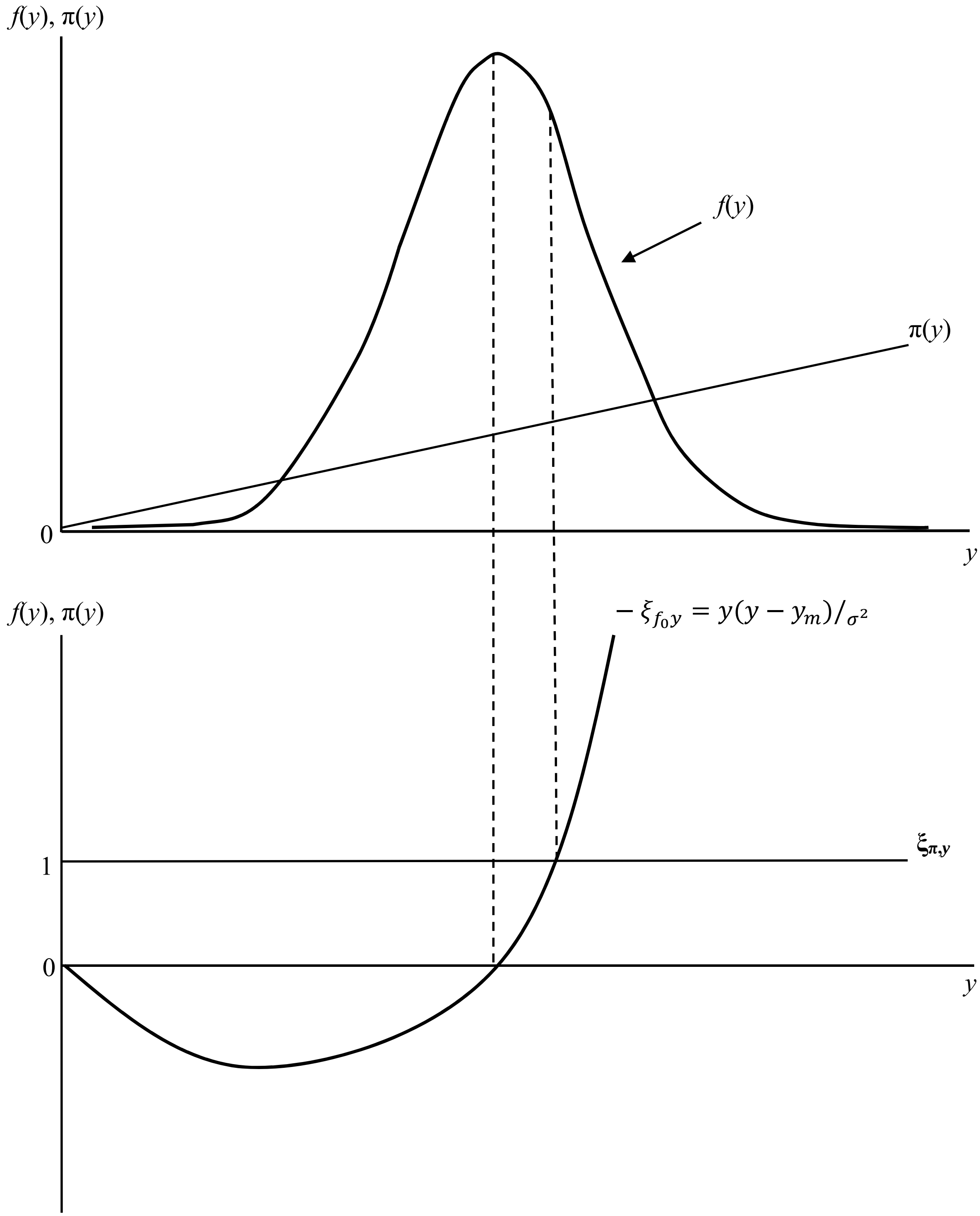
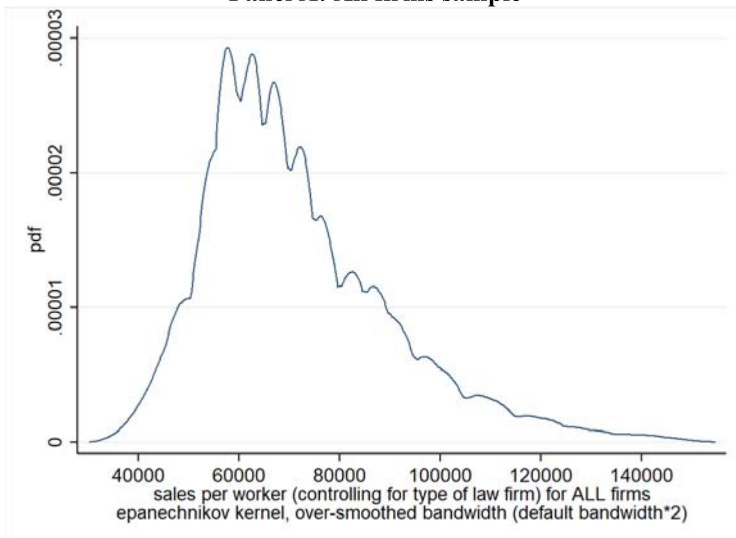
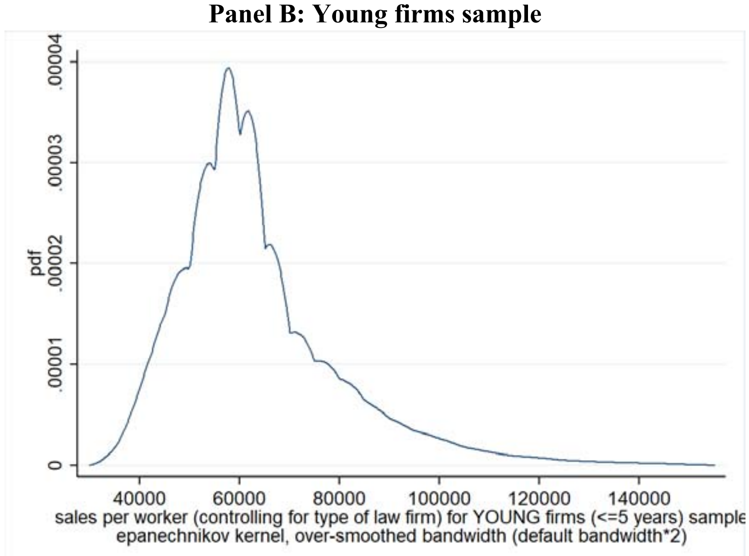


Figure 7a: Sale per worker kernel density estimation for law firms in the Unites States

Panel A: All firms sample



Panel B: Young firms sample



Panel C: Old firm sample

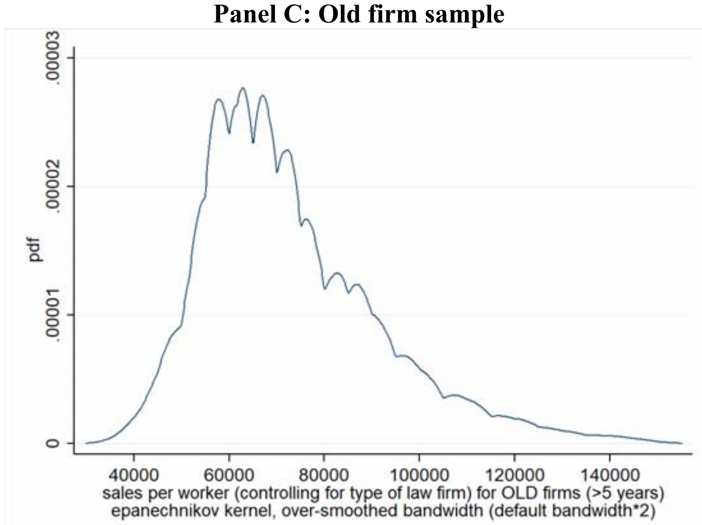
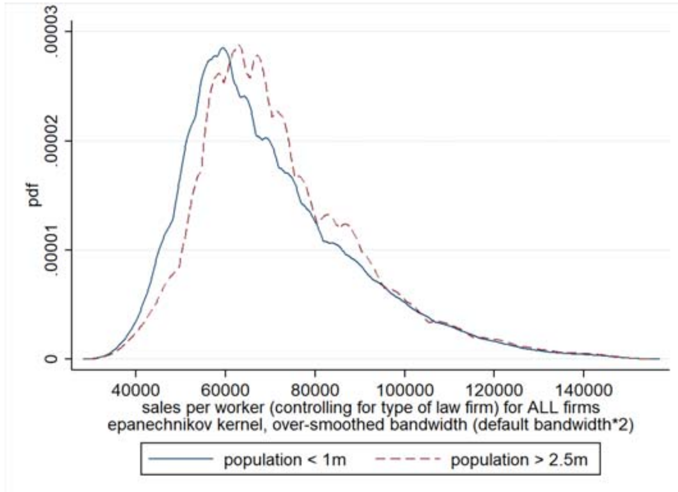
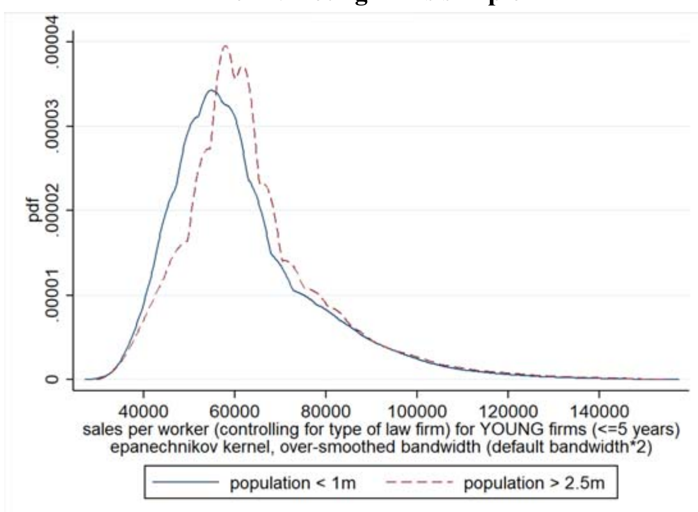


Figure 7b: Sale per worker kernel density estimation for law firms in small versus large cities

Panel A: All firms sample



Panel B: Young firms sample



Panel C: Old firm sample

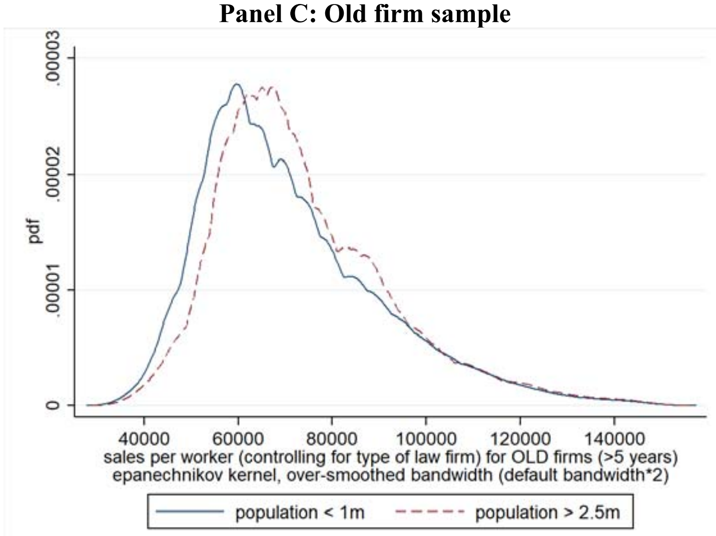
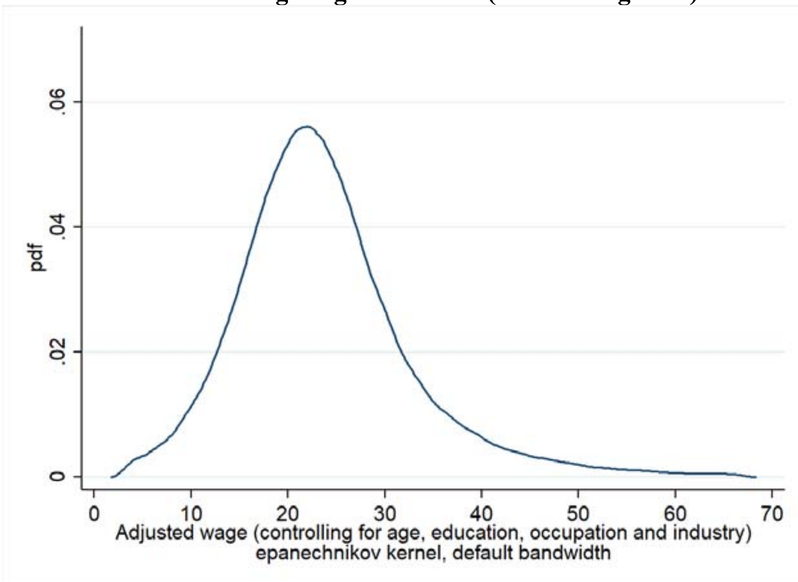


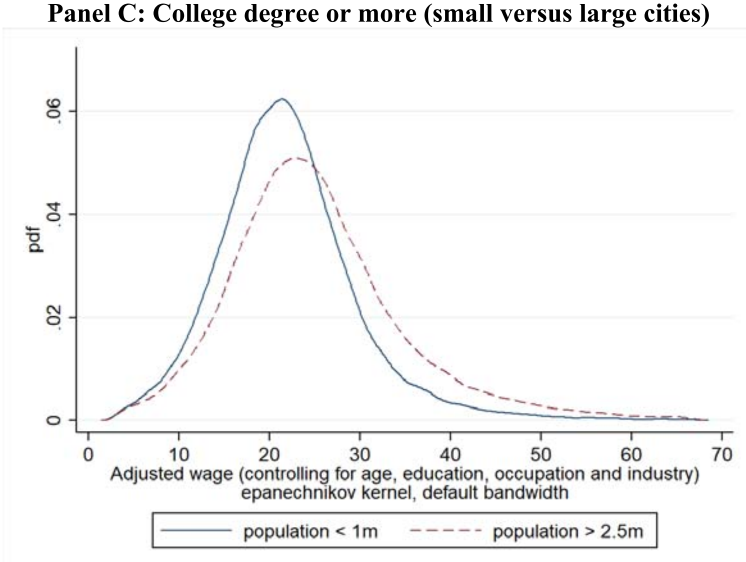
Figure 8: Adjusted wage kernel density estimation for married female non-Hispanic native-born white full-time workers, age 25-54

Panel A: College degree or more (all cities together)



Panel B: High school degree or less (all cities together)

Panel C: College degree or more (small versus large cities)



Panel D: High school degree or less (small versus large cities)

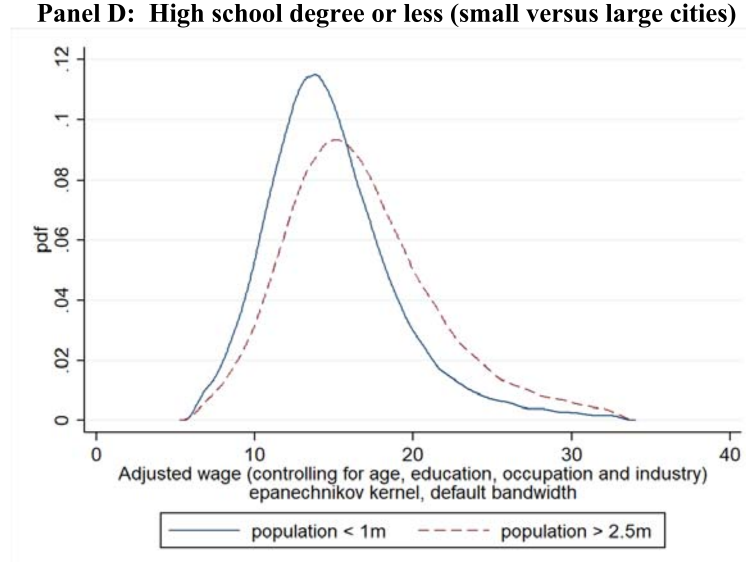
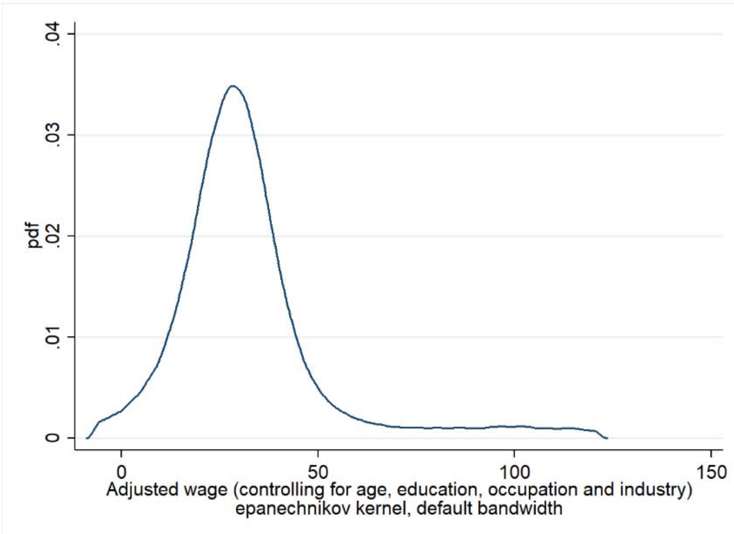
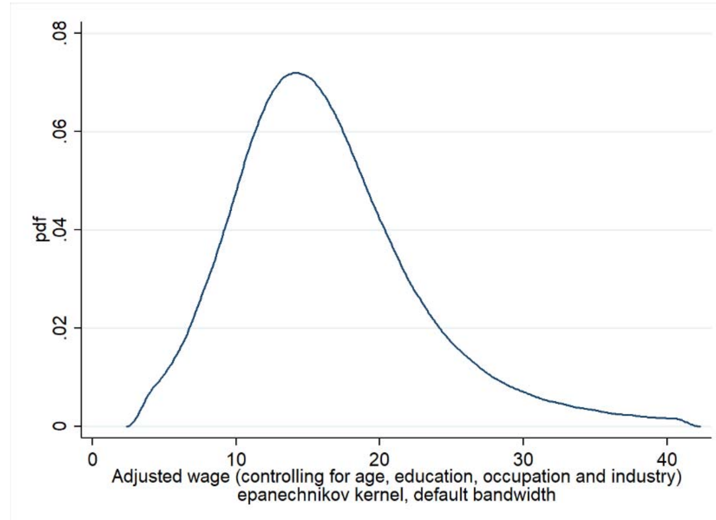


Figure 9: Adjusted wage kernel density estimation for male non-Hispanic native-born white full-time worker, age 25-54

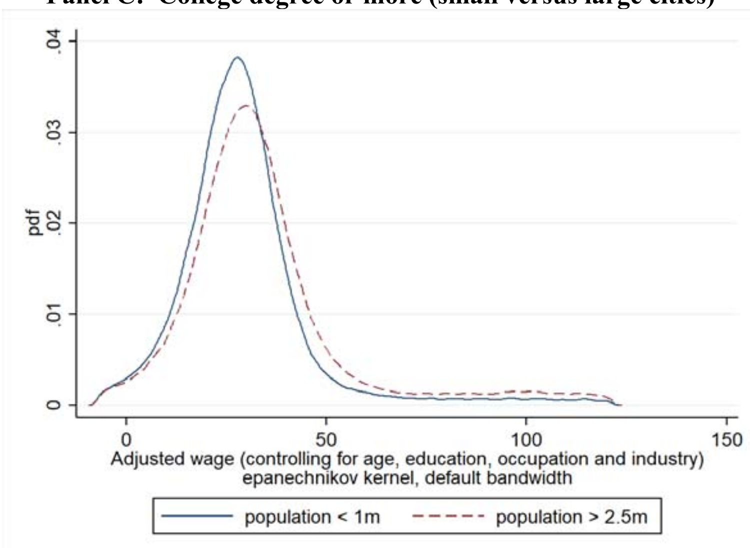
Panel A: College degree or more (all cities together)



Panel B: High school degree or less (all cities together)



Panel C: College degree or more (small versus large cities)



Panel D: High school degree or less (small versus large cities)

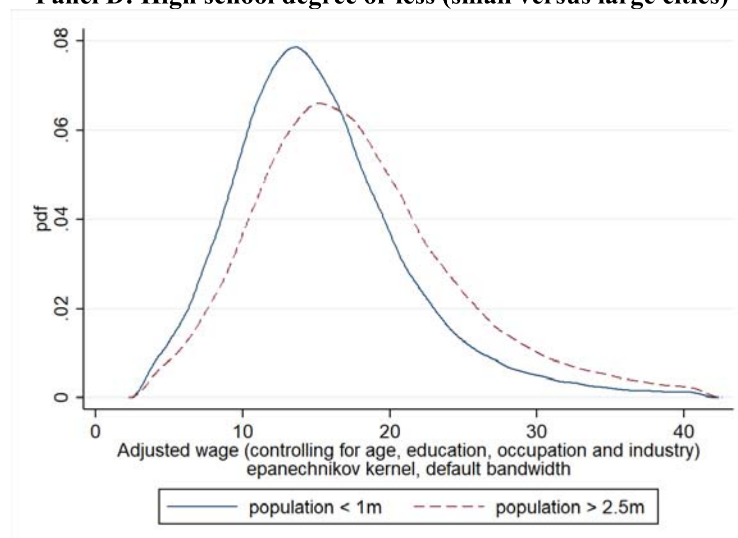
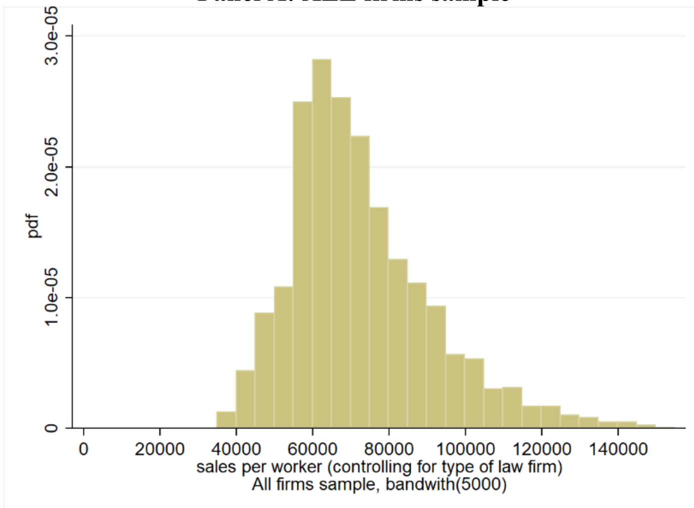
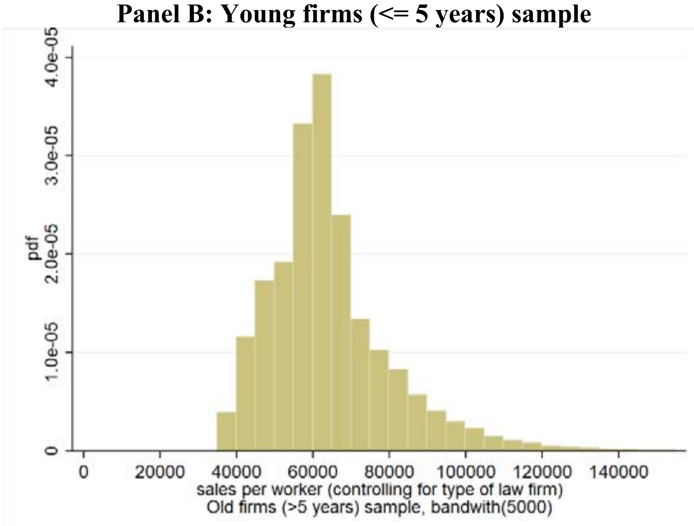


Figure 10a: Histogram estimation of sale per worker for law firms (bandwidth $5,000)

Panel A: ALL firms sample



Panel B: Young firms (<= 5 years) sample



Panel C: Old firms (> 5 years) sample

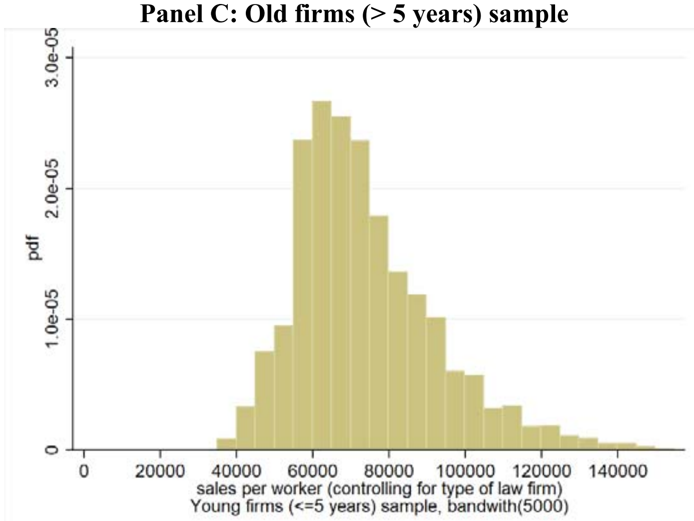
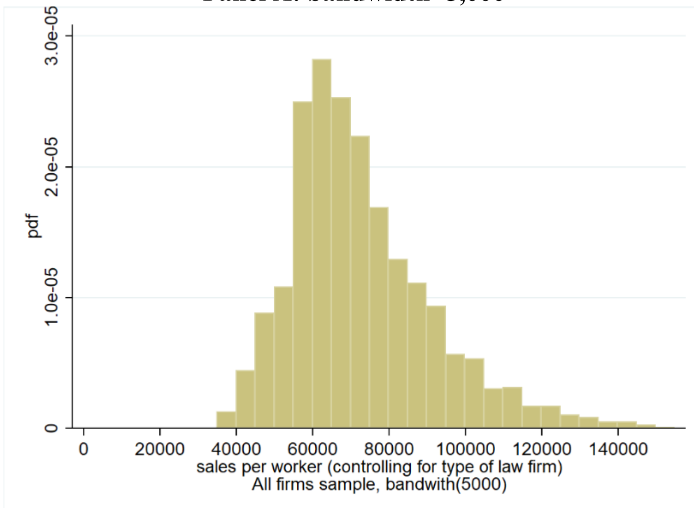
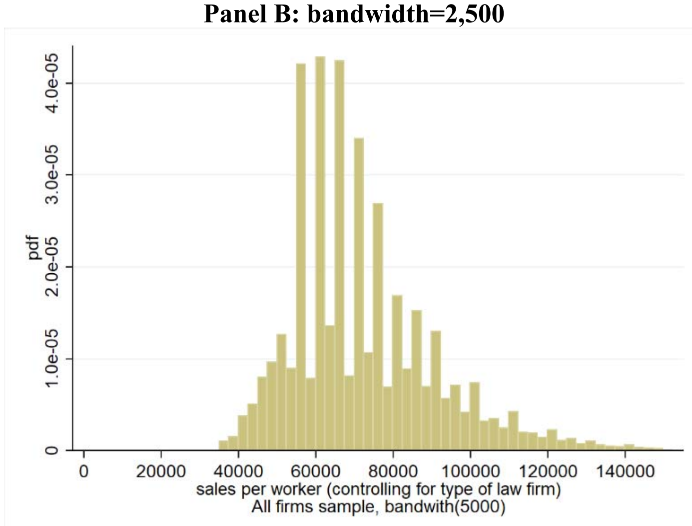


Figure 10b: Histogram estimation of sale per worker for law firm using different bandwidth

Panel A: bandwidth=5,000



Panel B: bandwidth=2,500



Panel C: bandwidth=7,500

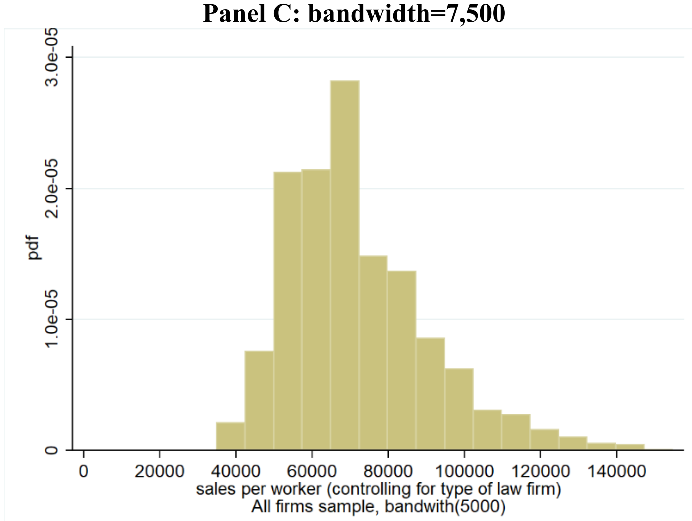


Figure 11a: Histogram estimation of adjusted wage for married female sample (bandwidth $3)

Panel A: College degree or more

Panel B: High school degree or less

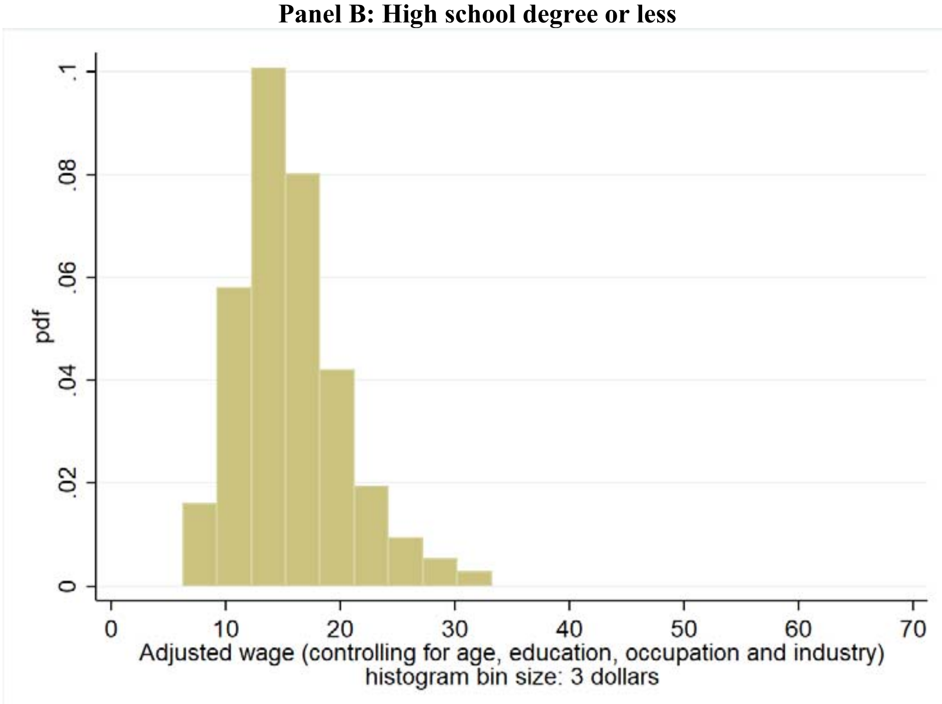
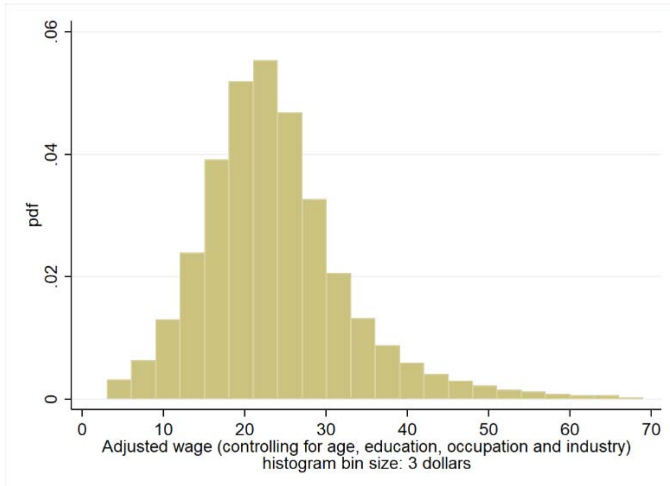
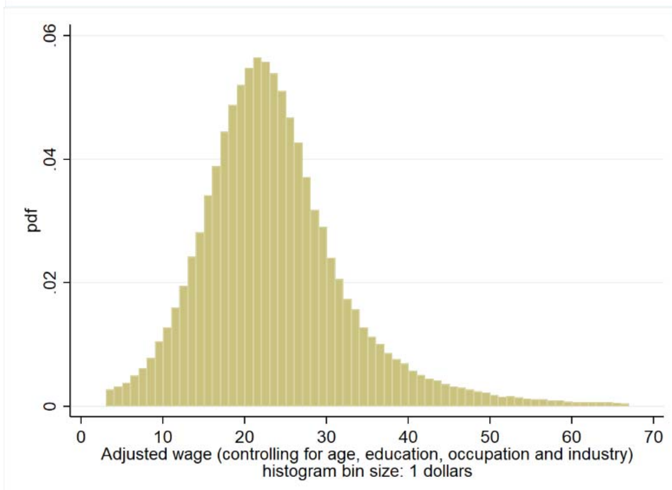


Figure 11b: Histogram estimation of adjusted wage for skilled (college degree or more) married female sample using different bandwidth

Panel A: bandwidth=3



Panel B: bandwidth=1



Panel C: bandwidth=5

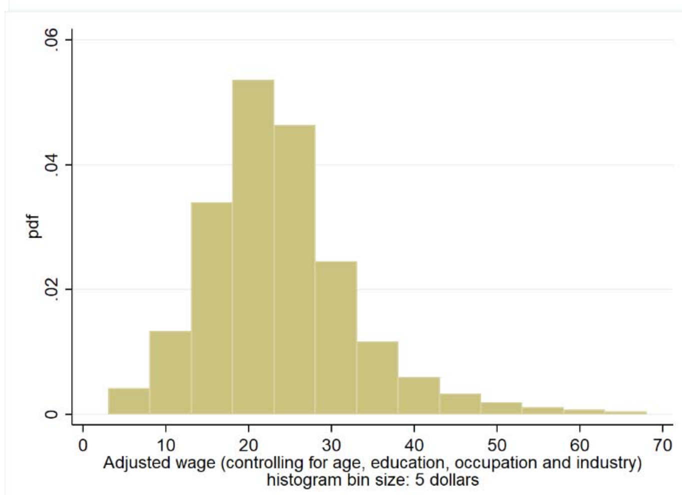
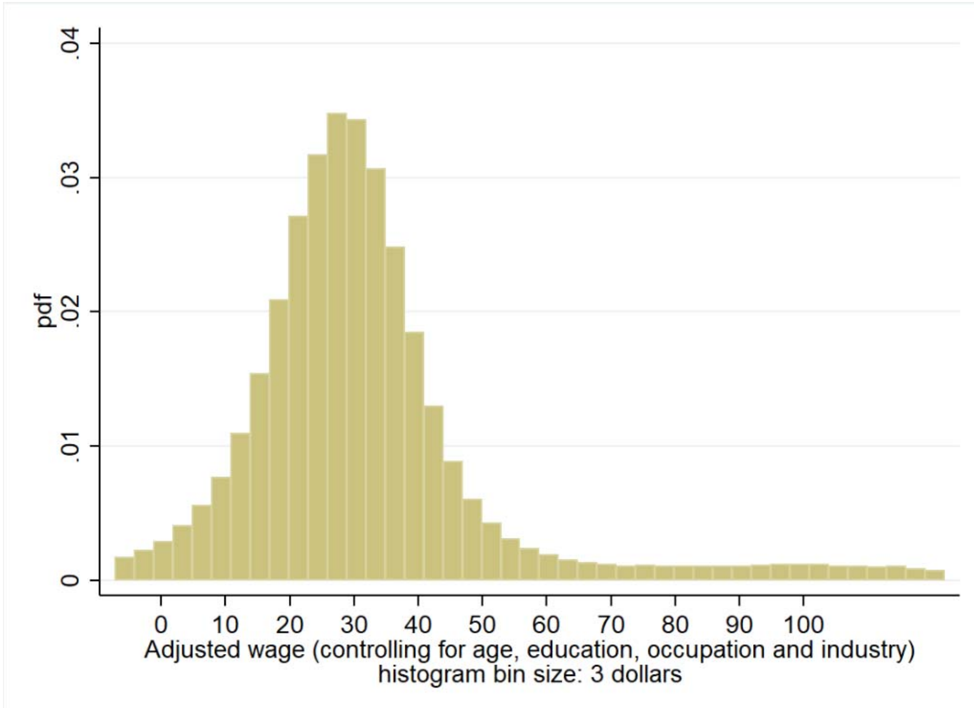


Figure 12a: Histogram estimation of adjusted wage for male sample (bandwidth $3)

Panel A: College degree or more



Panel B: High school degree or less

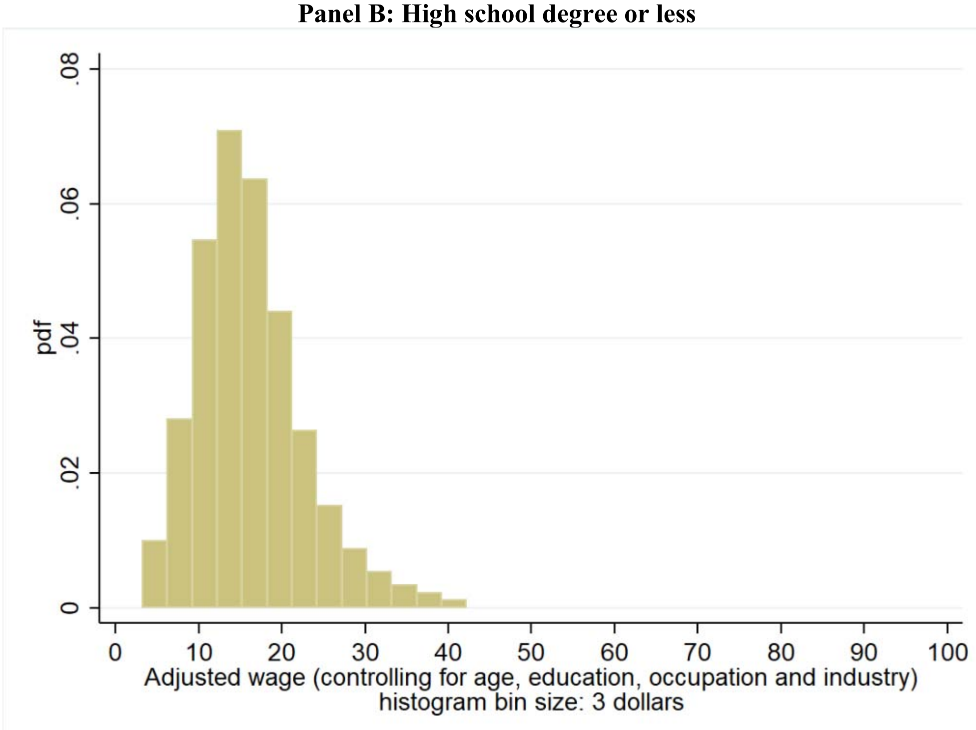
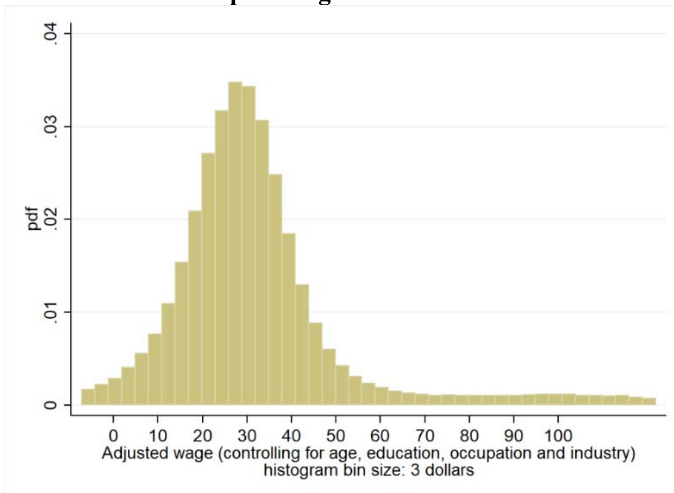
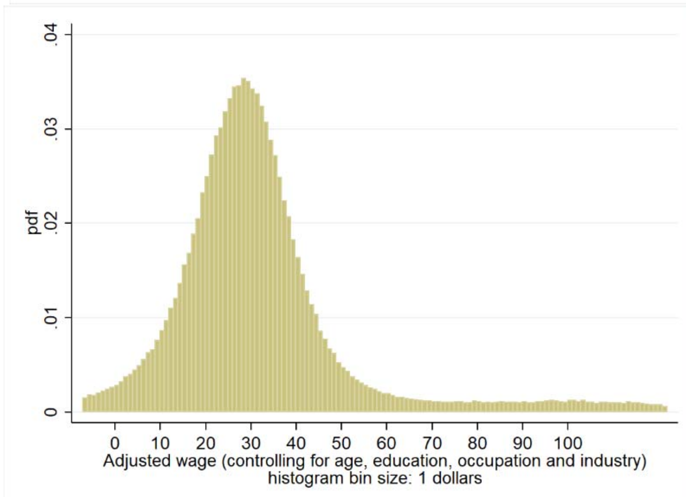


Figure 12b: Histogram estimation of adjusted wage for skilled (college degree or more) male sample using different bandwidth

Panel A: bandwidth=3



Panel B: bandwidth=1



Panel C: bandwidth=5

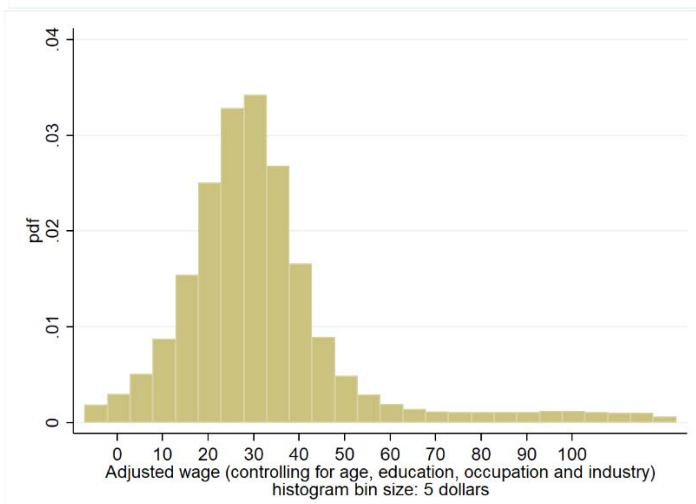
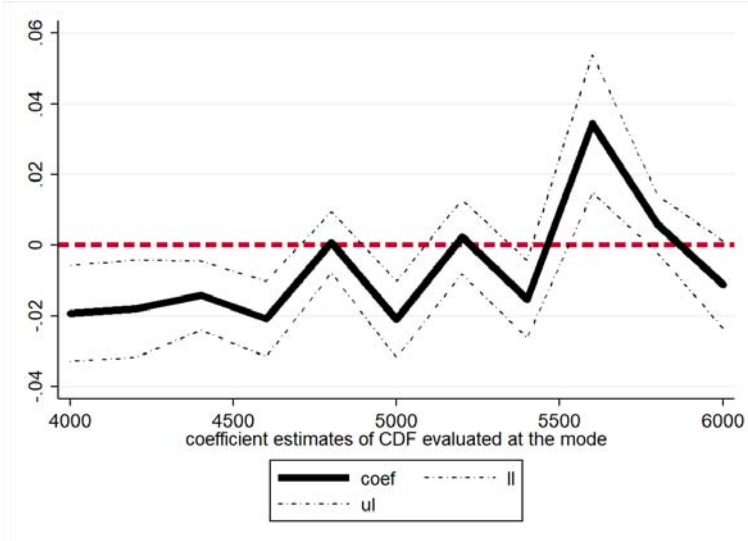
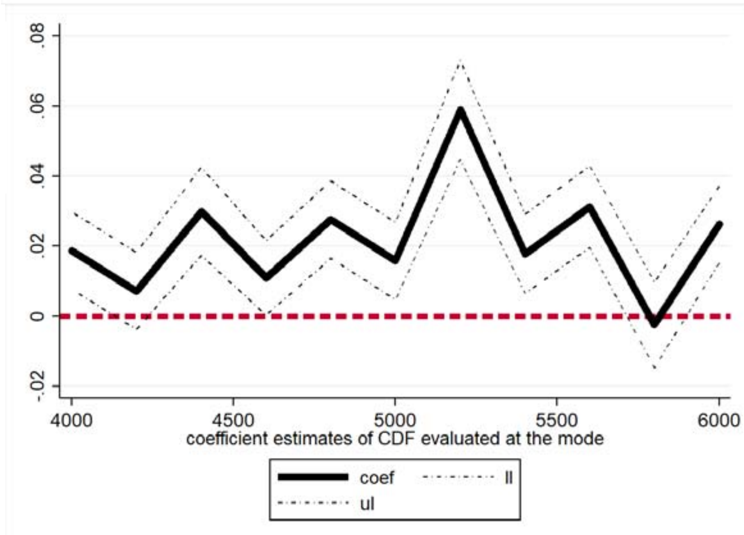


Figure 13a: Law establishment modal CDF robustness check using different bandwidth

Panel A: ALL firms sample.



Panel B: Young firms (<= 5 years) sample.



Panel C: Old firms (> 5 years) sample.

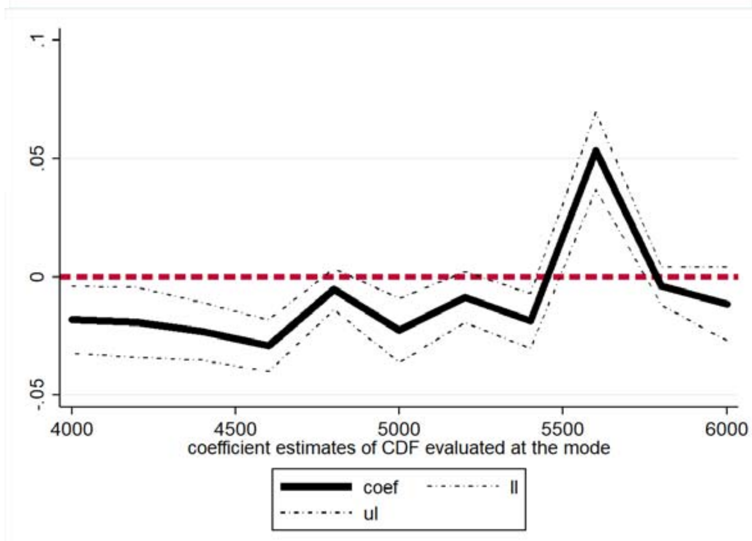
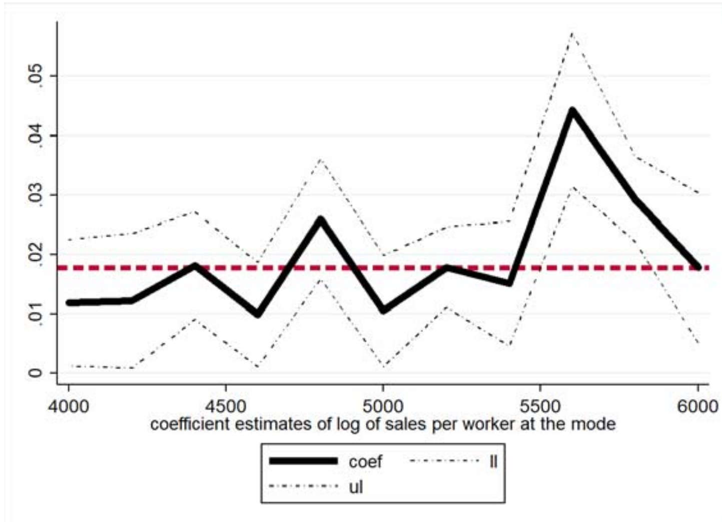
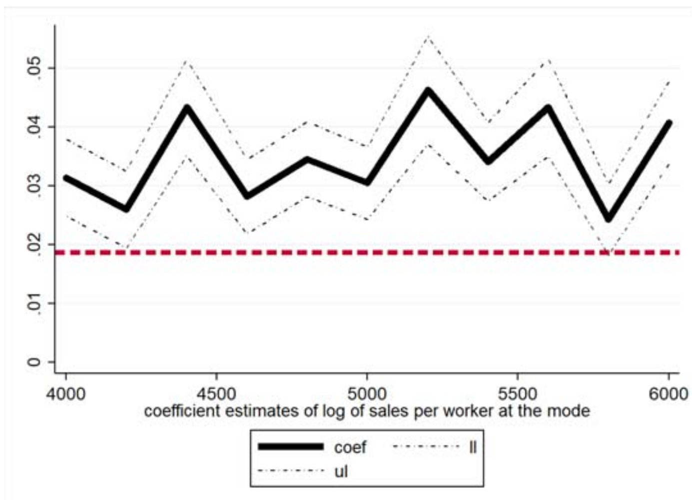


Figure 13b: Law establishment modal return robustness check using different bandwidth

Panel A: ALL firms sample.



Panel B: Young firms (<= 5 years) sample.



Panel C: Old firms (> 5 years) sample.

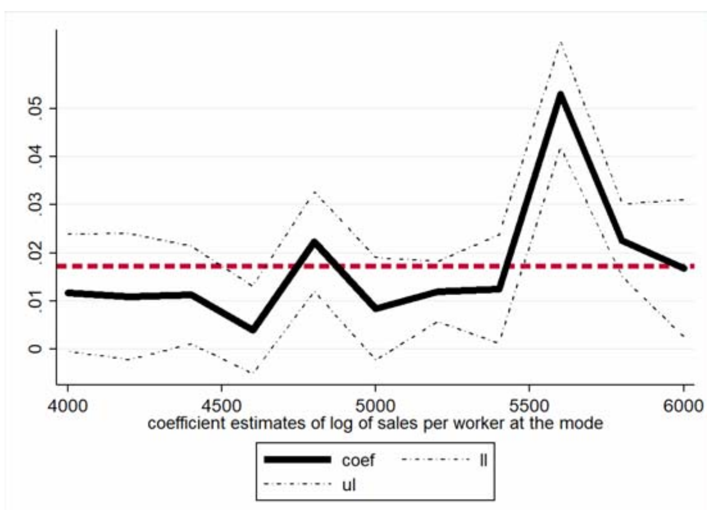
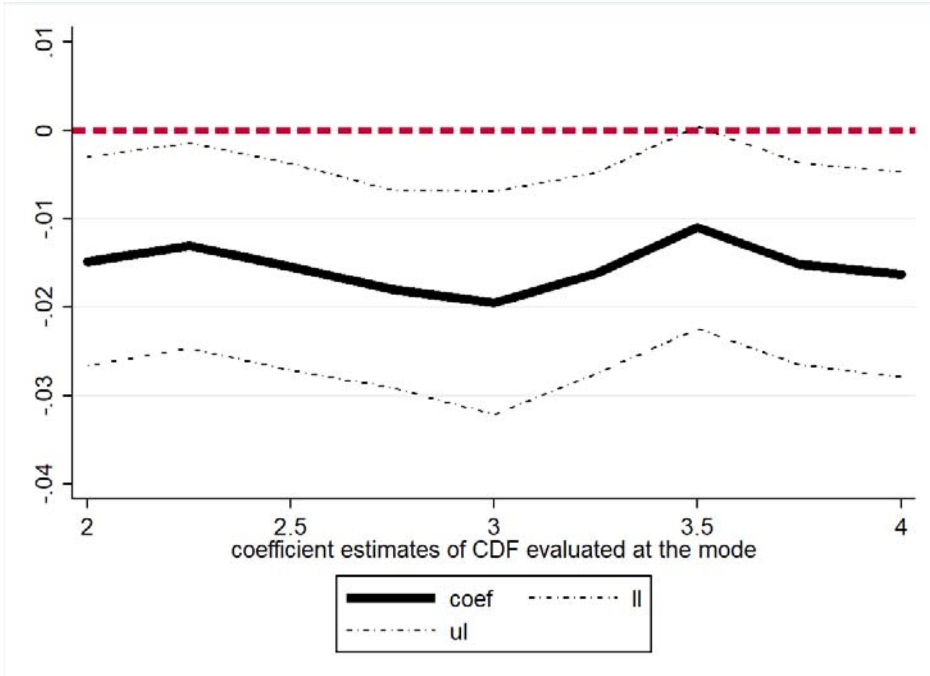


Figure 14a: Married sample modal CDF robustness check using different bandwidth

Panel A: College degree or more



Panel B: high school degree or less

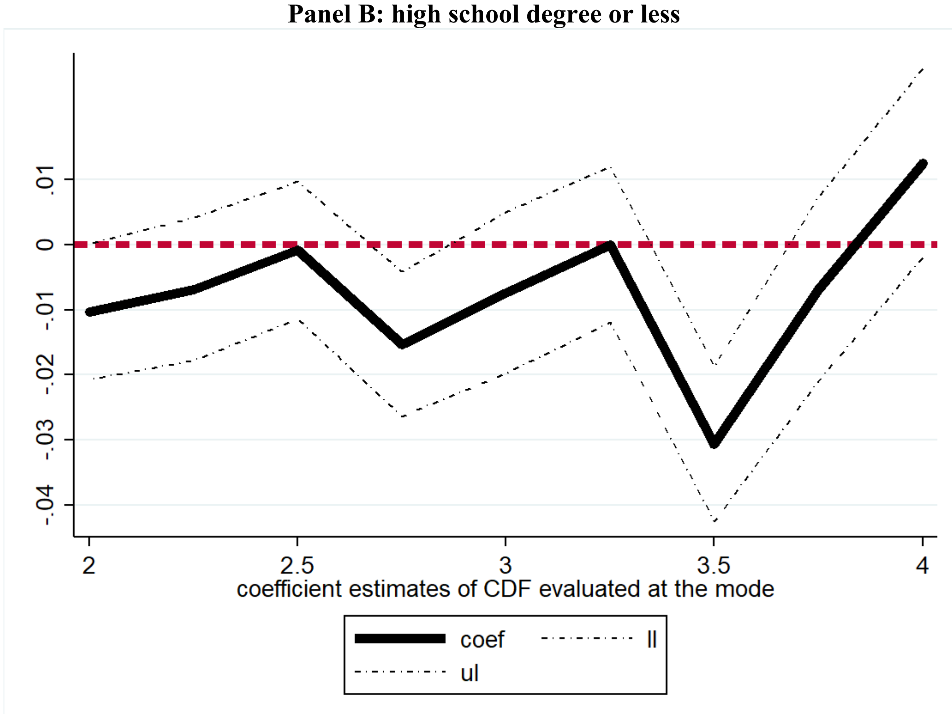
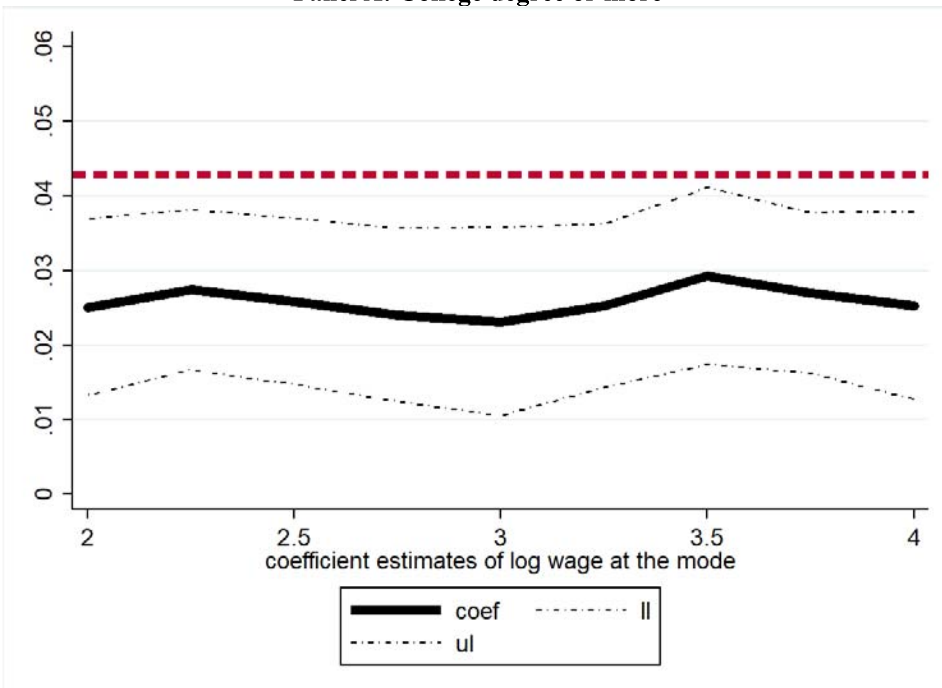


Figure 14b: Married female sample modal return robustness check using different bandwidth

Panel A: College degree or more



Panel B: high school degree or less

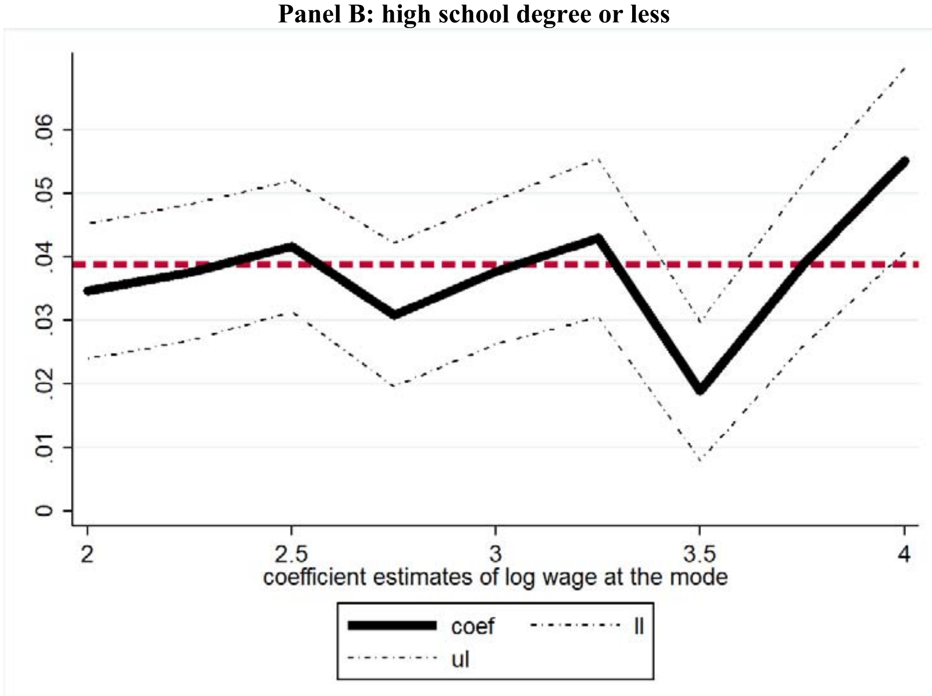
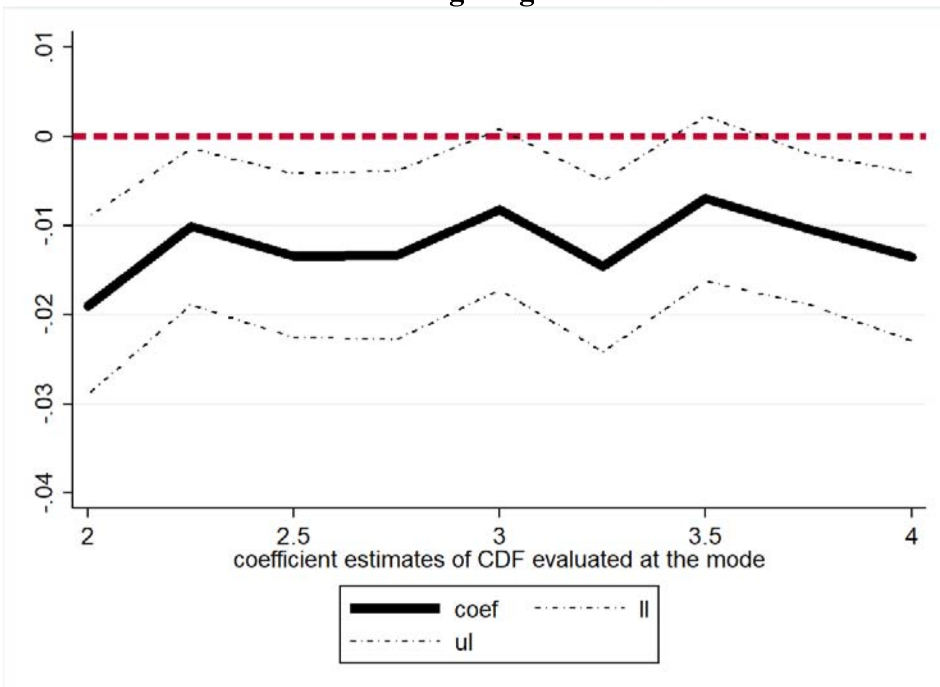


Figure 15a: Male sample modal CDF robustness check using different bandwidth

Panel A: college degree or more



Panel B: high school degree or less

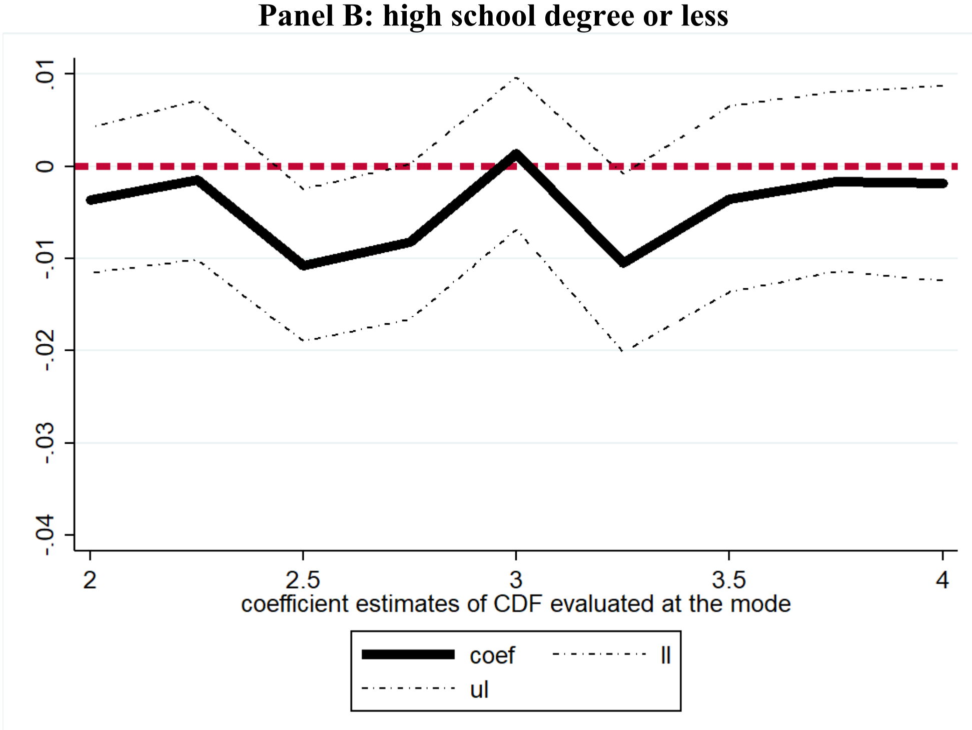
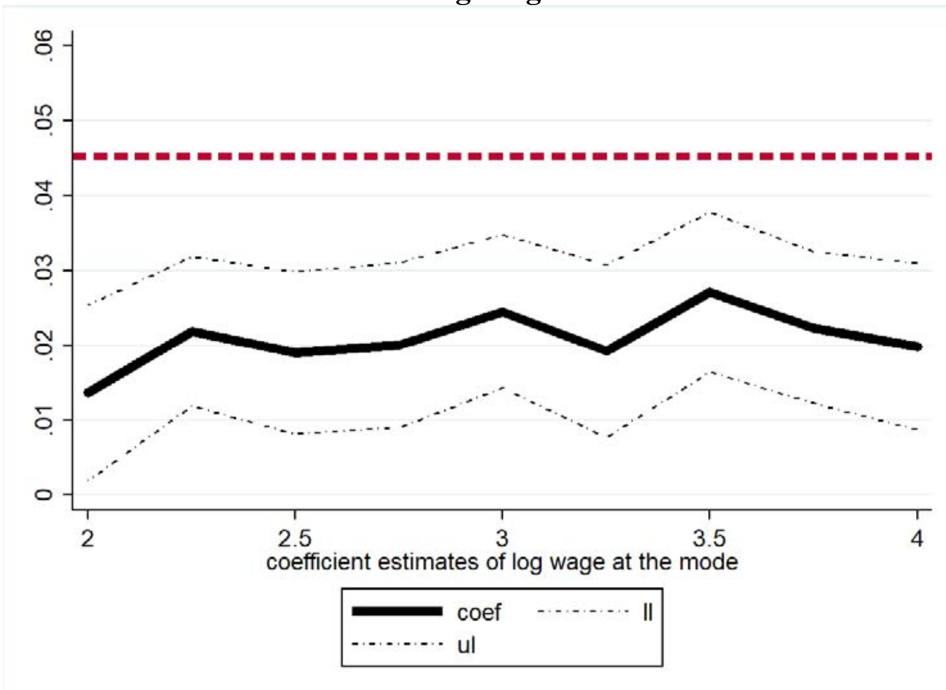


Figure 15b: Male sample modal return robustness check using different bandwidth

Panel A: college degree or more



Panel B: high school degree or less