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
Department of Economics and Center for Policy Research
Syracuse University Syracuse, New York, 13244-1020
Phone: (315) 480-9309
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

May 22, 2018

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 largely absent for low-skilled workers. Doubling city size causes skilled and low-skilled worker productivity to increase by roughly 2.4 and 4 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 impactful 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 law firms, the modal return from doubling city size is approximately 1% as compared to 1.7% 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 2.5% 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 largely absent 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 3.8% for women and 4.4% 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,

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

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 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.

[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 $y = 0.85$.

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 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.

[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.

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.

To begin, 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,

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

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,

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

Expression (2.6) indicates that at the mode, 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]

[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.

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 with y satisfies these conditions and ensures a unique solution for (2.7).

Figure 6 illustrates these principles. The upper panel displays a twice 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. If instead, the selection probability is constant, then the selection function will be flat and $\xi_{\pi,y} = 0$ for all y. This would occur when $\pi_1(y)$ is a constant less than 1, indicating that selection is random, as well as when the selection probability is equal to 1. In both cases, the selection function will not shift the mode.[10]

[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.

[10] Alternatively, if selection declines monotonically with y, expression (2.6) still holds but the mode in the conditional density function will shift to the left.

We now formalize three propositions that govern conditions under which the mode is a robust indicator of the presence of selection effects and can also be used as a way to control for selection.

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 disproportionately to the left of the mode and will increase if selection occurs disproportionately to the right of the mode.*
(iii) *These conditions hold regardless of the shape of the selection function.*

Proof: 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.

Proposition 2: *Given Assumptions 1 and 2, if the selection probability is constant for $y > y^*$ where y^* 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.*

Proof: If $\pi_1(y) = p < 1$ from a point y^* to the left of the mode and beyond, then selection is random for all $y > y^*$. 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 setting in which the selection function reaches an asymptote p to the right of the mode at $y^* > y_m$. 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, and as an illustration, 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 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 $\kappa < 1/2$ 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, $\mu + \sigma$) 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,

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

with a and b both positive and $y^* > y_m$.

Assumption 4 makes explicit that $\pi_1(y)$ reaches an asymptote $p \leq 1$ at y^* to the right of the latent density mode y_m. In addition, $\pi_1(y)$ has a positive vertical intercept $a > 0$ that governs the selection probability for the lowest skilled individuals. The selection elasticity in expression (2.7) is then,

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

where larger values for a reduce the magnitude of $\xi_{\pi,y}$ causing the selection function to become more inelastic. Note also that as a approaches 1, b must go to zero since $\pi_1(y)$ cannot exceed 1, in which case $\xi_{\pi,y} = 0$.

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 and κ, subject to the elasticity condition in (2.7). As an example, consider the extreme case where the shape parameter κ approaches zero. Then $f_0(y)$ converges to a constant equal to q for the relevant range of y, and from (2.5), $f_1(y) = \pi_1(y)\frac{q}{c}$. In this instance, the conditional density function takes on the shape of the selection function scaled by q/c. With $b > 0$, the mode of $f_1(y)$ shifts all the way to the right to the largest observed value for y, denoted by $y_{\max}$. Moreover, with positive mass to the left of $y_{\max}$, the mean of the conditional density shifts by less than the mode. This extreme example makes clear that as the density function becomes flatter the mode shifts further to the right for a given set of values for *a* and *b* (and provided the selection function reaches an asymptote to the right of y_m). In addition, with a sufficiently flat density function, the mode will shift by more than the mean. These principles are regulated, however, by a further condition. For a given latent density function, the rightward shift in the mode, both in absolute terms and relative to the mean, increases as the selection process hits an asymptote further to the right of y_m. These and related principles are formalized in Proposition 3 below.

> **Proposition 3:** *If the selection function reaches an asymptote p at $y^* > y_m$, and if assumptions 3 and 4 hold, then the mode of the conditional density $f_1(y)$ …*
>
> (i) *Shifts more in response to selection as y^* increases above y_m up to a point y^{**} that satisfies the elasticity condition in (2.7) beyond which the mode shifts no further.*
>
> (ii) *Shifts more in response to selection as κ becomes smaller and the density flatter.*

<u>Proof:</u> See appendix A.

For the class of models for which the selection probability reaches an asymptote to the right of the mode, Proposition 3 says that the extent to which the mode shifts increases as the selection function hits an asymptote further to the right of y_m and also as the density function becomes flatter. It is also worth emphasizing that as y^* approaches y_{max}, the conditional density takes on the shape of selection function scaled by q/c as described above. At the other extreme as $y^* - y_m$ approaches zero, the mode does not shift and Proposition 2 applies.

How sensitive then is the mode to selection? This will depend on the context in question which determines which set of modeling assumptions above are most appropriate and also the anticipated possible values for *a*, *b* and κ. For the application in this paper, we know from both casual and formal observation that small and large cities all have large numbers of very low-skilled individuals present. In addition, roughly 65 percent of the U.S. adult population does not have a college degree. A large share of low-skilled individuals therefore, participate in large city labor markets and this suggest that *a* is rather high which reduces the sensitivity of the mode to selection. Data plots presented later in the paper also clearly indicate that wage distributions among full-time skilled and low-skilled workers in the U.S.

display strongly single-peaked patterns with sharply defined modes. This further suggests that the mode for our present application is not very sensitive to 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.[11]

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.[12] 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.[13] 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

[11] 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.

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

[13] 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.

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 is approximately the same period as the 2016 law firm sample.[14] 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.[15]

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.[16] 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 the large-city distribution of sales per worker is clearly right-shifted as compared to small cities.

---

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

[15] 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).

[16] 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 married female non-Hispanic white native-born full-time workers (age 25-54) was obtained from the 2000 decennial census 5% public use micro sample (PUMS) from IPUMS.[17] Full-time workers were coded as those who report working at least 35 hours per week and 40 weeks per year.[18] 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.[19] 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.[20]

MSA population size was estimated using the 2000 census, the same year as the wage data were drawn from.[21] 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

[17] See Steven et al., 2015 and www.ipums.org. Observations from Alaska and Hawaii were excluded.
[18] 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.
[19] 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.
[20] Using this approach, a small number of observations had negative adjusted wage. Dropping these obsrevations did not affect our results.
[21] 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.

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.

4. Measuring the mode

Our estimation procedure requires that we measure the modal value of the outcome variables (e.g. sale/worker, wage) in each MSA. We illustrate how this is done using law establishment sale per worker data.

We first discretize the sales per worker distribution in each MSA by rounding the sales per worker values to the closest integral using a fixed bandwidth. The choice of rounding bandwidth will be discussed shortly. Then we define the modal sales per worker of each city as the mean of the cell that has the highest frequency in each MSA's discretized sales per worker distribution.

We use this method mainly for two reasons. First, as an example, a five-dollar increase in sale per worker has a cardinal interpretation that is readily understood in both small and large city contexts. Discretizing sales per worker using a fixed, common bandwidth across cities enhances comparability of estimates across MSAs and samples. Second, discretizing as above reduces measurement error associated with rounding when the raw data are reported. As shown in Figure 7a, the density plot of sales per worker includes a number of spikes that likely arise from rounding of reported values. Discretizing the data using a fixed bandwidth reduces the influence of such rounding.[22]

A key part of the procedure above is the choice of bandwidth used to discretize the data. If the bandwidth is too narrow, the discretized distribution will converge towards a uniform distribution with a poorly defined mode. If the bandwidth is too wide, variation in discretized values will be so reduced that it will not be possible to discern meaningful patterns since all of the data would eventually be coded to a single cell. It is necessary therefore, to select a bandwidth that balances these two extremes.

We begin by first documenting the inter-quartile range for the sales per worker distribution. In Table 1, Panel A, 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. From Figure 7b, observe also that for each law firm sample (all, young and old), the difference in modal sale per worker for large and small

[22] As a robustness check, we also identified the mode based on kernel density estimation, allowing the bandwidth to vary across cities. Results were similar to those reported above.

cities is less than $20,000. This suggests that any bandwidth larger than $10,000 would likely not preserve enough variation in the data to yield reliable results.

Figure 10a presents histograms of the aggregate sales per worker data using a $5,000 bandwidth. It is evident that there is a well-defined mode in the distributions for all three groups of firms (all law firms, young and old). 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.

For the reasons above, we choose a $5,000 bandwidth to discretize the sales per worker data in each city.[23] Using that bandwidth, Table 2, Panel A summarizes modal sales per worker estimates across MSAs. The difference in the minimum and maximum modal sales per worker is about $20,000 for all three groups of firms (all, young and old), consistent with the plots in Figure 7b.

The same procedure as above was used to select the bandwidths for the wage distributions for married women and men. In these instances, the bandwidth was set equal to $3. Observe also that for women, in Table 1, Panel B the interquartile range of adjusted wage is $10 for the skilled married female workers and $5 for the low-skilled workers. From the distribution plots in Figure 8, Panels C and D, the difference in the modal adjusted hourly wage between the small and the large cities is less than $5.[24] Figure 11a provides histograms of the adjusted hourly wage based on a $3 bandwidth. 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. We emphasize the $3 bandwidth results here and present extensive robustness checks based on alternate bandwidths later in the paper. Bearing that in mind, Table 2, Panel B summarizes the wage distributions for women using a $3 bandwidth. Across 216 MSAs, the minimum and maximum modal wages are $15 and $30 for skilled married women, and $12 and $18, respectively, for low-skilled married women.

[23] We discuss the robustness of the modal estimates to bandwidths from $4,000 to $6,000 later in the paper.
[24] We discuss the robustness of the results by varying the bandwidth from $2 to $4 in the empirical section.

For prime age men, Table 2 Panel C summarizes modal adjusted hourly wage estimates based on the discretized data. The minimum and maximum modal adjusted wages are $21 and $39 for skilled workers and $9 to $21 for low-skilled workers. Figure 12a plots histograms of the adjusted hourly wages for skilled and low-skilled workers, using a $3 bandwidth. Both panels have a well-defined mode. As a comparison, Figure 12b provides similar histograms using $1 and $5 bandwidths. The patterns are similar to those for women.

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 3 in Section 2 suggests 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, the CDF evaluated at the mode should decline with MSA size. Column (1) estimates in Panel A for all law firms confirm that prior. The coefficient on log population is -0.02 with a t-ratio of 3.81. This indicates that doubling city size reduces the CDF evaluated at the mode by 2 percentage points. This is consistent with our prior that weaker firms disproportionately select 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 0.0158 with a t-ratio of 2.81. This indicates that among newly established companies, weaker firms are *more likely* to be present in larger cities. In the Introduction we described a set of assumptions that would support this result. Specifically, that (i) entry and exit costs for opening a law firm are low, (ii) that lawyers only discover from experience whether operating their own firm dominates working for someone else's company, and (iii) that the rewards from operating a successful law firm increase with city size. While other modeling assumptions could also potentially generate the patterns in Panel B, the overall interpretation of the estimates would persist: among newly created law firms, the share of low performing companies increases with MSA size.[25] 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 both reasons, we expect that among older law firms, selection will disproportionately cause weaker law firms to drop out as MSA size increases. Panel C for older firms confirms that prior; notice that the column (1) coefficient is equal to -0.023 with a t-ratio of 2.73.

Consider now estimates of the elasticity of sales per worker with respect to city size. As discussed earlier, if weaker firms select out of larger MSAs, the estimated elasticity based on the mean should be larger and more upward biased than estimates based on the mode. In column (2) of Panel A (for all law firms), the elasticity estimate based on the mode is 1 percent while the column (3) estimate based on the mean is 1.69 percent. The difference between these two estimates in column (4) is also significant at the 5 percent level (based on a 1-tailed test) with a t-ratio of 1.68. These estimates reinforce the coefficient in column (1) and provide further evidence that weaker firms disproportionately select out in larger cities while also indicating that doubling MSA size increases law firm productivity by 1 percent.

[25] Recall from Proposition 3 in Section 2, that the same qualitative pattern in columns 1-4 of Panel B could arise if weaker firms disproportionately select out of larger cities, opposite from our interpretation above. However, for that to occur the selection process would have to reach an asymptote far to the right of the latent density mode and the latent density function would have to be quite flat. As demonstrated earlier, the law firm sale per worker densities are strongly single peaked. For that reason, we believe that the patterns in Panel B provide compelling evidence that among newly established law firms, weaker companies account for a disproportionate share of firms in larger cites.

A nearly identical set of elasticity estimates is obtained for older law firms as reported in Panel C. For younger law firms, however, a different pattern is again present. In Panel B, observe that the elasticity with respect to city size based on mode is 3.05 percent while the corresponding estimate based on the mean is 1.74. The estimate in column (4) also confirms that this difference is significant, with a t-ratio of -3.94. These estimates are consistent with the positive coefficient in column (1) and strengthen the view that as MSA size increases, the set of newly established law firms includes an increasing share of weaker companies.

Estimates in Table 3 are based on discretized distributions using a $5,000 bandwidth. To examine the robustness of the results with respect to bandwidth choice, we re-estimated our models varying the bandwidth from $4,000 to $6,000 in $200 increments. Figures 13a and 13b plot the resulting estimates.

The three panels in Figure 13a (for all law firms, young and old) report coefficient plots for the first stage regression with the CDF evaluated at the mode as the dependent variable. 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. In each panel, the elasticity based on the mean is indicated by the dashed horizontal line and provides a point of comparison. 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. From previous literature, (e.g. Rosenthal and Strange, 2004), the true elasticity of return with respect to city size is likely no larger than 5 percent. Relative to that benchmark, over or underestimating the elasticity at the mode by even a few percentage points is therefore empirically important. This highlights a power issue when attempting to use the law firm sample to estimate the return to MSA size based on the mode.

Summarizing, the estimates in Table 3 support the anticipated view that threshold-based selection causes weaker law firms to drop out over time. Moreover, controlling for threshold effects by using the mode, estimates indicate that among seasoned law firms, doubling city size increases productivity by roughly 0.84 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.

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 roughly -0.02 with a t-ratio of 3.02. 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 2.3 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 3.43). These patterns indicate that selection substantially 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 -1.17. This indicates that selection is largely absent among low-skilled workers, an implication of which is that the return to MSA size should be similar when evaluated at both the mode and the mean. This prediction is confirmed in columns 2-4. Based on the mode, the estimated elasticity with respect to MSA size is 3.8 percent, while based on the mean, the corresponding estimate is 3.9 percent. This small difference is also insignificant, with a t-ratio in column (4) of 0.24.

Results in Table 3 were obtained for a discretized wage distribution using a $3 bandwidth. Figure 14a plots coefficient estimates based on the CDF evaluated at the mode for 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.[26] 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.

Summarizing, for college educated married women, there is compelling evidence that selection effects in larger cities disproportionally drive less productive workers out of the full-time labor market. Consistent with that pattern, doubling city size increases wages for college educated women by 2.3 percent measured at the mode and 4.3 percent measured at the mean. For low-skilled married women, evidence of selection in a manner associated with city size is largely absent. For this portion of the work force, doubling city size increases wage by roughly 3.8% at both the mode and the mean.

[26] 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.

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 -0.008 with a t-ratio of 1.79. Consistent with that estimate, the wage elasticity with respect to city size is 2.5% when measured at the mode (column 2) and 4.5% when measured at the mean (column 3). The difference in these estimates is also significant with a t-ratio of 3.87 in column (4). Together, these estimates confirm that among college educated men, unusually productive individuals tend to sort into larger cities.

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 0.0014 with a t-ratio of 0.34. In the absence of systematic sorting, the return to city size measured at the mode and the mean should be similar. That prior is partly supported by estimates in columns 2 and 3. Notice that the estimated elasticities based on the mode and mean are 4.4% and 3.6%, respectively while the t-ratio on that difference in column (4) is -1.75. Although there is some difference in the two estimates, the magnitude of the difference is small relative to that for skilled individuals (both for men and women). In addition, the larger elasticity based on the mode relative to the mean is consistent with the sign predicted by the positive (though insignificant) coefficient in column 1. These conclusions are also supported by robustness checks based on different bandwidths.

Figure 15a plots estimates of the MSA size coefficient for the CDF regression (column (1) in Table 5) using different 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 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.

To summarize, the patterns for male full-time workers suggest that among college educated individuals, migration tends to draw unusually productive individuals to larger cities. The elasticity of wage with respective city size is 2.5% measured at the mode compared to 4.5 percent measured at the mean. As with the married women sample, ignoring selection nearly doubles the estimated return to MSA size among skilled individuals. For low-skilled workers, however, selection effects appear to be slight if at all.

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 2.5% for skilled workers and 4.4% for low-skilled. Similar estimates are obtained based on 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% with larger effects among newly originated firms and more pronounced selection among older companies as weaker firms 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

Check back soon!

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

Panel A
Modal sales per worker estimates for law establishments

	(1)	(2)	(3)	(4)	(5)	(6)
	Min	Max	Median	Mean	Std.	Observation
All Firms	50,000	70,000	55,000	57,562	4,383	239
Young Firms (<= 5 years)	45,000	65,000	55,000	55,659	3,781	239
Old Firms (> 5 years)	50,000	75,000	55,000	58,075	4,867	239

Panel B
Modal wage estimates for married non-Hispanic native-born white full-time female workers, aged 25-54

	(1)	(2)	(3)	(4)	(5)	(6)
	Min	Max	Median	Mean	Std.	Observation
College degree or more	15.00	27.00	21.00	21.27	2.29	216
High school degree or less	12.00	18.00	15.00	14.03	1.56	216

Panel C
Modal adjusted wage summary statistics for male non-Hispanic native-born white full-time worker, age 25-54

	(1)	(2)	(3)	(4)	(5)	(6)
	Min	Max	Median	Mean	Std.	Observation
College degree or more	21.00	39.00	27.00	27.52	2.98	262
High school degree or less	9.00	21.00	15.00	13.78	1.79	262

[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] 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.

[c] Bandwidth used to define modal wage is $3. Bandwidth used to define modal sales per worker is $5,000.

Table 3: OLS results based on 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	-0.0210	0.0105	0.0169	0.0064
	(-3.81)	(2.20)	(7.23)	(1.68)
R-squared	0.058	0.021	0.164	0.012
Observations	239	239	239	239
Panel B: Young Firms (<= 5 years)				
Log population in MSA	0.0158	0.0305	0.0174	-0.0131
	(2.81)	(9.75)	(6.44)	(-3.94)
R-squared	0.022	0.224	0.003	0.041
Observations	239	239	239	239
Panel C: Old Firms (> 5 years)				
Log population in MSA	-0.0226	0.0084	0.0163	0.0080
	(-3.28)	(1.54)	(6.41)	(1.69)
R-squared	0.054	0.012	0.132	0.154
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.

Table 4: OLS results based on married non-Hispanic female native-born white 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	-0.0195	0.0232	0.0428	0.0197
	(-3.02)	(3.60)	(9.98)	(3.43)
R-squared	0.033	0.049	0.312	0.043
Observations	216	216	216	216
Panel B: High school degree or less				
Log population in MSA	-0.0074	0.0377	0.0388	0.0012
	(-1.17)	(6.50)	(10.03)	(0.24)
R-squared	0.005	0.126	0.291	0.000
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.

Table 5: OLS results based on male non-Hispanic native-born white 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	-0.0082	0.0245	0.0452	0.0207
	(-1.79)	(4.70)	(12.58)	(3.87)
R-squared	0.009	0.055	0.352	0.041
Observations	262	262	262	262
Panel B: High school degree or less				
Log population in MSA	0.0014	0.0440	0.0361	-0.0079
	(0.34)	(6.88)	(7.46)	(-1.75)
R-squared	0.000	0.125	0.173	0.009
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.

Figure 1: Productivity Distributions with Agglomeration Economies

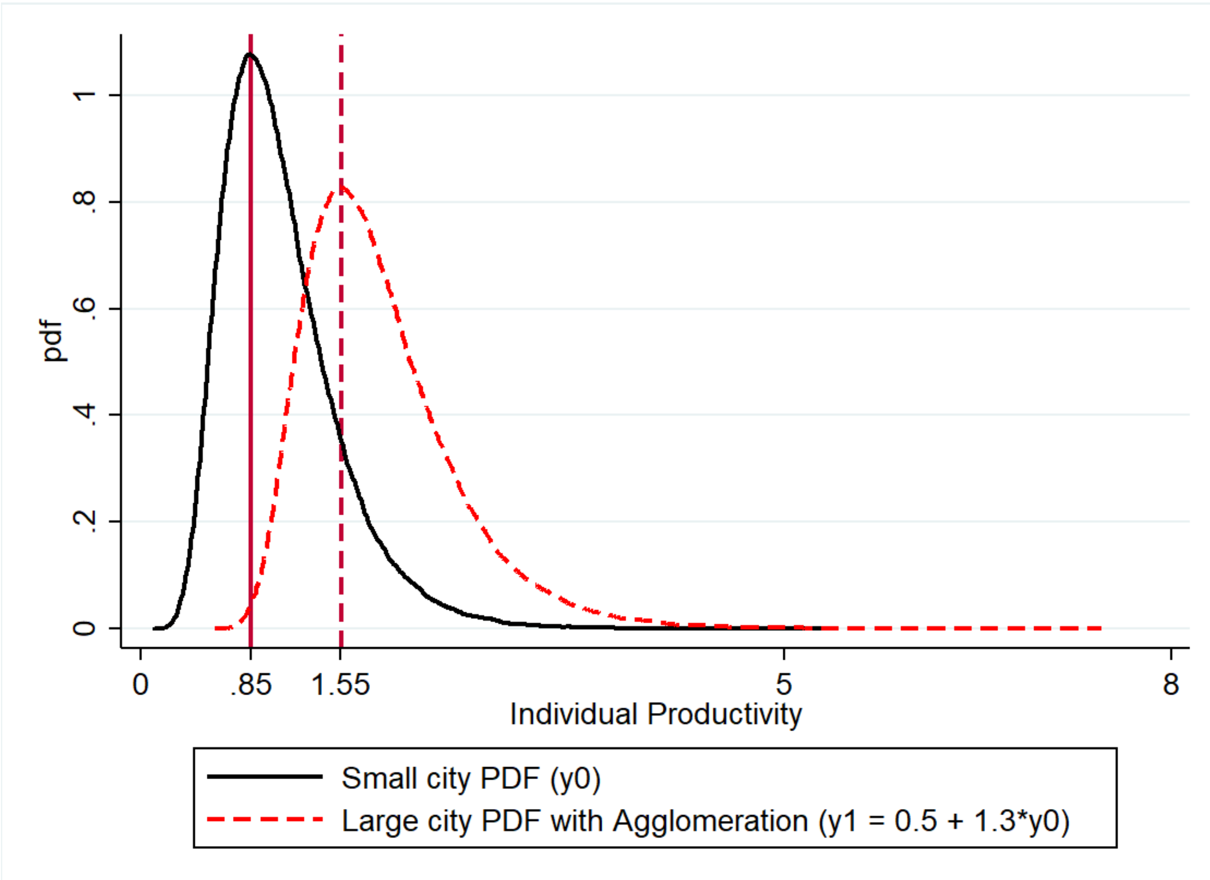
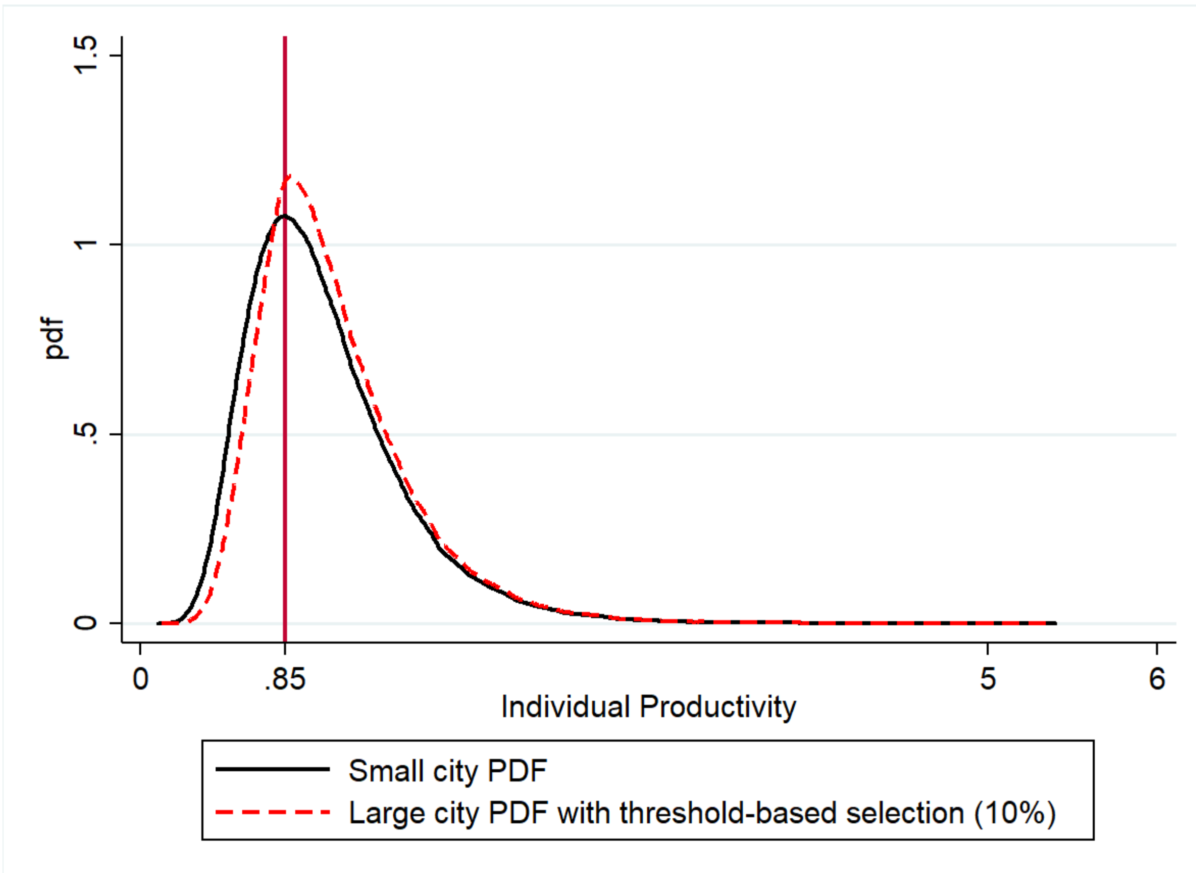


Figure 2: Productivity Distributions with Threshold Effects



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

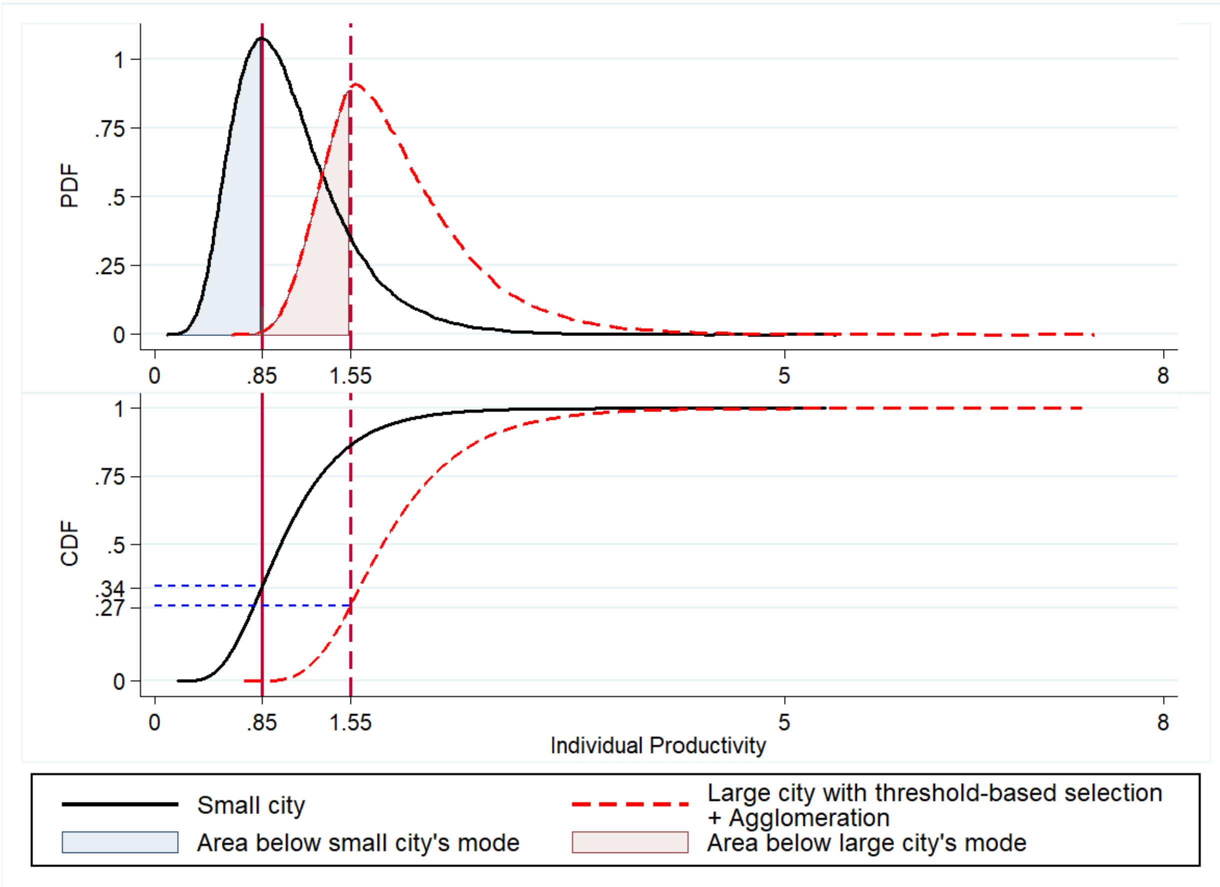
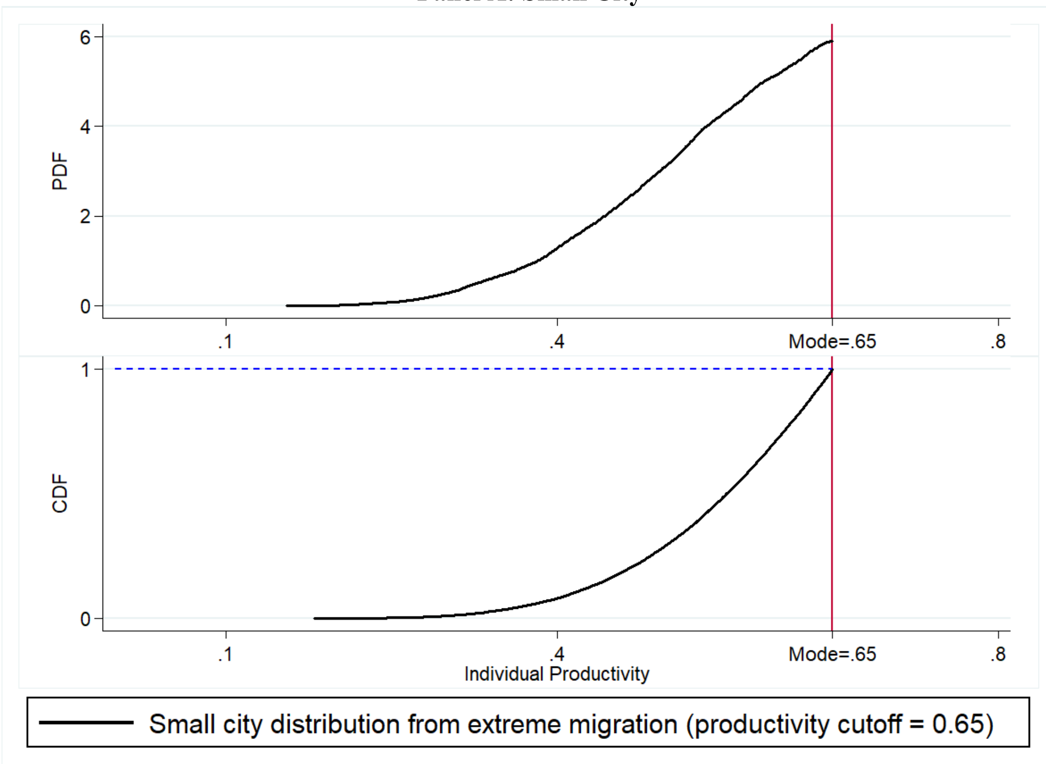


Figure 4a: Small and Large City Productivity Distributions from Extreme Migration Effects (Original Mode=0.85 and Migration Productivity Cutoff = 0.65)

Panel A: Small City



Panel B: Large City

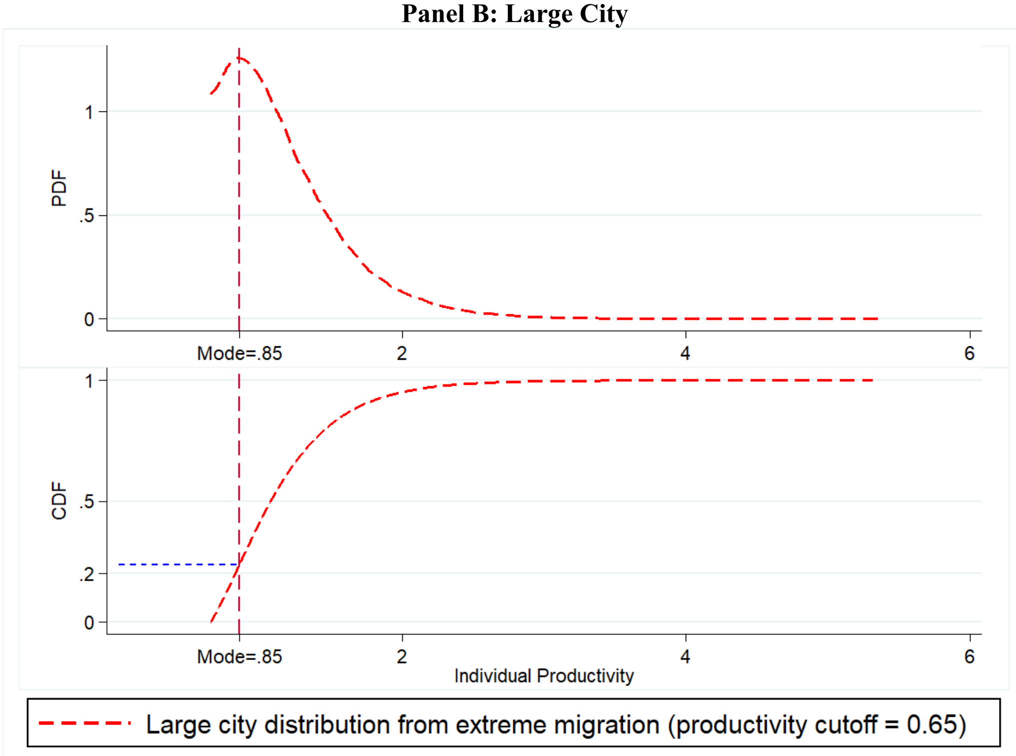
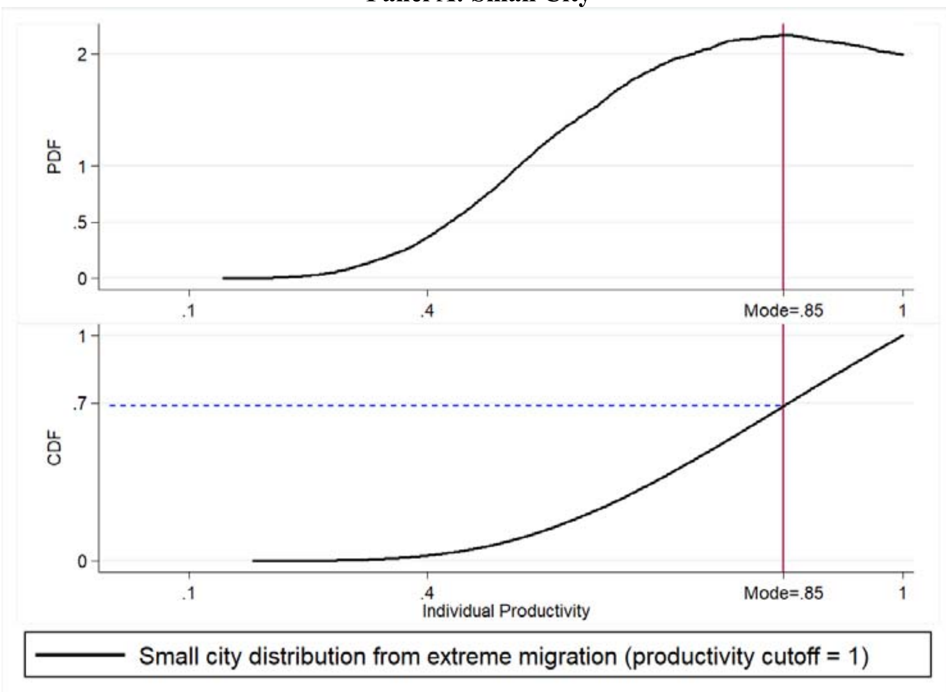


Figure 4b: Small and Large City Productivity Distributions from Extreme Migration Effects (Original Mode=0.85 and Migration Productivity Cutoff = 1)

Panel A: Small City



Panel B: Large City

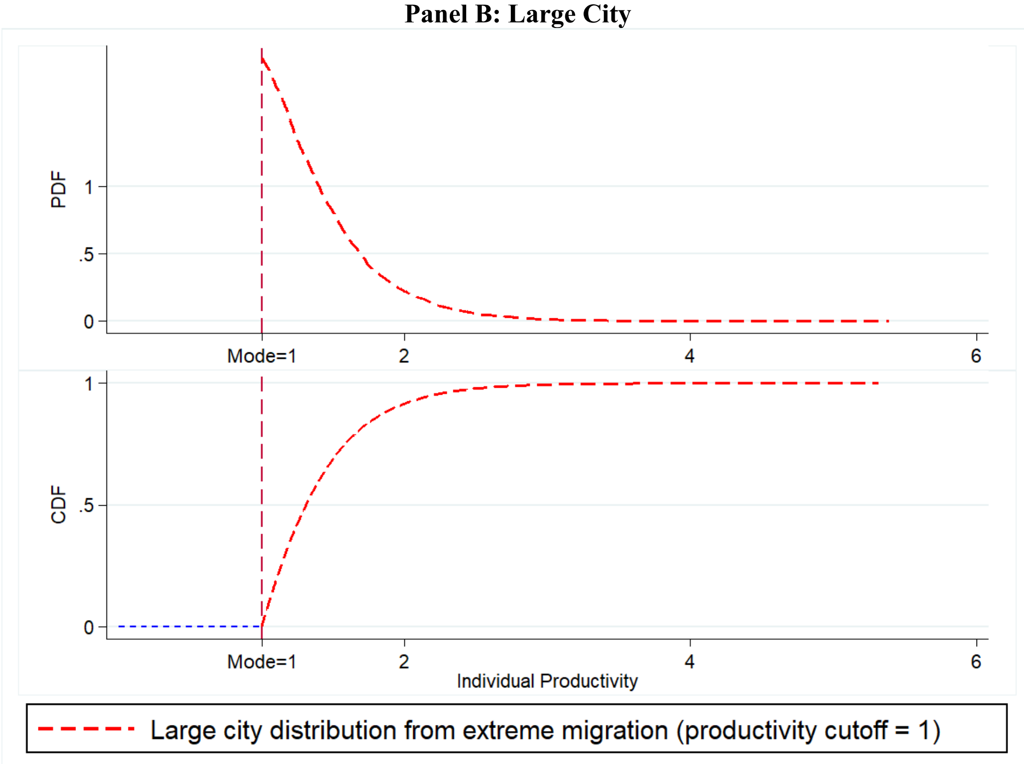


Figure 5a: Productivity Distributions with Migration Effects

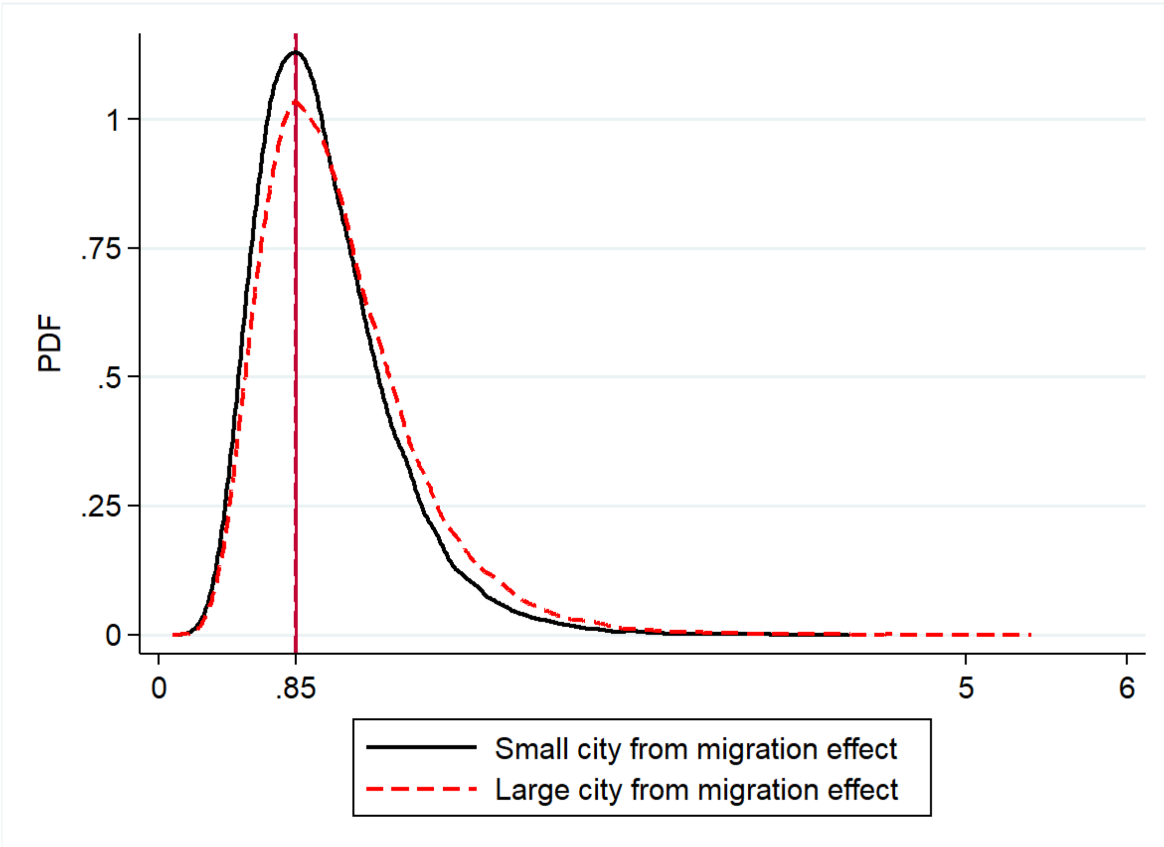


Figure 5b: Productivity Distributions with Agglomeration Economies and Migration Effects

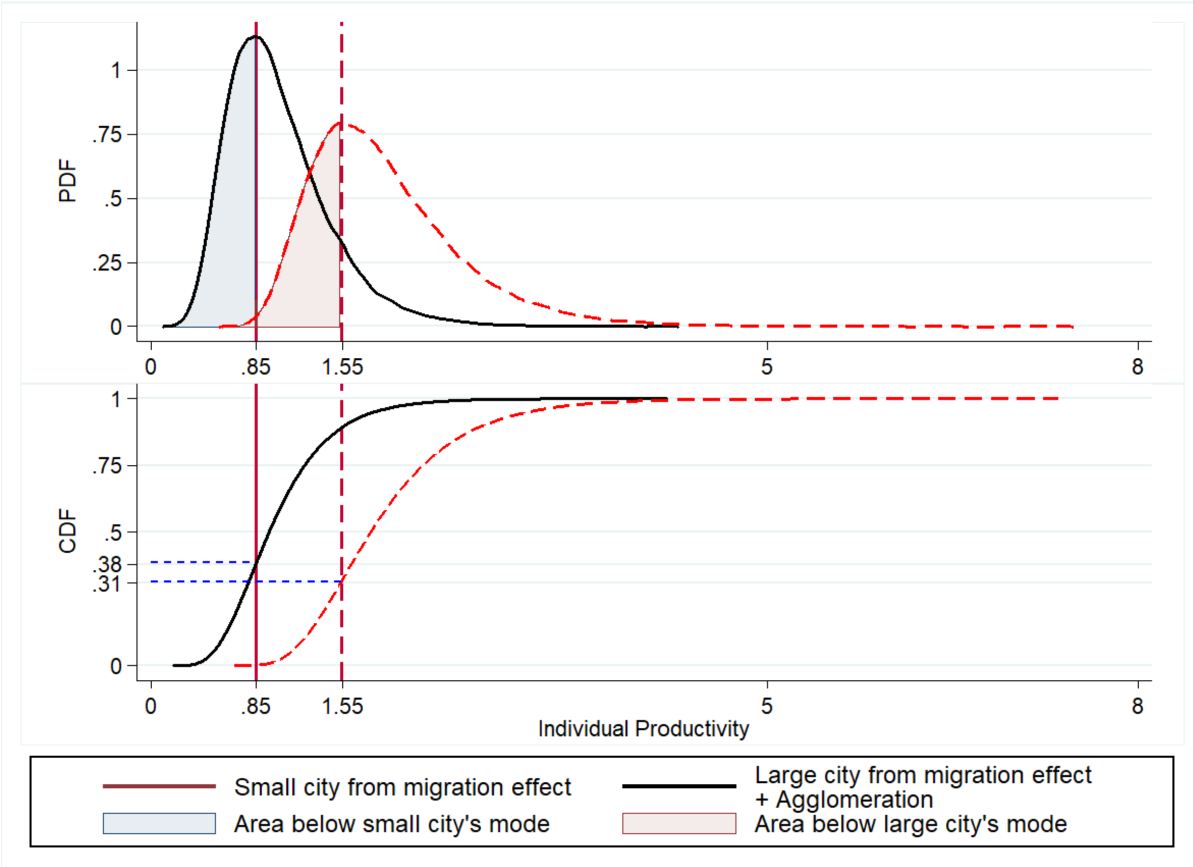


Figure 6: Slope Conditions and Shifts in the Mode

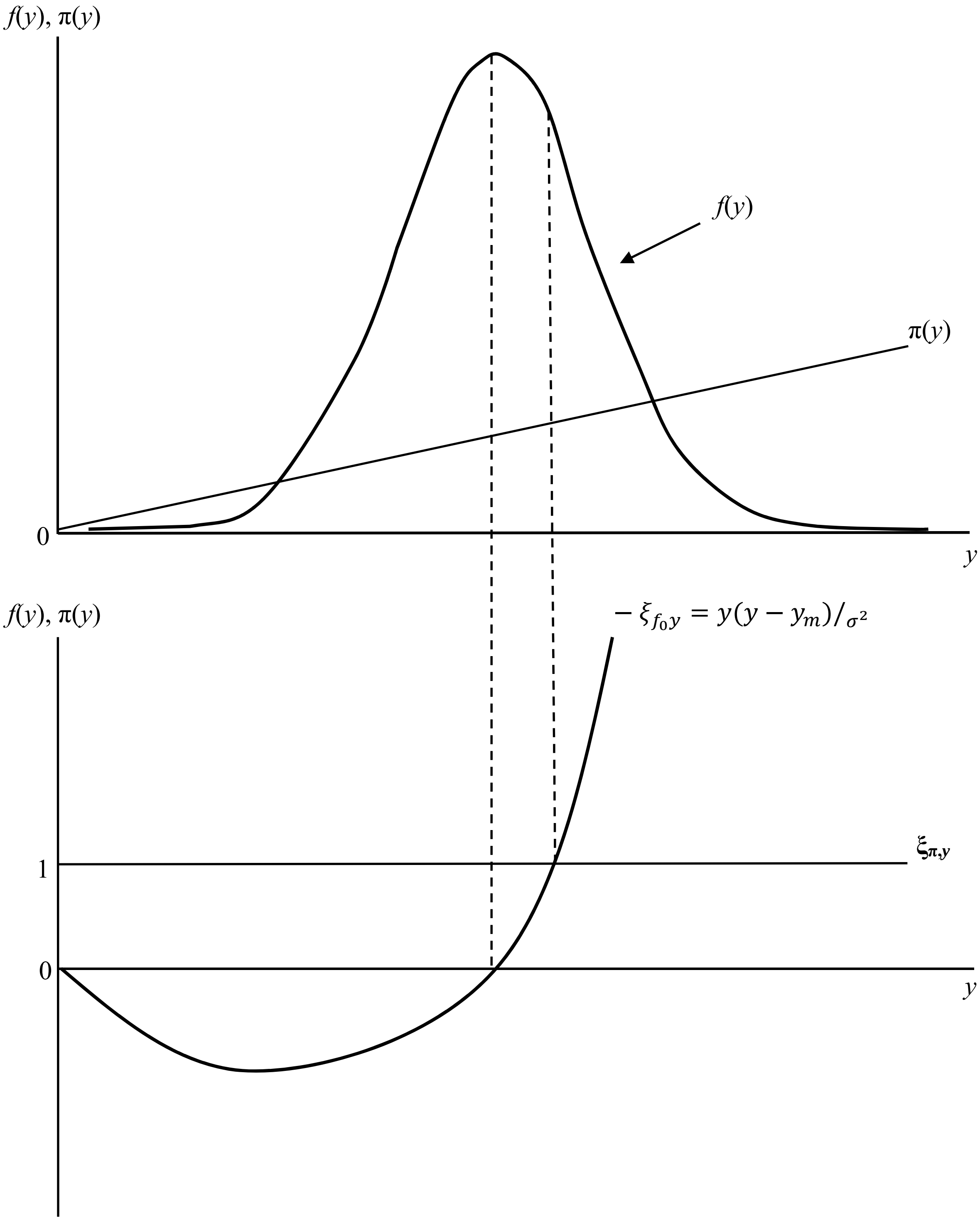
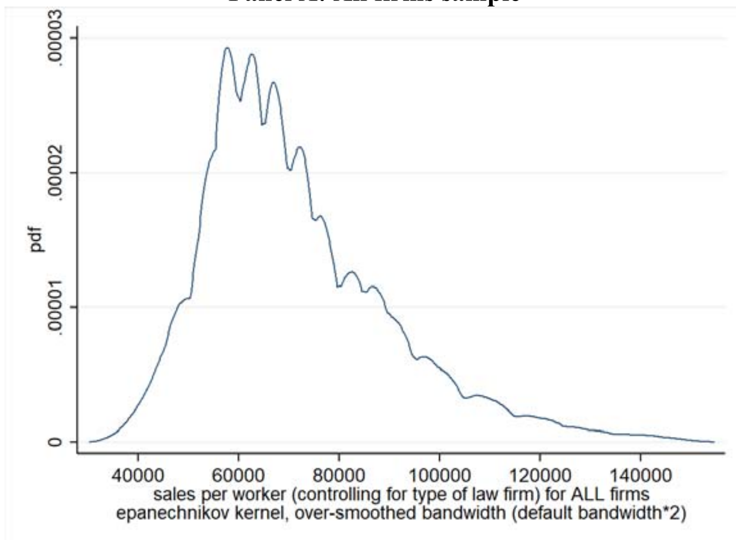
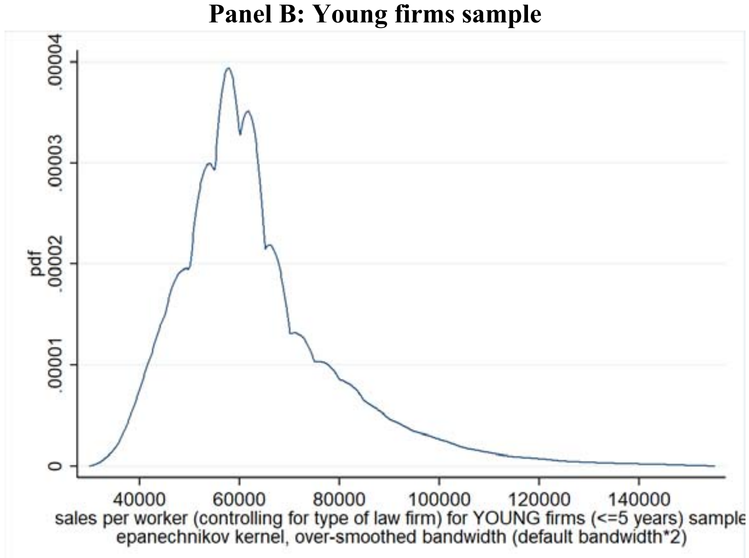


Figure 7a: Sale per worker kernel density estimation for law firms in the Unites States

Panel A: All firms sample



Panel B: Young firms sample



Panel C: Old firm sample

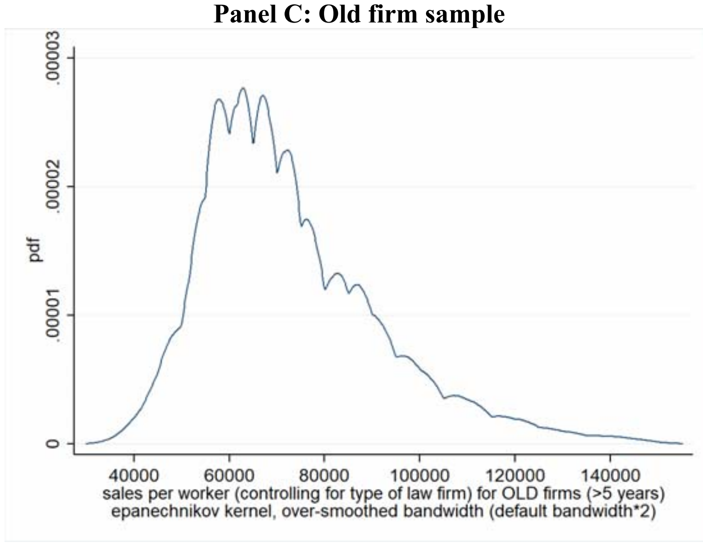
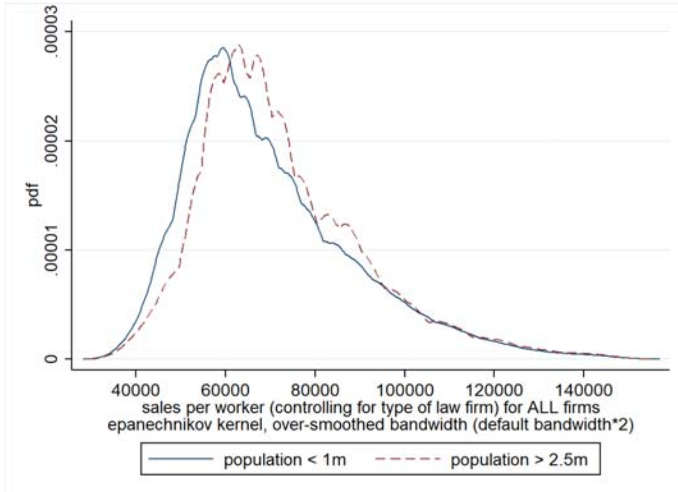
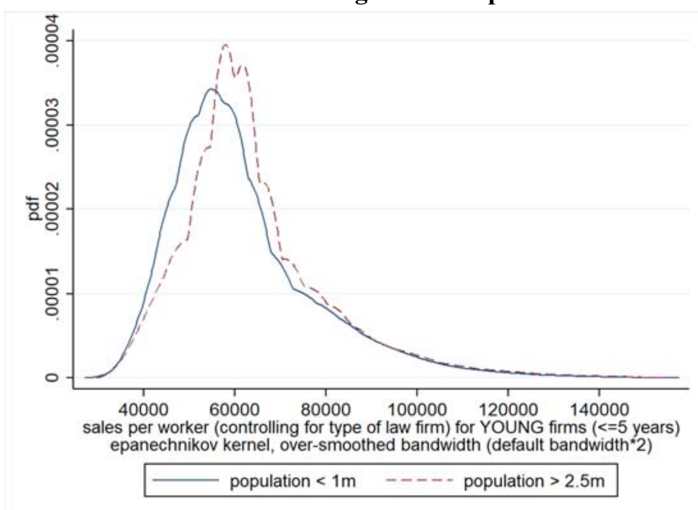


Figure 7b: Sale per worker kernel density estimation for law firms in small versus large cities

Panel A: All firms sample



Panel B: Young firms sample



Panel C: Old firm sample

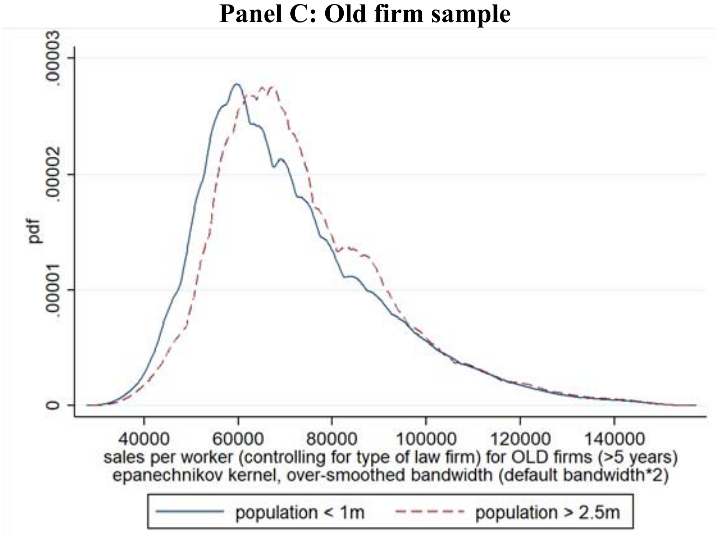
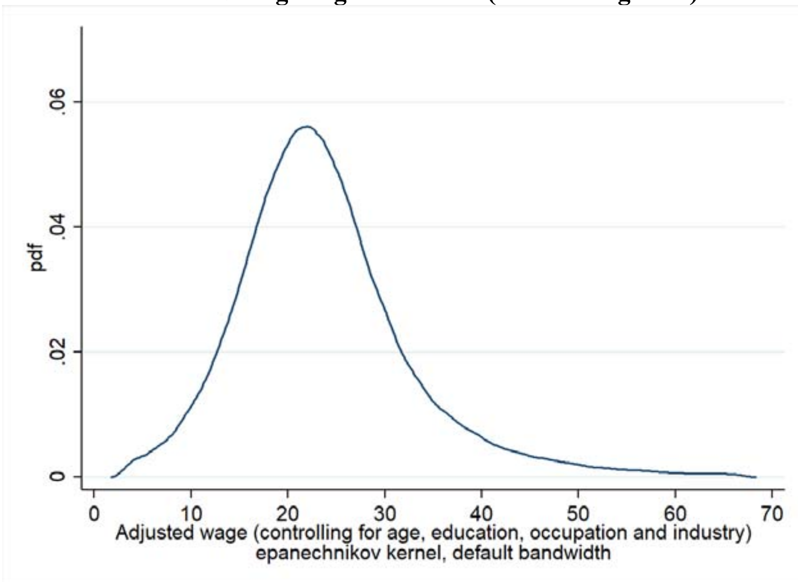
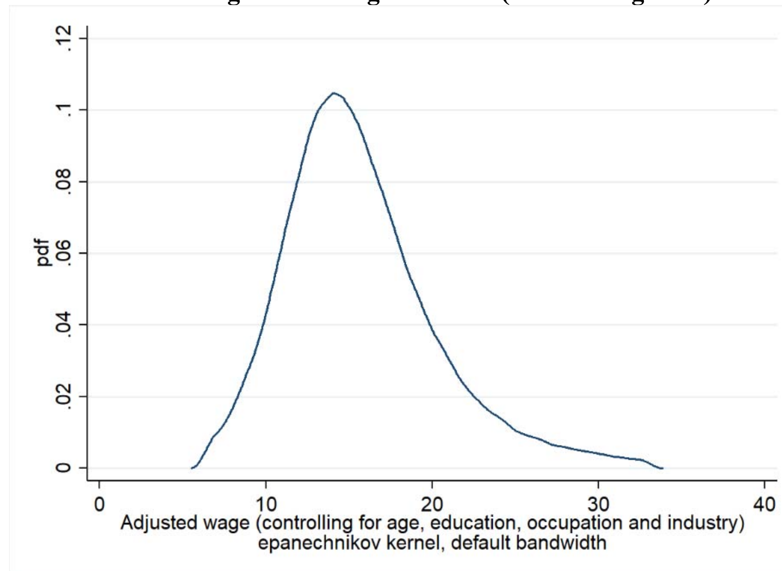


Figure 8: Adjusted wage kernel density estimation for married female non-Hispanic native-born white full-time workers, age 25-54

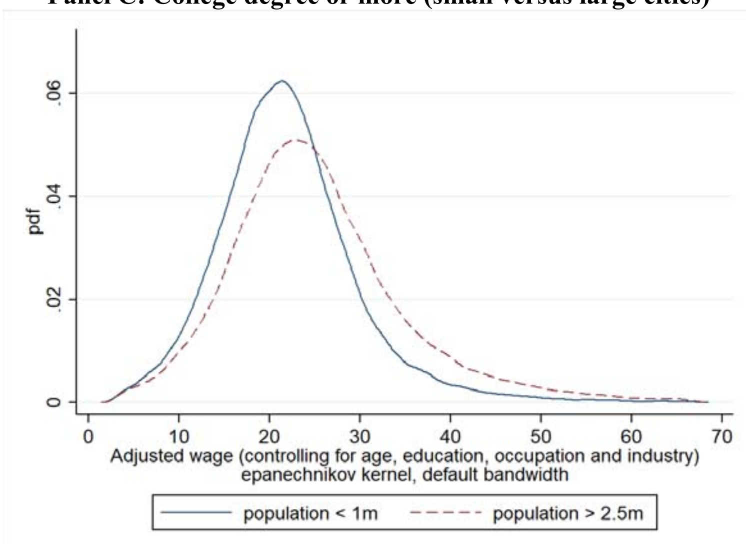
Panel A: College degree or more (all cities together)



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)

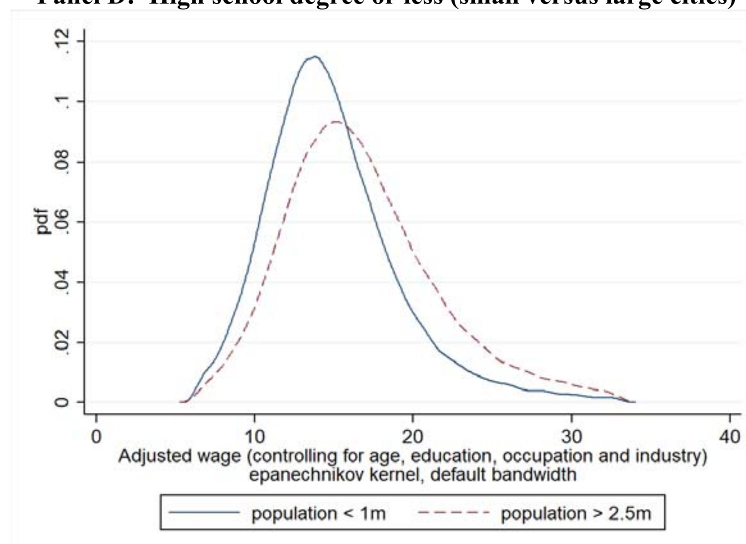
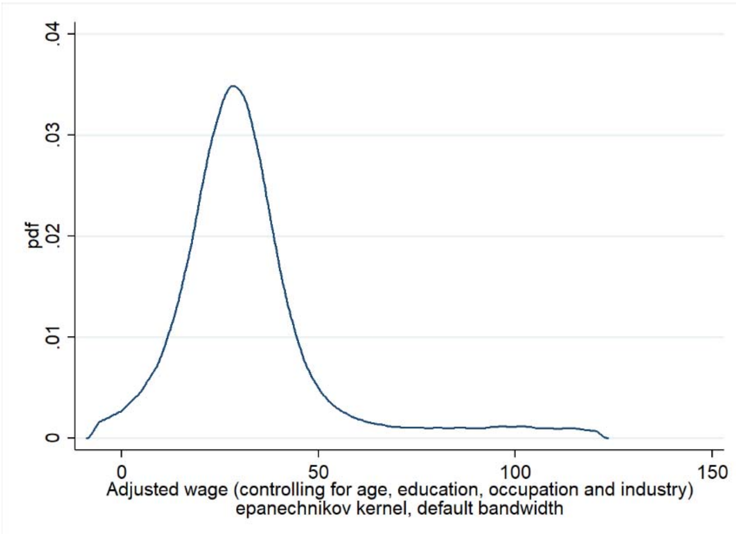
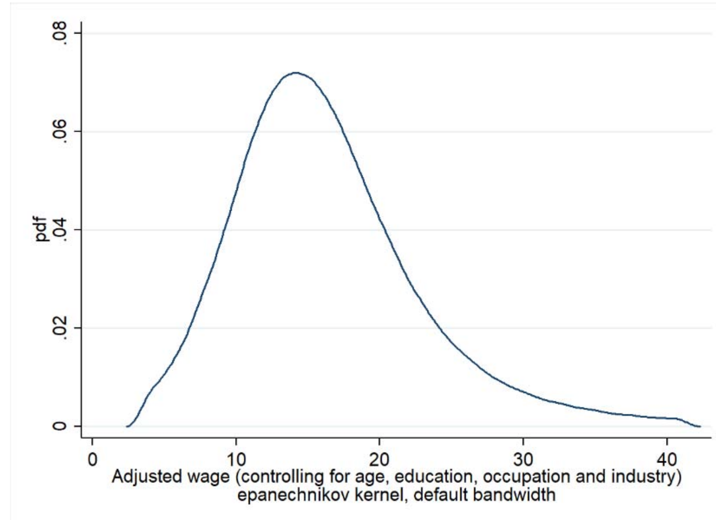


Figure 9: Adjusted wage kernel density estimation for male non-Hispanic native-born white full-time worker, age 25-54

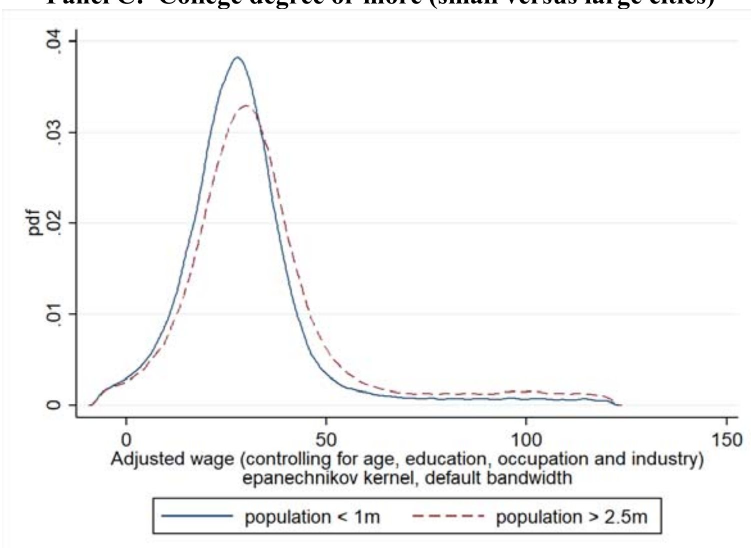
Panel A: College degree or more (all cities together)



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)

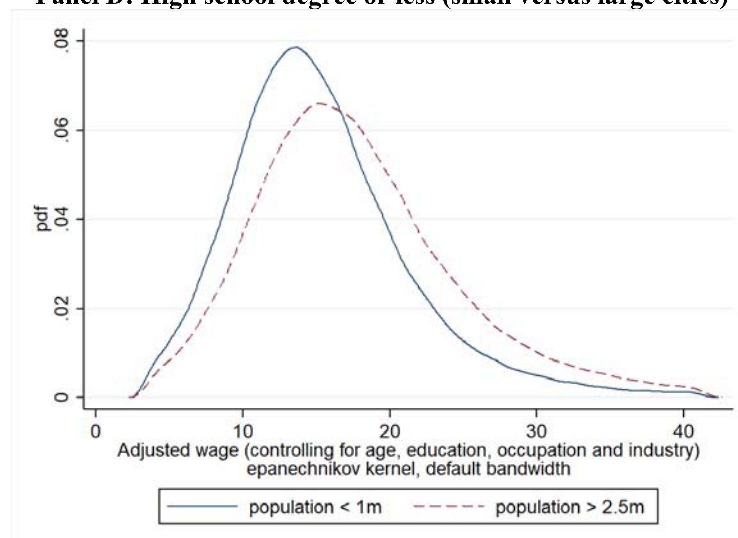
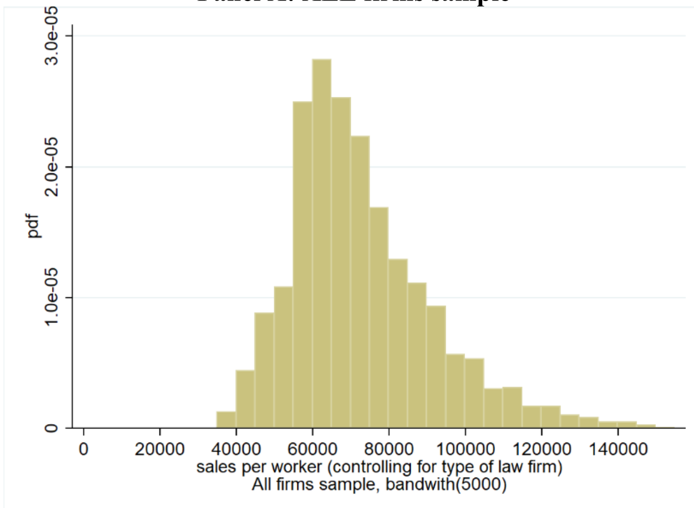
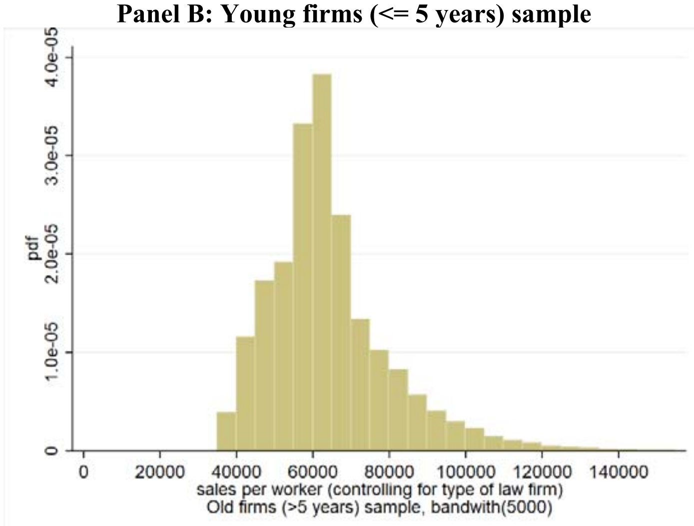


Figure 10a: Histogram estimation of sale per worker for law firms (bandwidth $5,000)

Panel A: ALL firms sample



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



Panel C: Old firms (> 5 years) sample

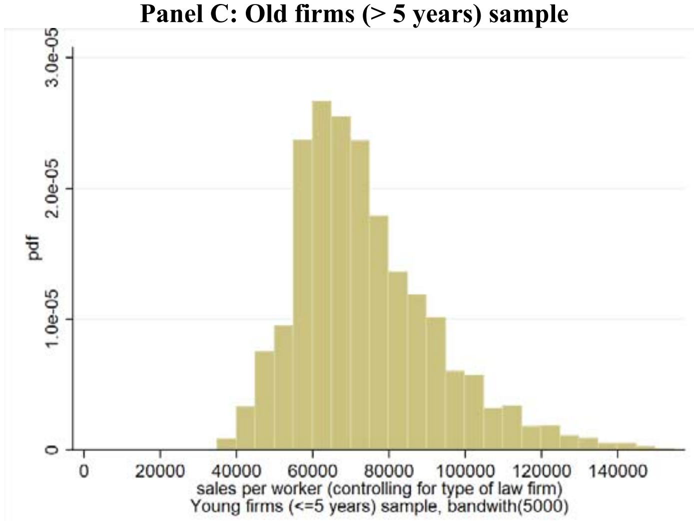
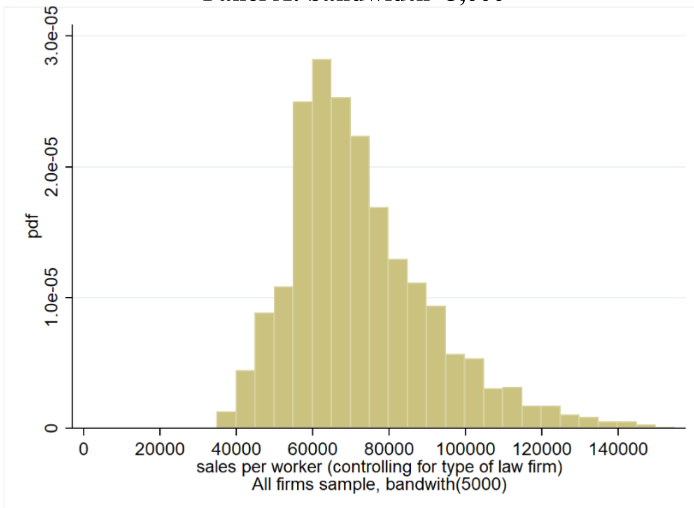
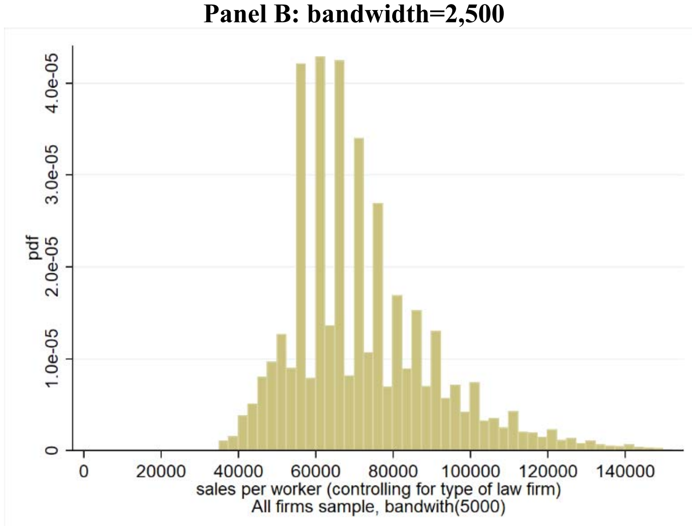


Figure 10b: Histogram estimation of sale per worker for law firm using different bandwidth

Panel A: bandwidth=5,000



Panel B: bandwidth=2,500



Panel C: bandwidth=7,500

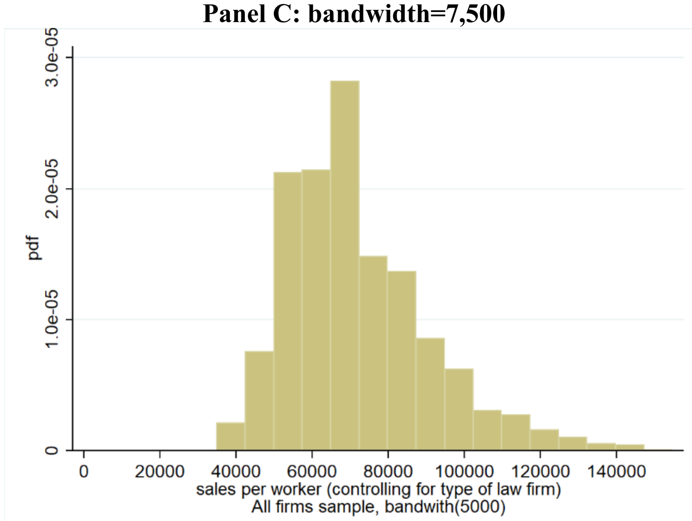


Figure 11a: Histogram estimation of adjusted wage for married female sample (bandwidth $3)

Panel A: College degree or more

Panel B: High school degree or less

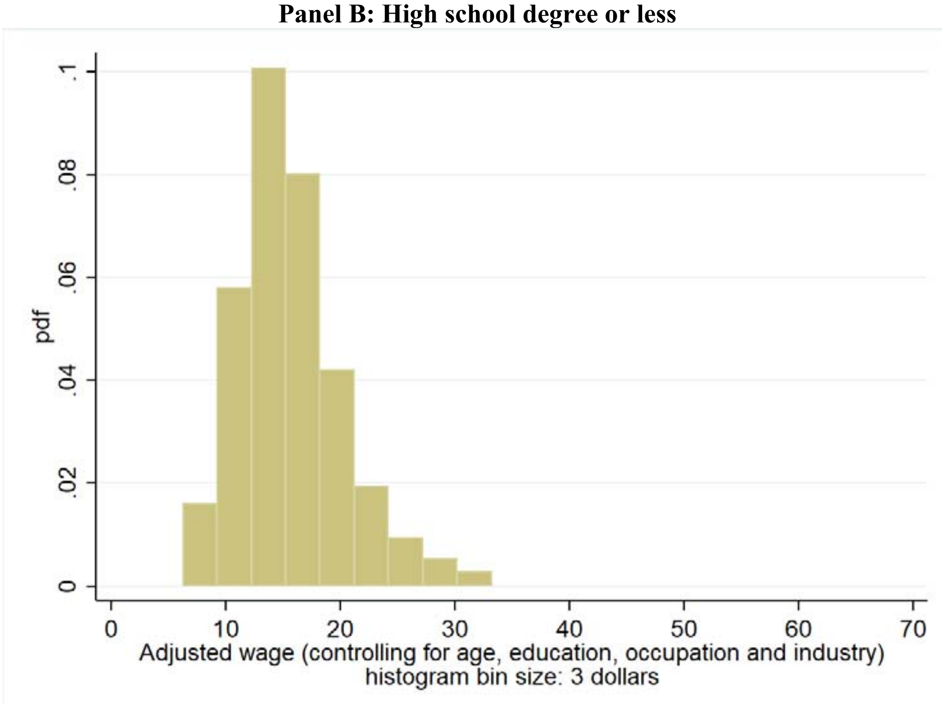


Figure 11b: Histogram estimation of adjusted wage for skilled (college degree or more) married female sample using different bandwidth

Panel A: bandwidth=3

Panel B: bandwidth=1

Panel C: bandwidth=5

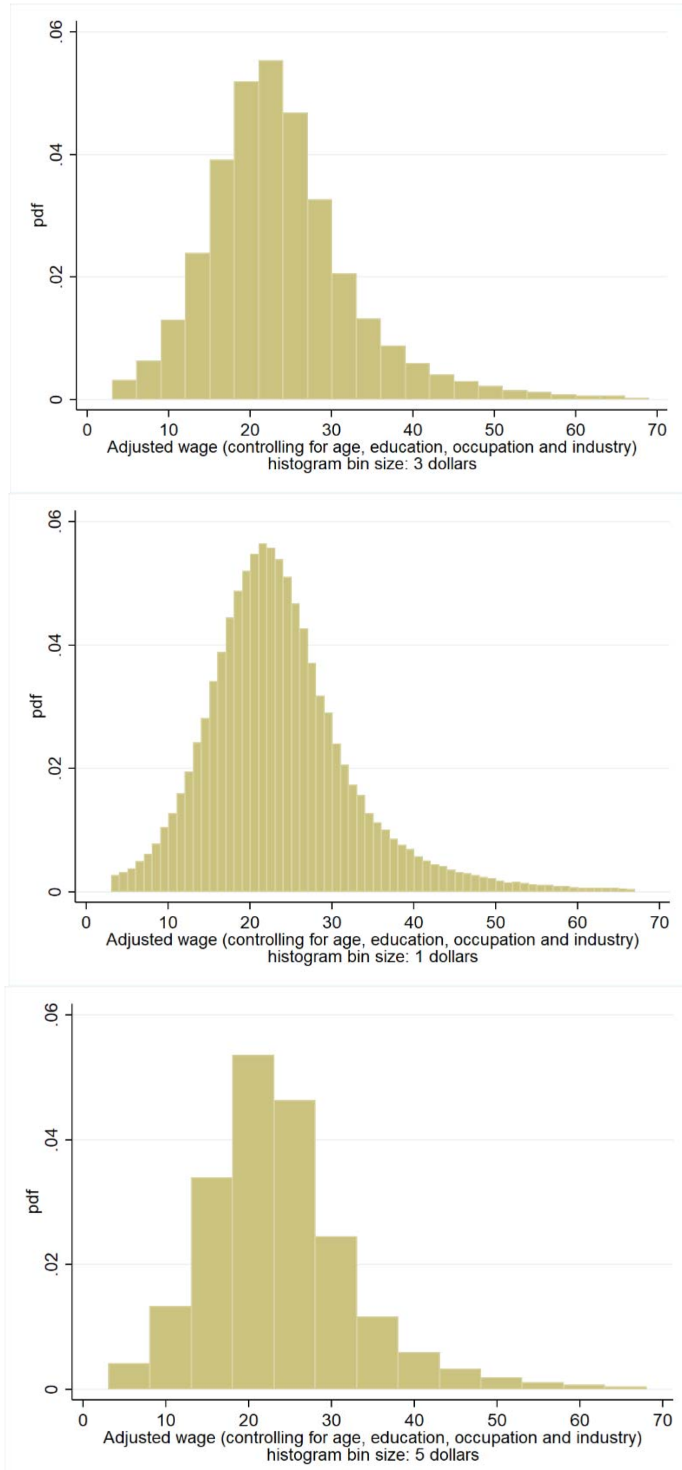
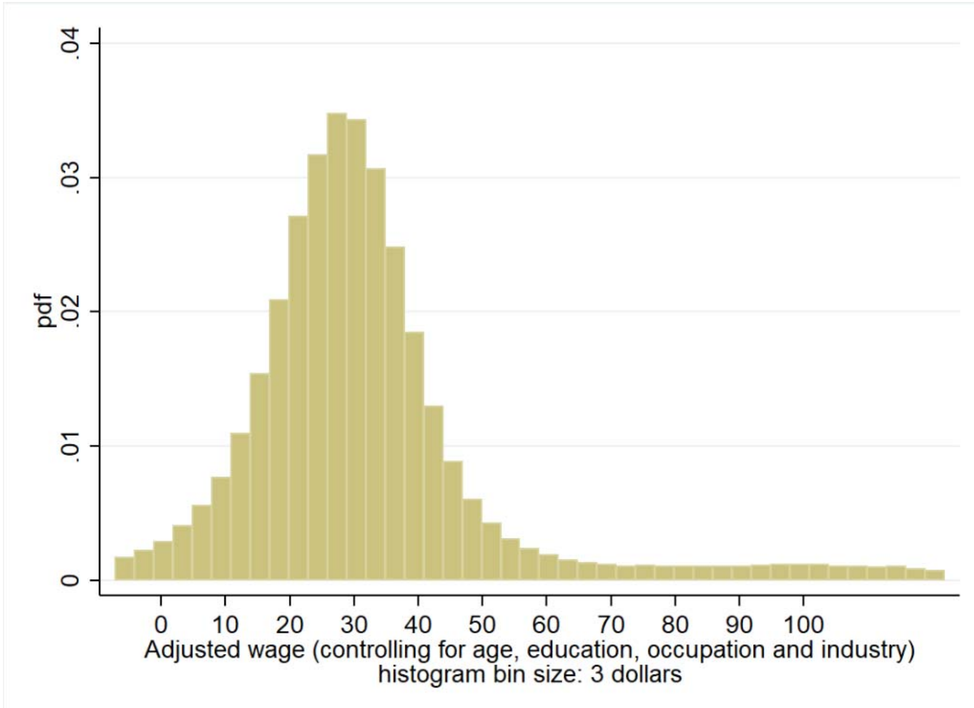


Figure 12a: Histogram estimation of adjusted wage for male sample (bandwidth $3)

Panel A: College degree or more



Panel B: High school degree or less

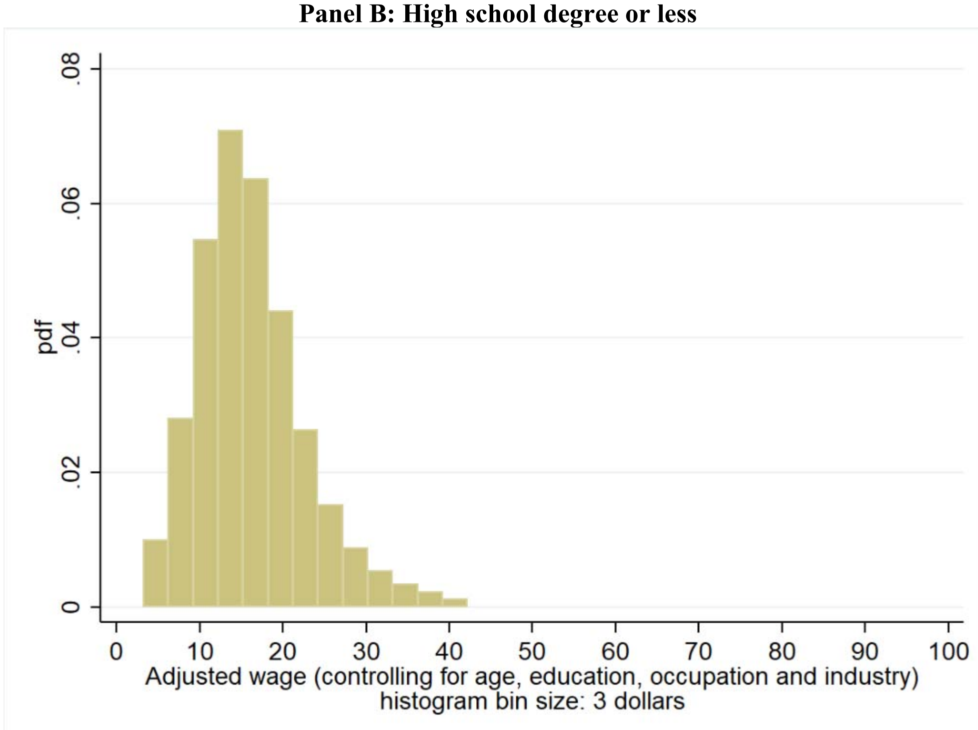
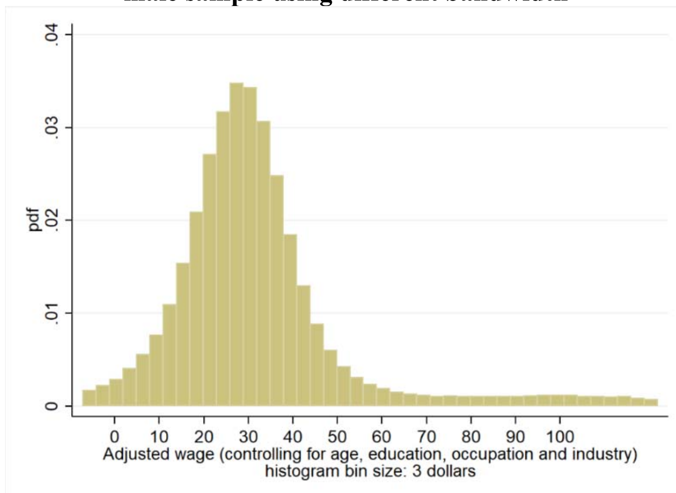
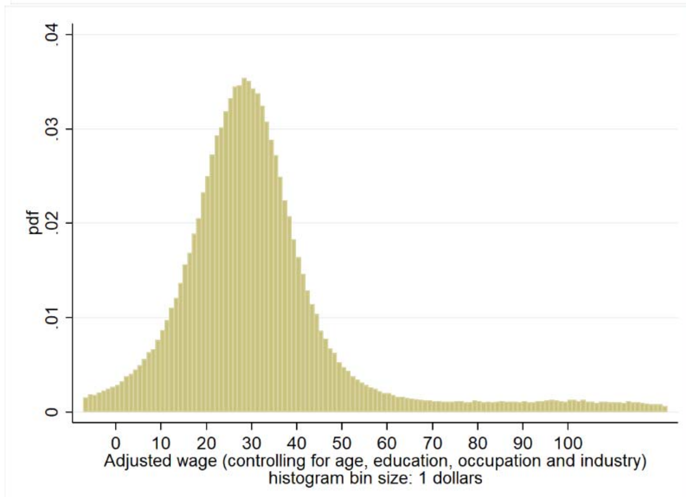


Figure 12b: Histogram estimation of adjusted wage for skilled (college degree or more) male sample using different bandwidth

Panel A: bandwidth=3



Panel B: bandwidth=1



Panel C: bandwidth=5

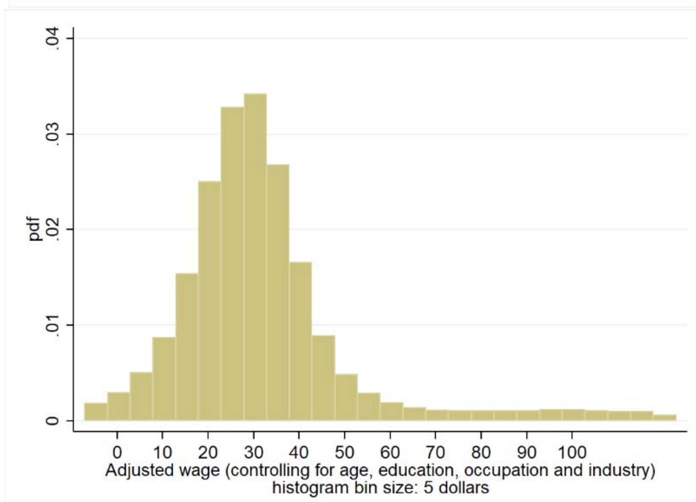
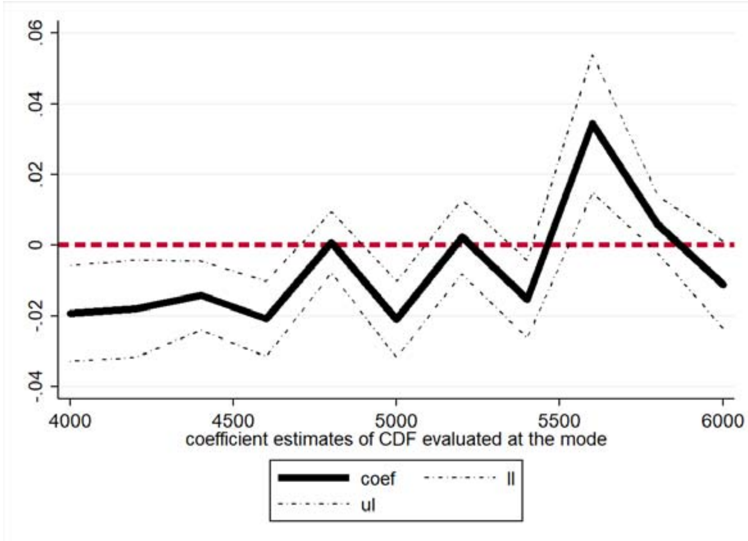
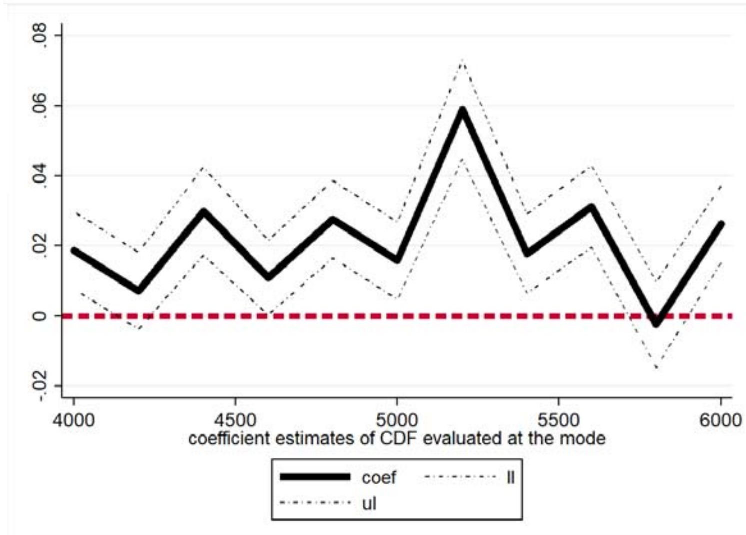


Figure 13a: Law establishment modal CDF robustness check using different bandwidth

Panel A: ALL firms sample.



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



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

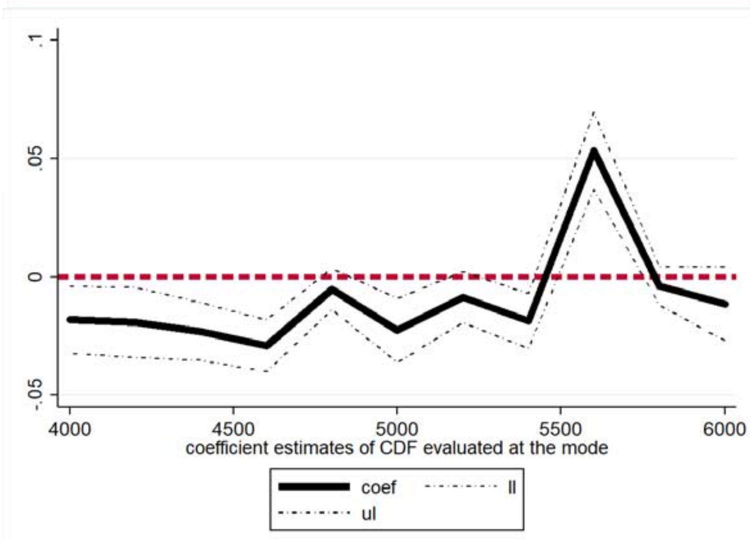
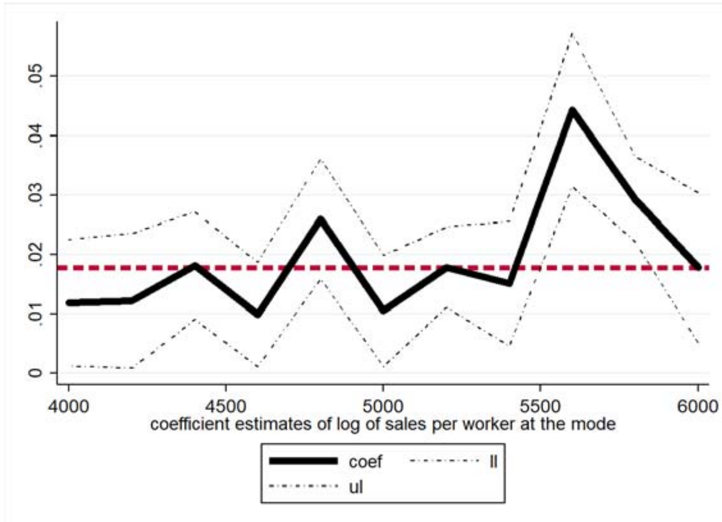
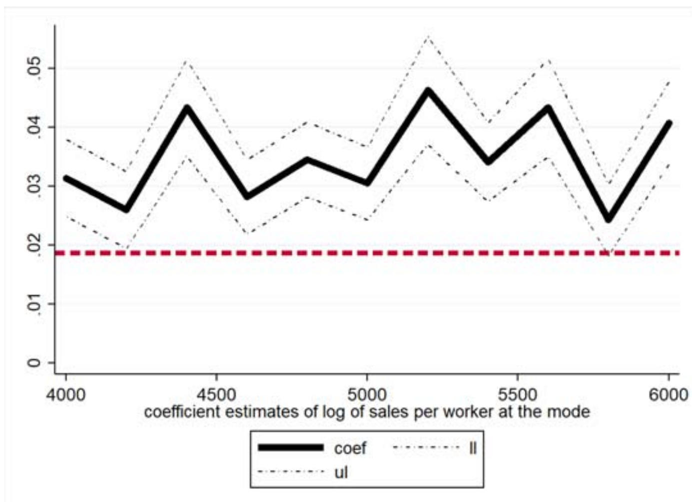


Figure 13b: Law establishment modal return robustness check using different bandwidth

Panel A: ALL firms sample.



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



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

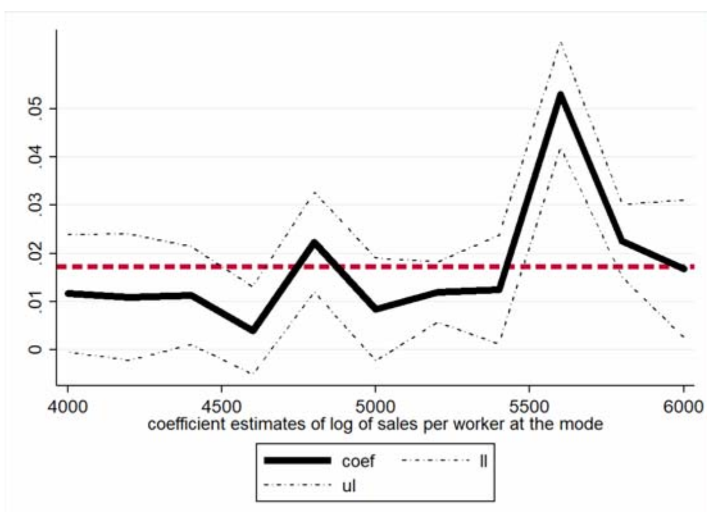
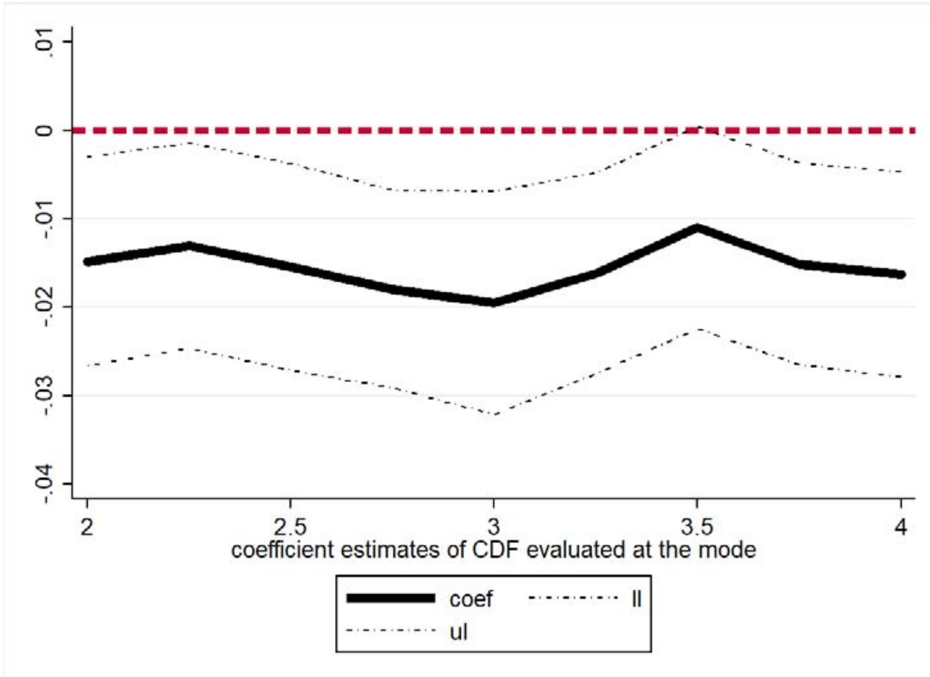


Figure 14a: Married sample modal CDF robustness check using different bandwidth

Panel A: College degree or more



Panel B: high school degree or less

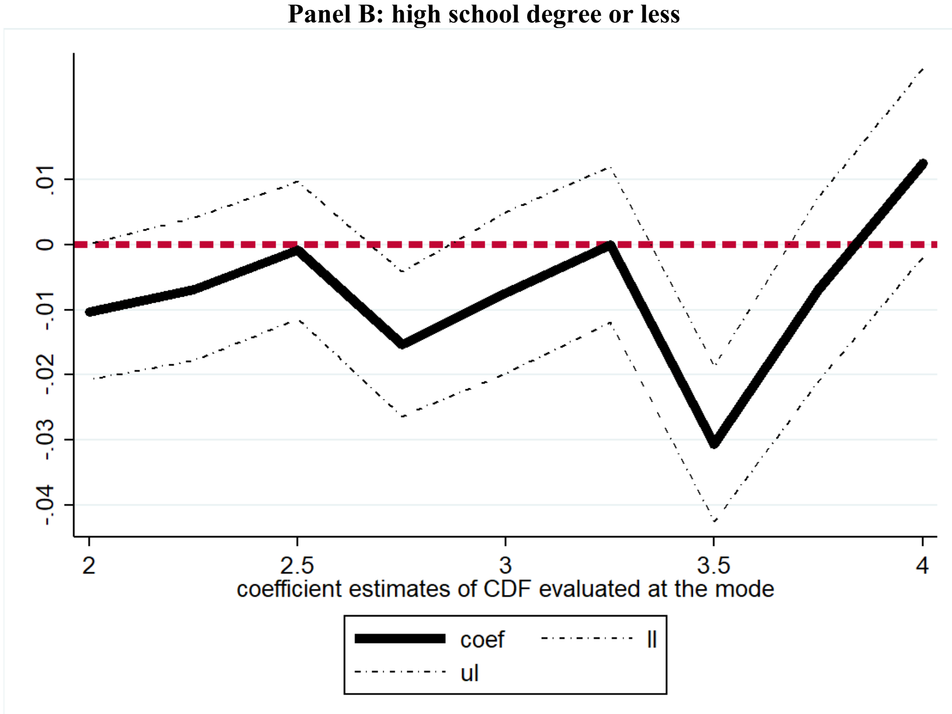
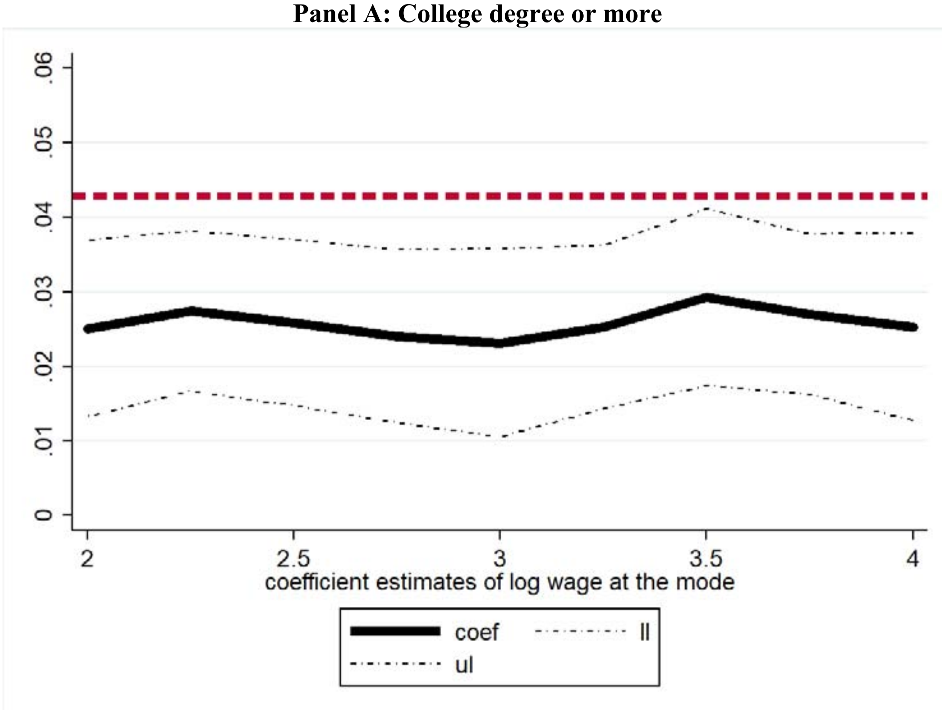


Figure 14b: Married female sample modal return robustness check using different bandwidth

Panel A: College degree or more



Panel B: high school degree or less

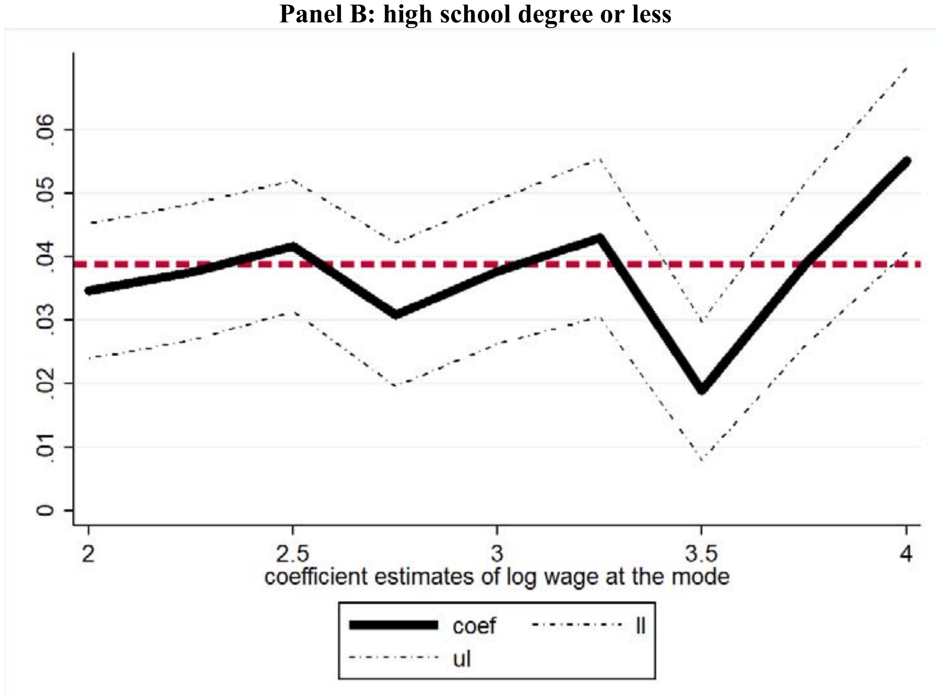
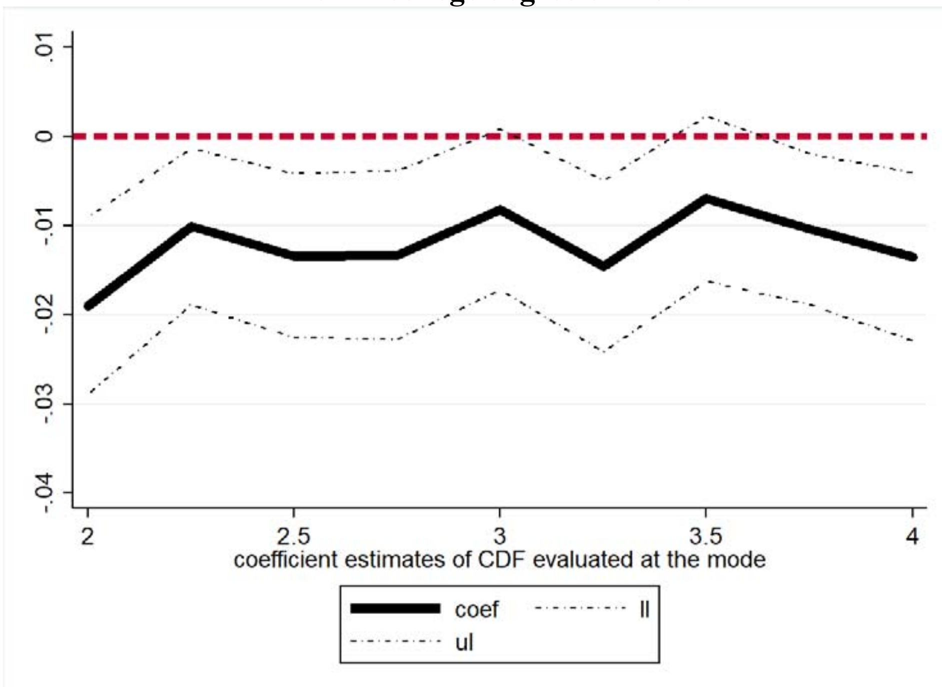


Figure 15a: Male sample modal CDF robustness check using different bandwidth

Panel A: college degree or more



Panel B: high school degree or less

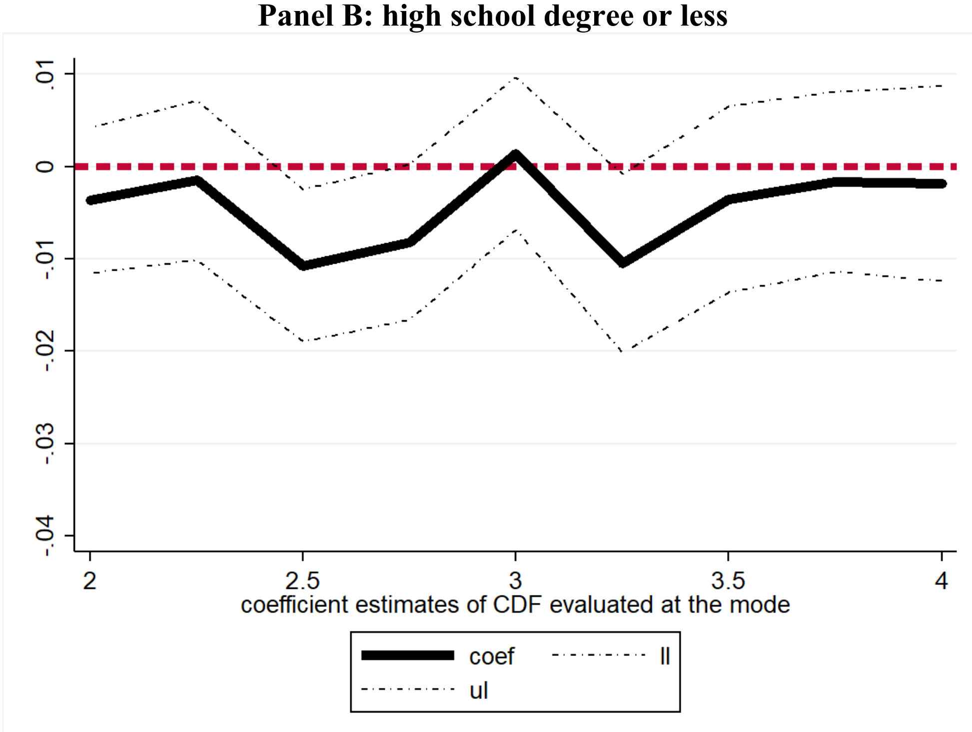
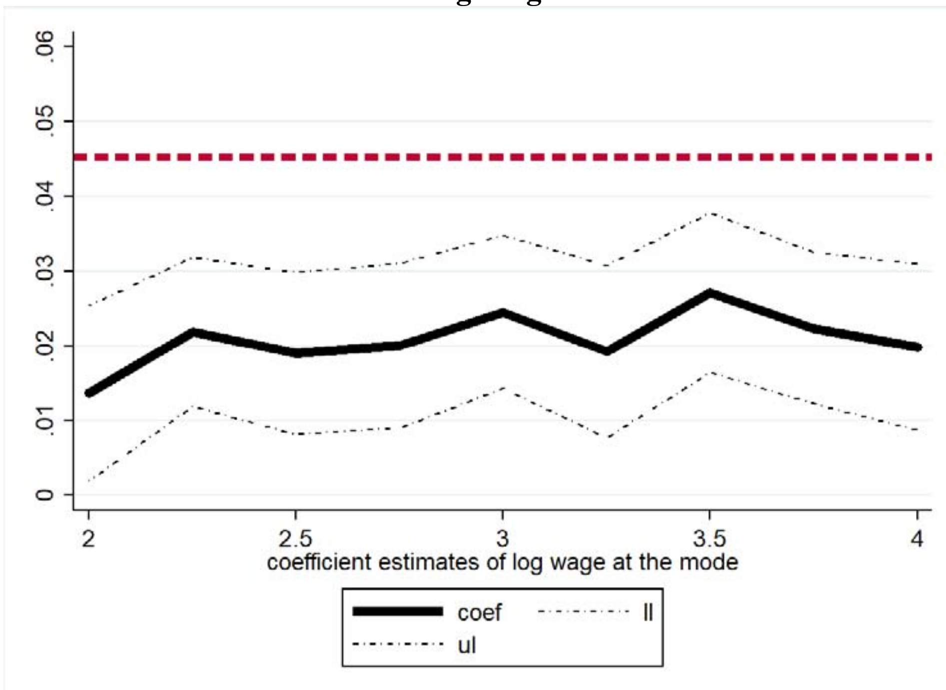


Figure 15b: Male sample modal return robustness check using different bandwidth

Panel A: college degree or more



Panel B: high school degree or less