Forthcoming: *Journal of Urban Economics*

JUE Insight: Using the Mode to Test for Selection in City Size Wage Premia

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

Boqian Jiang
Amazon Inc., Seattle
boqianj@gmail.com

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

July 26, 2022

Helpful comments are gratefully acknowledged from seminar participants at Camp Econometrics, the Urban Economic Association conference in Vancouver, the Institute for Economics at Barcelona, AREUEA International conference at Jinan University, the NBER urban workshop, the Cleveland Federal Reserve Bank, and the University of Colorado. We also thank Nate Baum-Snow, Pierre Philippe Combes, Laurent Gobillon, Tom Holmes, Carlos Martins-Filho, Daniel McMillen and Ronni Pavan for helpful comments. All errors are our own.

Abstract

The causal effect of city size on urban wage premia has been difficult to measure because unusually skilled workers may select into large city labor markets. We propose a new approach to this challenge. For single-peaked wage distributions, if individuals left of the mode disproportionately select out of large city labor markets, the CDF evaluated at the mode shrinks as city size increases. Among college trained, white full-time US workers, evidence of selection is present even after conditioning on extensive observable attributes. Among individuals with a high school degree or less, selection is absent. Additional estimates indicate that for college trained workers, 3.5% is an upper bound on the modal worker's wage elasticity with respect to city size. For those with limited education we can be more precise: modal wage elasticity is 3.9% for men and 5.2% for married women.

JEL Codes: R00, J30, C24
Key Words: Selection, Mode, Agglomeration Economies

1. Introduction

Despite broad consensus that cities enhance productivity (see Rosenthal and Strange (2004, 2020) and Combes and Gobillon (2015) for reviews), the magnitude of this effect is difficult to confirm because unusually skilled workers may select into large city labor markets.[1] Motivated by studies that draw on the shape of factor return distributions for identification (e.g. Saez, 2010; Combes et al, 2012; Eeckhout, Pinheiro and Schmidheiny, 2014; Kleven, 2016; Jales and Yu, 2017; and Jales 2018), we propose a new approach to test for whether selection is present and to estimate the effect of city size on a typical worker's wage. We show that for single-peaked wage distributions with a well-defined interior mode, if workers left of the mode disproportionately drop out of large city labor markets the CDF evaluated at the mode will shrink as city size increases, and sensitivity of modal wage to selection diminishes as the mode becomes more prominently shaped.

We highlight four populations in the U.S. that are likely to differ in their sensitivity to selection because of differences in labor force participation and mobility across MSAs.[2] These include married women with a college degree or more, married women with a high school degree or less, and analogous samples for men (grouping married and single individuals together). For prime age individuals, labor supply is elastic for married women but inelastic for men (e.g. Heim, 2007; Blau and Kahn, 2007). Highly educated individuals may also be more willing to relocate across MSAs (e.g. Molloy et al, 2011; Balgova, 2022), in part because they are more likely to benefit from skill bias in large city labor markets that has become more prominent in recent decades (e.g. Autor, 2019). These mechanisms suggest that selection is most likely to be present for married college educated women and least relevant for men with limited education, priors that we consider empirically.

[1] Throughout, we remain agnostic as to the underlying mechanisms that may drive higher returns in cities, focusing instead on selection. Other studies highlighted in the review papers above consider the nature of underlying microfoundations of agglomeration economies and how that contributes to different patterns of agglomeration and productivity effects across industries and workers. See also Duranton and Puga (2001) for an industry-based example and Bacolod et al (2009) for an example that focuses on different types of individuals.
[2] As an example, Black et al. (2014) argue that higher commuting costs in large cities discourage married women from working. Costa and Kahn (2000) provide evidence that dual career college educated couples are more likely to locate in large cities to help address job market co-location challenges.

Throughout, our focus is on the effect of unobserved factors that affect wage as it is unobservables that drive concern about selection. Using individual level data from the 2000 U.S. census, we begin by creating relatively homogeneous samples with observations restricted to full-time working individuals, age 25-55, who are white, non-Hispanic and native born. For each of our four samples, log wage is regressed on age fixed effects, occupation fixed effects, industry fixed effects, years of schooling fixed effects (for the high school or less samples) and marital status (for men). The residuals from these regressions capture unobserved factors that affect wage and are used in all of the analysis that follows. To facilitate comparison of summary statistics across our four samples, we add the sample-specific raw log wage mean values to the residuals. This preserves sample means but has no effect on our primary estimates because the variance and shapes of the residual distributions are driven entirely by unobservables. Formed as above, the adjusted residuals are sometimes referred to as conditional wage or simply wage.

For each of our four samples, the conditional wage distributions are single peaked with prominent interior modes in small (less than 1 million) and large (more than 2.5 million) urban areas. This is evident in Figure 1 which displays density plots having converted the log wage measures to wage levels. Plots are in separate panels for the four demographic groups highlighted above, each of which includes a separate density for small and large MSAs. If the underlying latent densities were not single peaked, then selection would be the primary driver of the single peaked shape of the densities in the figure. Although we cannot rule out that possibility – latent densities are unobserved – the pronounced interior mode for each of the densities in Figure 1 suggests that the underlying latent densities are similarly shaped.[3]

Bearing this in mind, in the simplest case, if only workers to the left of the mode drop out of large city labor markets, the mode does not shift. In more general cases, selection extends to the right of the mode. In that instance, for a linear, monotonic selection process, elasticity conditions developed later in

[3] It is worth noting, that wage and earnings data have long been recognized as single peaked, examples of which appear in Chotikapanich et al (1997), Clementi and Gallegati (2005), Lopez et al (2006), Sala-i-Martin and Pinkovsky (2009), Eeckhout, Pinheiro and Schmidheiny (2014), and many other studies. Density plots of college and graduate school entrance exam scores (e.g. SAT and GRE) are also strongly single peaked.

the paper determine the extent of modal shift. Moreover, that shift shrinks to zero as the mode becomes increasingly sharp (as with a Laplace distribution). These principles point to three complementary regressions that we estimate. The first regresses the CDF of modal wage in the city on log city size. This regression serves as a diagnostic tool to discern whether selection is present. The second and third regressions replace the dependent variable with log modal wage in the city and log mean wage in the city, respectively. These regressions shed further light on how the mode can be used to evaluate the effect of city size on worker wage.

Results from the CDF regressions yield compelling evidence that among college trained workers, including both married women and men, selection contributes to higher wages in larger cities. Among individuals with a high school degree or less, evidence of selection is absent. These patterns are broadly consistent with our priors. They also echo patterns highlighted by Autor (2019). Autor documents the increase in skill bias in large U.S. cities in the last several decades, with a growing role for college trained, knowledge-oriented workers. He also shows that the nature of work for noncollege workers has shifted dramatically in large cities, becoming increasingly similar to work performed in small cities, and more reliant on generic as opposed to differentiated skill. These trends may be contributing to differences in selection that we estimate for college and noncollege workers.

Estimates from the mode and mean wage regressions are also revealing. For college educated workers, the elasticity of modal wage with respect to MSA size is roughly 3.5% for married women and for men. Having confirmed that selection is present in both samples, that estimate should be interpreted as an upper bound on the effect of MSA size on wage for the typical (modal) worker.

For these same two samples, estimates based on log mean wage in an MSA are notably higher. The estimates are 5.78% and 5.47% for college educated married women and for men, respectively. The larger elasticities for mean wage could arise from two sources. One is selection. Given evidence of strongly single-peaked distributions with interior modes (in Figure 1), selection likely pulls the mean up relative to the mode. A second mechanism is dilation which would occur if high performing individuals based on unobserved factors derive greater productivity boosts from operating in a larger MSA (e.g.

Combes et al, 2012; Gaubert, 2018). This would also pull mean wage up relative to modal wage. Although we cannot confirm whether dilation is present in these samples (for reasons described later), our estimates suggest that selection is contributing to the spread between mode and mean wage effects.

The pattern for those with a high school degree or less is different. As noted above, for this group, evidence of selection is absent, both for married women and for men. The estimated elasticities of modal and mean wage with respect to MSA size are also quite similar, 5.23% and 5.26% respectively for married women and 3.94% and 4.00% for men.[4] We show later that absence of selection effects and similar estimates of the wage elasticities at the mode and mean implies that dilation is not present. This allows us to be more definitive about city size effects: among workers with limited education, an approximate doubling of city size increases wage for the typical (modal) married woman by roughly 5% and among men by roughly 4%.

At face value, the wage elasticities above suggest that the urban premium for a modal noncollege worker is higher than for a modal college trained individual. This does not, however, necessarily indicate that workers with limited education derive a greater productivity boost from city size. Instead, as suggested by Autor (2019), if noncollege individuals are not notably more productive in large cities, they may require compensation to remain in large urban markets where they serve as complements to college trained workers. This is because large cities tend to be expensive.

Using the mode as above has limitations. Most prominently there must be a prior that the underlying latent density is single peaked with a well-defined interior mode. This is consistent with Figure 1 and characteristic of considerable economic data. The modal individual or group must also be informative. For symmetric distributions, the mode is the same as the mean and the median. More generally, the mode represents the most common occurrence. In politics, this could sometimes be a candidate's "base" whose support is needed to secure a majority. In education, resource constrained schools may skew teaching and related opportunities towards students with more typical (modal)

[4] In Table 2, t-ratios on these differences are 0.08 and 0.15, confirming that the differences are not significant.

attributes. In such instances where the most common individual matters for policy or behavior, the mode will tend to be of intrinsic interest.[5]

Several alternate approaches have been used to address possible selection of unusually skilled workers into larger cities, each with advantages and challenges. Eckert et al (2022), De la Roca (2017), De la Roca and Puga (2017), and Glaeser and Mare (2001) use large individual-level panel data files for Denmark, Spain, and the U.S., each of which allows the authors to control for person fixed effects and model movement into and away from large cities. Person fixed effects do much to capture unobserved skill of an individual, but this approach is not possible for widely accessible cross-sectional data. Pseudo-random experiments have also been used to help identify causal effects. Eckert et al (2022) document outcomes for refugees randomly assigned to different cities in Denmark; Ahlfeldt et al (2015) consider changes after the Berlin Wall came down, and Greenstone et al (2010) compare outcomes in locations selected by major companies as compared to runner-up locations that were nearly chosen. Pseudo random events are not always available, however, and are often idiosyncratic in ways that may limit generalization to other settings. Rosenthal and Strange (2008) instrument for local agglomeration using geologic features as cost shifters for construction of tall buildings. Instrument validity must still be defended, however, and instrument variation may not be sufficient in some locations (e.g. landslide hazard is similar throughout much of the U.S. Midwest). Gaubert (2018), Ahlfeldt et al (2015) and Baum-Snow and Pavan (2012) use structural methods to estimate productivity gains from agglomeration and proximity. These papers stand out for their successful integration of economic theory with estimation. Applying structural methods, however, is not straightforward as it requires careful balancing of parsimony that is needed to make the models tractable with enough generality to allow for robustness.

[5] Our focus on the mode has precedents. Lee (1989) showed that the mode from a truncated distribution can be a consistent estimate of the conditional mean from the non-truncated distribution. Our work is also related to modal regression literature in statistics and econometrics (Lee, 1993; Kemp et al, 2012; Huang et al. 2013; Yao and Li, 2014; Chen et al., 2016, Honoré, 1989). Cardoso and Portugal (2005) show that modal wage is a better measure of the central tendency of a 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.

Different from the strategies above, our use of the mode is most similar to the approach adopted by Combes et al (2012), which is among the few studies in the agglomeration literature to draw on the shape of the factor return distribution.[6] Combes et al (2012) divide French cities into those below versus those above median size. Focusing mostly on manufacturing plants, they minimize mean squared deviations between TFP quantiles for plants organized into the two groups of cities. Their model is designed in a manner that identifies parameters of the underlying city size productivity function, including a shift factor and dilation, in addition to a TFP threshold value below which companies select out of large city markets. An appealing feature of the Combes et al (2012) model is that it does not impose any structure on the underlying latent productivity distribution although it does embed a particular selection process in the estimation routine which makes their estimates potentially sensitive to departures from that structure. While our model based on the mode is not as general, it can be easily implemented for cross-sectional datasets, drawing on variation across many locations. This makes the mode an accessible diagnostic tool to test for the presence of selection when using cross-sectional data, and in some instances, a straightforward way to estimate the effect of city size.

2. Model

2.1 Mode location when city size enhances productivity and selection is present

This section establishes conditions that determine the location of the mode in each city's wage distribution when city size enhances productivity, selection sorts workers between cities, and the underlying aggregate latent wage distribution is single peaked with a well-defined interior mode. We begin with notation.

In the discussion below, y is a worker's skill endowment, which is interpreted in logs, with $y \geq 0$. The term s is city size, also in log form, where $s = 0$ denotes the smallest city in the system. We use $f_s(y)$ to represent the density of log productivity among individuals in a city of size s ($s \geq 0$), where $f_s(y)$ is

[6] See also Eeckhout, Pinheiro and Schmidheiny (2014) for related work.

assumed to be continuous and differentiable. For a given skill endowment, productivity varies with city size in a manner described below and is given by y_s.

Three assumptions drive features of our model that help to guide and interpret the empirical work that follows. The first is that large city productivity gains take a linear form that preserves an individual's productivity rank within a given city, similar to Combes et al (2012), Gaubert (2018), and Firpo (2007).[7]

Assumption 1: For a given skill endowment *y*, productivity in a city of size *s* is given by:

$$y_s = \beta_0 s + (1 + \beta_1 s)y, \text{ where } \beta_0 \geq 0 \text{ and } \beta_1 \geq 0. \tag{2.1}$$

In (2.1), in the absence of city size effects, $\beta_0 = \beta_1 = 0$ and $y_s = y$, in which case an individual's productivity is the same in all cities. If instead increasing city size increases productivity by a common percentage for all workers, $\beta_0 > 0$, $\beta_1 = 0$, and $f_s(y)$ shifts to the right as *s* increases. If in addition $\beta_1 > 0$, the productivity gains from city size increase with an individual's endowment, *y*. This creates a dilation effect so that as city size increases, $f_s(y)$ becomes right skewed with an elongated right tail.

For (2.1) above, the cumulative distribution function (*F*) for productivity up to a given skill endowment, y, is the same in each city, denoted as F_0 for the smallest city and F_s for a city of log size *s* for $s > 0$,

$$F_s\big(y_s(y)\big) = F_0(y) \tag{2.2}$$

Substituting for y_s from expression (2.1) and taking derivatives with respect to *y*, the relationship between large and small city productivity densities is given by,

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

Our next modeling assumption is to presume a linear monotonic selection process that governs participation in different size city labor markets.

Assumption 2: The probability of selecting into and participating in a size-*s* city labor market is given by:

[7] The *rank preservation* assumption is standard in the literature on quantile treatment effects. See Firpo (2007) and the references therein for discussion.

$$\pi_s(y) = \begin{cases} a + by, \text{ for y} < y_p \\ a + by_p \equiv p \leq 1, \text{ for y} \geq y_p \end{cases} \tag{2.4}$$

where $0 \leq \pi_s(y) \leq 1$ and $b > 0$.

In (2.4), the term $\pi_s(y)$ is indexed by s to indicate that the selection probability for a given y differs with city size. Implicitly, a, b, and y_p are also indexed by s but the subscript on these terms is suppressed for simplicity. Assumption 2 also makes explicit that $\pi_s(y)$ reaches an asymptote at y_p with $p \leq 1$ since $\pi_s(y)$ can never exceed 1. This assumption reflects in part the sense that operating costs are higher in larger cities and/or the environment more competitive. This discourages weaker workers from participating in larger city markets (e.g. Combes et al (2012)).

Allowing for selection effects as above and also the possibility of productivity spillovers, expression (2.3) becomes,

$$f_s(y) = \frac{\pi_s(y)}{(1+\beta_1 s)c_s} f_0\left(\frac{y-\beta_0 s}{1+\beta_1 s}\right) \tag{2.5}$$

where $c_s = \int \pi_s(u) f_0(u) du$ ensures that the density in the city s integrates to 1.[8]

Suppose now that selection effects are present but there are no productivity gains from agglomeration. Then $\beta_0 = \beta_1 = 0$, and the density in (2.5) becomes,

$$f_s(y) = \frac{\pi_s(y)}{c_s} f_0(y) \, . \tag{2.6}$$

Our third and most important modeling assumption concerns the shape of the wage density function:

Assumption 3: *$f_0(y)$ is single peaked with a well defined mode at an interior location.*

Because $f_0(y)$ is assumed to be differentiable and single peaked, its slope at the mode is zero. Differentiating (2.6) with respect to y and setting the derivative to zero, the modal value for y (denoted by y_m) in the conditional density, $f_s(y)$ must satisfy,

$$\psi_{\pi,y} = -\psi_{f,y} \tag{2.7}$$

---

[8] Notice that for the special case when selection is independent of y, $\pi_s(y) = c_s = p$, and in the absence of spillover effects ($\beta_0 = \beta_1 = 0$), expression (2.4) simplifies to $f_s(y) = f_0(y)$.

where $\psi_{\pi,y} \equiv \frac{\pi_s^{'}(y)}{\pi_s(y)}$ and $\psi_{f,y} \equiv \frac{f_o{'}(y)}{f_o(y)}$ are the semi-elasticities of the selection and latent density functions, respectively. Multiplying both sides of (2.7) by y (for $y \neq 0$) this can also be expressed as an elasticity condition,

$$\xi_{\pi,y} = -\,\xi_{f_0,y} \tag{2.8}$$

with $\xi_{\pi,y} \approx \frac{\%\Delta\pi(y)}{\%\Delta y}$ and $\xi_{f_0,y} \approx \frac{\%\Delta f_0(y)}{\%\Delta y}$. Note that these expessions vary with city size but the s subscripts are supressed for simplicity. Expression (2.8) indicates that at the mode of the conditional distribution, a small change in y yields equal magnitude but opposite signed percentage changes in the selection probability and the density of y.

2.2 Magnitude of mode shift

The previous section established the elasticity condition that holds at the mode of a conditional density function. Here we focus on the amount by which selection causes the mode to shift.

The simplest case is when all non-random selection occurs on just one side of the mode, which we will treat as being on the left side given our city size context. In this instance, the CDF evaluated at the mode shrinks in response to selection and the mode does not shift. To understand why, recall that the mode of a single-peaked density function has the highest density. Removing mass only from one side of the mode scales up the density elsewhere in the distribution – including at the original mode – by a common factor so that the conditional density integrates to one. Although the mode itself does not shift, the CDF evaluated at the mode shrinks. In settings where this case applies, it is not necessary to impose any structure on the selection function when using the mode to address selection.

A more general case is when the selection function in (2.4) reaches an asymptote p to the right of the mode at $y_p > y_m$. If the asymptote is reached before the elasticity condition in (2.8) is satisfied, then the mode changes by y_p - y_m. If that is not the case, then further structure on the shape of the density function is needed to clarify the extent of modal shift. With that in mind, in Appendix A we develop additional properties of our model using the generalized error distribution (GED) to characterize the latent skill distribution. The GED is a flexible three-parameter single-peaked function for which the shape parameter governs how sharply defined is the mode.[9]

Several properties of our model are highlighted in the appendix and reinforce intuition already developed. Most important, modal shift declines monotonically as the mode becomes more sharply defined, converging to zero as the density converges to a Laplace distribution. At the other extreme, as the latent density function converges to a uniform, the mode disappears. In that instance, the density $f_0(y)$ becomes a constant f_0 over the relevant range of y. From (2.6) it follows that $f_s(y) = \pi_s(y)\frac{f_0}{c}$. The conditional density function therefore takes on the shape of the selection function scaled by f_0/c. With $b > 0$, the mode of $f_s(y)$ must shift all the way to the right edge of the distribution.[10] These two extremes underscore a central premise: for the mode to be an effective device to address selection, there must be a clearly defined interior mode.

2.3 Estimating equations

As described earlier, we estimate three regressions in the empirical work to follow. The first uses $F(y_{m,i})$ as the dependent variable, the CDF evaluated at the modal wage for MSA i. The other two use modal log wage and average log wage in MSA i as the dependent variables, $y_{m,i}$ and $\bar{y}_i$, respectively. For all three regressions, log population in MSA i is the righthand side control and is represented as s_i. The estimating equations are then given by:

$$F\left(y_{m,i}\right) = c_{0,1} + c_{1,1}s_i + e_i \tag{2.9a}$$

$$y_{m,i} = c_{0,2} + c_{1,2}s_i + e_i \tag{2.9b}$$

$$\bar{y}_i = c_{0,3} + c_{1,3}s_i + e_i \tag{2.9c}$$

[9] The shape parameter in the GED ranges from 0 to infinity. As the shape parameter approaches 0, the GED converges to a uniform distribution and the mode disappears. The normal distribution has a shape parameter of one-half, and the Laplace has a shape parameter of 1.

[10] This is confirmed in Appendix A. For a shape parameter $k = 0$, the GED flattens to the uniform distribution and modal shift equals σ, the scale parameter.

The constants in (2.9a-c) capture features that are not associated with MSA size and are not of interest for that reason. Instead, it is the slope coefficients that are the primary focus. These allow us to infer evidence of whether selection is present and depending on the combination of estimates, provide differing degrees of information about the causal effect of city size on wage. We highlight three scenarios that are of particular interest.

Consider first the case where selection is absent, with $b = 0$ in the selection equation (2.4). This implies that $c_{1,1} = 0$. Taking derivatives of (2.1), (2.9b) and (2.9c) with respect to s yields, respectively, $\frac{\partial y}{\partial s} = \beta_0 + \beta_1 y$, $\frac{\partial y_m}{\partial s} = c_{1,2}$, and $\frac{\partial \bar{y}}{\partial s} = c_{1,3}$. Evaluating the first of these derivatives at both y_m and $\bar{y}$ and substituting from the latter two, we can solve for β_0 and β_1 as,

$$\beta_0 = c_{1,2} - \left[\frac{c_{1,3} - c_{1,2}}{\bar{y} - y_m}\right] y_m \tag{2.10a}$$

$$\beta_1 = \frac{c_{1,3} - c_{1,2}}{\bar{y} - y_m} \tag{2.10b}$$

where $\bar{y}$ and y_m are based on values for a reference city (e.g. the smallest city). Moreover, if $c_{1,2} = c_{1,3}$, then $\beta_1 = 0$ in (2.10b) and dilation in (2.1) is absent. In that case, $c_{1,2} = c_{1,3} = \beta_0$.

Suppose next that $c_{1,1} < 0$ but there is a prior that selection does not extend beyond the mode ($0 \leq y_p < y_m$). In that instance, we do not need to specify the shape of the selection process. Instead, the mode does not shift in response to selection and $c_{1,2}$ measures the causal effect of MSA size on modal wage.[11]

A third and more general scenario arises when $c_{1,1} < 0$ and we cannot rule out that selection extends beyond the mode. In this instance, modal wage will shift right in response to selection. As a result, estimates of $c_{1,2}$ and $c_{1,3}$ are biased upwards and provide upper bounds on the causal effect of MSA size on productivity at the mode and mean, respectively.[12]

[11] In this instance, $c_{1,1}$ would also approximate the loss of mass to the left of y_m as MSA size increases.

[12] A further point is that the degree to which $c_{1,1}$ is less than zero understates whether selection is present. This is because although $c_{1,1} < 0$ indicates that selection removes more mass left of the mode – implying that b in (2.4) is positive – selection will also draw the mode to the right. This secondary effect increases the CDF evaluated at the mode causing $c_{1,1} < 0$ to understate evidence of selection.

A few final points remain that affect the viability of using the mode to address selection. First, large samples are needed so that the mode can be estimated with reasonable precision. This will improve precision of the subsequent regressions in (2.9a) and (2.9b) that use the estimated mode when forming the dependent variable. Second, provided estimation error around the mode is uncorrelated with MSA size, estimates from the modal wage regression in (2.9b) will be unbiased because the estimated mode appears on the lefthand side of the regression. Third, this same point holds as an approximation for the CDF regression in (2.9a). This is because the derivative of the density function at the mode is zero, and so upon taking a second-order Taylor expansion around the CDF at the mode, the second order term drops out. The CDF evaluated at the estimated mode is then approximately equal to its true value plus the same linear error term (scaled by a constant in this instance) as in the modal wage regression.[13]

3. Data, measuring modal wage, and summary statistics

For reasons described in the Introduction, we focus on four samples, married women with a college degree or more, men (grouping singles and married together) with a college degree or more, and analogous samples for those with a high school degree or less. In each case, individual level data were drawn from the 5% 2000 U.S. census, downloaded from IPUMS.[14] For each sample, we restrict observations to full-time working individuals, age 25-55, who are white, non-Hispanic and native born.[15] Full-time workers were coded as those who work at least 35 hours per week and 40 weeks per year.

[13] Taking a second-order Taylor expansion around the true mode and rearranging, the CDF evaluated at the estimated mode can be expressed as below, where $e = \hat{y}_m - y_m$ and F and f are the CDF and density functions, respectively,

$$F(\hat{y}_m) \approx F(y_m) + f(y_m)e + \frac{1}{2}f'(y_m)e^2 \quad . \qquad \text{(N.1)}$$

Because $f'(y_m) = 0$ at the true mode, the second order term in (N.1) drops out leaving:

$$F(\hat{y}_m) \approx F(y_m) + f(y_m)e \quad . \qquad \text{(N.2)}$$

Expression (N.2) indicates that the CDF evaluated at $\hat{y}_m$ is approximated by the CDF evaluated at the true mode $F(y_m)$ plus the same error e as in the modal wage regression scaled by the density at y_m.

[14] See Ruggles et al., 2015. Year-2000 census data can be downloaded at no charge from the IPUMs website at https://usa.ipums.org/usa/ .

[15] Observations from Alaska and Hawaii were excluded.

Hourly wage was computed by dividing annual earnings by annual hours worked. Each sample is further restricted to (i) individuals who earn at least two-thirds of the federal minimum wage in 2000 ($5.15 in year-2000 dollars)[16]; (ii) who live in MSAs with more than 100,000 population in 2000, and (iii) for which 100 or more individuals remain in the sample after these other criteria are applied.[17]

To focus on the effect of unobserved factors, for each sample we regressed log wage level on 15 age fixed effects, 359 5-digit occupation fixed effects, 94 industry fixed effects, years of schooling fixed effects (for the high school or less samples) and marital status (for the male sample). The mean of log wage was then added to the residuals to preserve mean values across the samples. We sometimes refer to the adjusted residuals as conditional wage or simply wage. In instances where we focus on wage levels (as opposed to log values), we exponentiated the log wage measures. All of the analysis that follows is based on these conditional wage measures which were converted to May 2002 dollars to facilitate interpretation.

Having prepared the conditional log wage measures as above, we estimate modal wage values for each of the four main samples for each MSA. This was done using a kernel density estimator with the Epanechnikov kernel function and Silverman's rule of thumb for the bandwidth.

Table 1 provides summary statistics. We focus first on individual level conditional wage rates having pooled data across MSAs. Measures for married women are in Panel A and for men in Panel B. In both cases, values for the college educated and high school or less are in separate rows. Values for the 5th, 25th, 50th, 75th and 95th quantiles are in columns. As would be expected, wage is higher among college educated workers and among men. For the high school or less population, median wage is $20.41 per hour for married women and $26.55 per hour for men. Among the college educated, analogous values are $34.93 and $44.59. More substantive for our purposes, there is much more variation in wage among college educated workers as compared to those with high school or less. Among men, for example, the

[16] The federal minimum wage was $5.15 in 2000 ($8.92 in 2022 dollars). Results were unchanged when we dropped observations with wage below the federal minimum wage.

[17] MSA population is based on the 2000 census and was obtained from IPUMS at: https://usa.ipums.org/usa/resources/volii/MSA2013_PUMA2000_pop2000_crosswalk.xls .

interquartile ranges for the two groups are, respectively, \$27.04 and \$14.2. That difference is driven entirely by differences in unobserved factors.

Panels C and D report summary measures for MSA-specific modal wage rates, where each observation is a single MSA. The distributions mirror those at the individual level, with higher wage rates for college educated workers and for men. Also, there is considerably more cross-MSA variation in modal wage rates among college educated workers as compared to high school or less workers.

4. Estimates

Table 2 reports estimates of expressions 2.9a, 2.9b and 2.9c in columns 1-3, respectively. Column 4 reports estimates based on the difference between the mean and modal wage. In each case, log of MSA population in 2000 is the primary control measure. Constants are included in the regressions but are not reported in the table. Estimates for married women with a college degree and those with a high school degree or less are in Panels A and B, respectively. Corresponding estimates for men are in Panels C and D.

Consider first the evidence of selection across the four panels. This is assessed based on the coefficients on MSA population size in the CDF regression in column 1. For college educated workers, clear evidence of selection is present for both married women and for men. The corresponding coefficients are -0.0122 and -0.0074, with t-ratios of -3.80 and -2.93, respectively. For individuals with a high school degree or less, evidence of selection is absent. For married women the coefficient is -0.0006 while for men the estimate is -0.0007 (with t-ratios of -0.19 and -0.28, respectively). Overall, these estimates conform to our priors.

Focusing on the wage elasticities, estimates for college educated married women (Panel A) and men (Panel C) are nearly identical in magnitude. For modal wage, the elasticities are 3.43% and 3.65%, respectively (with t-ratios of 5.67 and 7.12). Given evidence of selection, these should be interpreted as upper bound measures of the effect of MSA size on the typical (modal) worker's wage. For mean wage, the corresponding estimates are higher, 5.78% and 5.47% (with t-ratios of 10.93 and 12.47) and significantly so relative to the modal wage elasticities (the t-ratios in column 4 on the difference in estimates are 5.94 and 3.94, respectively).

For workers with a high school degree or less, the pattern is once again different for both married women (Panel B) and men (Panel D). For both samples, the mode and mean wage elasticities are nearly identical, differing by less than 0.1 percentage point (see column 4 in the table). Recall from our model that absence of selection and evidence that the mode and mean wage are similar suggests that dilation is not present. The patterns in Panels B and D are consistent with that interpretation. This also suggests that we can be more precise in interpretating the elasticities for those with limited education: for married women, an approximate doubling of MSA size increases wage by roughly 5% while for men the corresponding estimate is 4%.

5. Conclusion

Identifying the effect of city size on an individual's wage is central to understanding the benefits of urbanization. Nevertheless, studies struggle to develop such estimates because unusually skilled workers may select into large city labor markets. We propose a new approach to this challenge. For single peaked wage distributions with a well-defined interior mode, if workers left of the mode disproportionately select out of large city labor markets, the CDF evaluated at the mode will shrink.

Drawing on the principle above, estimates indicate that selection contributes to urban wage premia for college trained workers but is absent for individuals with a high school degree or less. This pattern adheres to priors that selection will be most pronounced for individuals with more elastic labor supply and for those who are more moble across cities. It also echos evidence from Autor (2019) that increasing skill bias in the last few decades has amplified the role of knowledge-oriented college trained workers in large cities, while at the same time, tasks performed by noncollege workers in large cities have become more generic and similar to work in small cities. These trends may be contributing to differences in selection effects that we find for college workers versus those with limited education.

Additional estimates indicate that for college educated workers, an upper bound on the elasticity of modal wage with respect to MSA size is roughly 3.5%. Among workers with limited education, the absence of evidence of selection allows us to be more definitive. For married women, an approximate doubling of city size increases a typical individual's wage by roughly 5% while for men the corresponding estimate is roughly 4%. Although in principle these estimates could indicate that workers with limited education derive a greater productivity boost from city size as compared to college trained individuals, shifts in the nature of work described by Autor's (2019) suggest an alternate explanation. Given the similarity of large and small city work among individuals with limited education, noncollege workers may require compensation for the high cost of living in large cities if they are to remain active in such markets where they often serve as complements to more highly trained workers (e.g. Eeckhout et al, 2014).

Using the mode to test for selection is easy to implement and can serve as a diagnostic tool. While not a panacea – there must be a credible prior of a single peaked latent distribution with a well-defined interior mode – economic data often exhibit that property. Focusing on the mode also highlights outcomes for the most common occurrence or individual. This will be of intrinsic interest when behavior or policy is driven by the most populous group as compared to average or median outcomes.

References

Ahlfeldt, Gabriel, Stephen J. Redding, Daniel M. Sturm and N. Wolf (2015). "The economics of density: evidence from the Berlin Wall." *Econometrica* 83(6), 2127-2189.

Autor, David H. (2019). "Work of the Past, Work of the Future," *AEA Papers and Proceedings*, Vol 109, 1-32.

Bacolod, Marigee, Bernardo S. Blum and William C. Strange (2009). "Skills in the City," *Journal of Urban Economics*, 65(2), 136-153.

Balgova, Maria (2022). "Moving to Look or Looking to Move? The Role of Job Search in Migration Decisions," working paper.

Baum-Snow, Nathaniel, and Ronni Pavan (2012). "Understanding the city size wage gap." *The Review of Economic Studies* 79.1: 88-127.

Behrens, Kristian and Frederic Robert-Nicoud (2015). "Agglomeration theory with heterogeneous agents." Gilles Duranton, J. Vernon Henderson and William C. Strange (eds), Handbook in Regional and Urban Economics, Amsterdam (Holland), Elsevier Press.

Black, Daniel A., Kolesnikova, Natalia, Lowell J. Taylor (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, Francine D., Lisa M. Kahn (2007). "Changes in the labor supply behavior of married women: 1980–2000." *Journal of Labor Economics*, 25(3), 393-438.

Cardoso, Ana Rute and Pedro Portugal, (2005). "Contractual wages and the wage cushion under different bargaining settings." *Journal of Labor economics*, 23(4), 875-902.

Chen, Yen-Chi, Christopher R. Genovese, Ryan J. Tibshirani and Larry Wasserman (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, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, Sebastian Roux (2012). "The productivity advantages of large cities: Distinguishing agglomeration from firm selection." *Econometrica* 80.6: 2543-2594.

Combes, Pierre-Philippe and Laurent 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.

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

Eckert, Fabian, Mads Hejlesen, and Conor Walsh (2022). “The Return to Big City Experience: Evidence from Refugees in Denmark.” *Journal of Urban Economics*.

Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny (2014), “Spatial Sorting”, *Journal of Political Economy*, 122, 554–620.

Firpo, Sergio. (2007) “Efficient semiparametric estimation of quantile treatment effects.” *Econometrica* 75.1: 259-276.

Gaubert, Cecile (2018). “Firm sorting and agglomeration.” *American Economic Review*, 108(11), 3117-3153.

Glaeser, Edward L. and David C. Mare. (2001). “Cities and skills” *Journal of Labor Economics* 19.2: 316-342.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti (2010). “Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings.” *Journal of Political Economy* 118.3: 536-598.

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

Honoré, Bo. E. (1988). “Using Modes to Identify and Estimate Truncated Regression Models.” *Working paper.*

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

Jales, Hugo (2018). “Estimating the effects of the minimum wage in a developing country: A density discontinuity design approach.” *Journal of Applied Econometrics*, *33*(1), 29-51.

Jales, Hugo and Zhengfei Yu (2017). “Identification and estimation using a density discontinuity approach”, *in* Regression Discontinuity Designs: Theory and Applications (pp. 29-72). Emerald Publishing Limited.

Kemp, Gordon C. J.M.C. Santos Silva (2012). “Regression towards the mode.” *Journal of Econometrics,* 170(1), 92-101.

Kleven, Henrik J. (2016). “Bunching.” *Annual Review of Economics*, 8, 435-464.

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

Lee, Myoung-Jae (1993). “Quadratic mode regression.” *Journal of Econometrics*, 57(1-3), 1-19.

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

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak (2011). "Internal Migration in the United States," *Journal of Economic Perspectives*, Vol 25(3), 173-196.

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.

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

Rosenthal, Stuart S. and William C. Strange (2020). "How Close is Close? The Spatial Reach of Agglomeration Economies," *Journal of Economic Perspectives*, Vol 34(3), 27-49.

Saez, Emmanuel (2010). "Do taxpayers bunch at kink points?" *American Economic Journal: Economic Policy*, 2(3), 180-212.

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

Yao, Weixin and Longhai Li (2014). "A new regression model: modal linear regression." *Scandinavian Journal of Statistics*, 41(3), 656-67.

Table 1: Individual and MSA Modal Wage Distributions Based on Unobserved Factors[a]

Panel A: Individual Married Female Wage

	(1) 5th quantile	(2) 25th quantile	(3) 50th quantile	(4) 75th quantile	(5) 95th quantile	(6) Observations
College degree or more	16.65	26.94	34.93	44.92	71.28	155,456
High school degree or less	10.81	16.02	20.41	25.88	39.09	153,015

Panel B: Individual Male Wage

	(1) 5th quantile	(2) 25th quantile	(3) 50th quantile	(4) 75th quantile	(5) 95th quantile	(6) Observations
College degree or more	18.48	33.08	44.59	60.12	123.60	388,091
High school degree or less	12.60	20.21	26.55	34.37	52.83	396,955

Panel C: MSA Modal Married Female Wage

	(1) Min	(2) Max	(3) Median	(4) Mean	(5) Std. Dev.	(6) Observations
College degree or more	22.92	41.82	30.49	31.05	3.30	216
High school degree or less	13.53	24.77	17.61	18.08	2.17	216

Panel D: MSA Modal Male Wage

	(1) Min	(2) Max	(3) Median	(4) Mean	(5) Std. Dev.	(6) Observations
College degree or more	27.44	52.12	37.96	37.65	3.80	263
High school degree or less	18.04	33.20	22.73	23.23	2.94	263

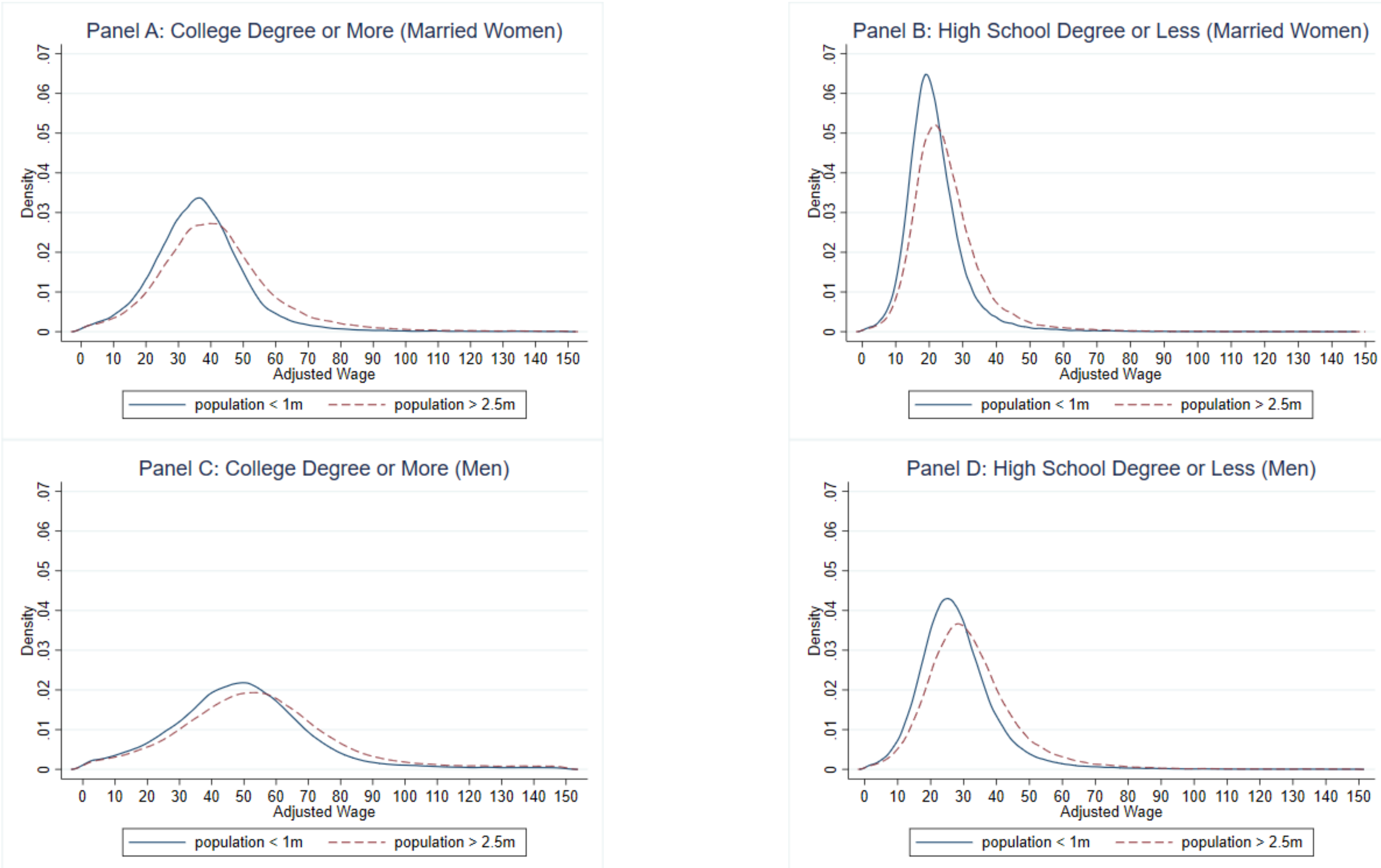
[a] Individual-level data are from the 2000 Census with samples restricted to non-Hispanic native-born white, full-time workers, age 25-54 who earn a wage equal to or above two-thirds of Federal minimum wage in 2000. Samples are further restricted to MSAs with at least 100,000 or more population in 2000 with at least 100 or more observations for the target group in the sample. Wage is in year-2022 (May) dollars and is calculated by dividing earned income by hours worked. Wage measures are exponentiated residuals from a log wage regression on fixed effects for industry, occupation, age, years of schooling (for high school or less) and marital status (for men) as described in the text. For each sample, raw log wage sample-specific means were added to the residuals. MSA modal wage is obtained from kernel density estimates using the Epanechnikov kernel with the Silverman rule of thumb bandwidth.

Table 2: Wage Regressions Based on Unobserved Factors[a]

Panel A: Married women with college degree or more	(1) CDF of wage evaluated at the Mode	(2) Log(wage) at the MSA Mode	(3) Log(wage) at the MSA Mean	(4) Coefficient difference (3) - (2)
Log population in MSA	-0.0122	0.0343	0.0578	0.0235
	(-3.80)	(5.67)	(10.93)	(5.94)
R-squared	0.042	0.113	0.348	0.103
Observations	216	216	216	216
Panel B: Married women with high school degree or less				
Log population in MSA	-0.0006	0.0523	0.0526	0.0003
	(-0.19)	(8.16)	(10.19)	(0.08)
R-squared	0.000	0.205	0.282	0.000
Observations	216	216	216	216
Panel C: Men with college degree or more				
Log population in MSA	-0.0074	0.0365	0.0547	0.0182
	(-2.93)	(7.12)	(12.04)	(3.94)
R-squared	0.028	0.142	0.365	0.050
Observations	263	263	263	263
Panel D: Men with high school degree or less				
Log population in MSA	-0.0007	0.0394	0.0400	0.0006
	(-0.28)	(6.09)	(7.94)	(0.15)
R-squared	0.000	0.109	0.180	0.024
Observations	263	263	263	263

[a] T-ratios based on robust standard errors in parentheses. Individual-level data are from the 2000 Census with samples restricted to non-Hispanic native-born white, full-time workers, age 25-54 who earn a wage equal to or above two-thirds of Federal minimum wage in 2000. Samples are further restricted to MSAs with at least 100,000 or more population in 2000 with at least 100 or more observations for the target group in the sample. Wage is in year-2022 (May) dollars and is calculated by dividing earned income by hours worked. Wage measures are residuals from a log wage regression on fixed effects for industry, occupation, age, years of schooling (for high school or less) and marital status (for men) as described in the text. For each sample, raw log wage sample-specific means were added to the residuals. MSA modal wage is obtained from kernel density estimates using the Epanechnikov kernel with the Silverman rule of thumb bandwidth.

Figure 1: Kernel Density Estimates of Wage Distributions Based on Unobserved Factors (Large and Small MSAs)[a]



[a] Individual-level data are from the 2000 Census with samples restricted to non-Hispanic native-born white, full-time workers, age 25-54 who earn a wage equal to or above two-thirds of Federal minimum wage in 2000. Samples are further restricted to MSAs with at least 100,000 or more population in 2000 with at least 100 or more observations for the target group in the sample. Wage is in year-2022 (May) dollars and is calculated by dividing earned income by hours worked. Wage measures are exponentiated residuals from a log wage regression on fixed effects for industry, occupation, age, years of schooling (for high school or less) and marital status (for men) as described in the text. For each sample, raw log wage sample-specific means were added to the residuals. MSA modal wage is obtained from kernel density estimates using the Epanechnikov kernel with the Silverman rule of thumb bandwidth. Samples used to estimate kernel densities are restricted to observations for which adjusted wage level is no larger than $150/hour.

Online Appendix A: Modal Shift With a Generalized Error Distribution

This appendix develops further properties of our model using the generalized error distribution (GED) to characterize the latent wage density. We make the following assumption:

Assumption A1: $f_0(y)$ belongs to the family of symmetric, generalized error distributions (GED):

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

where μ is the center of the distribution (equal to the mode, median and mean, and y_m above). The scale parameter σ governs the degree of dispersion in the distribution, κ is a shape parameter that affects the degree to which the mode is sharply defined, with range from 0 to ∞, and Γ denotes the gamma function.

For this density function, $\kappa = 1/2$ corresponds to the normal distribution. For $\kappa = 1$, the resulting distribution is a double exponential or Laplace distribution which has a sharply defined mode, 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 $U(\mu - \sigma, \mu + \sigma)$ and at the other extreme, as $\kappa \to \infty$, $f_0(y)$ becomes degenerate with all mass concentrated at a single value for y.

We consider two cases based on whether κ is strictly less than 1 or $\kappa \geq 1$. For the former, elasticity conditions govern modal shift. However, when $\kappa \geq 1$, the GED has a non-differentiable point at the mode and the mode is as least as sharply defined as for the Laplace. For this case, we show that the mode does not shift for any continuous and differentiable selection process like the one described in equation (2.4).

The extent to which the mode shifts in response to selection is governed by four parameters: a and b from the selection process in expression (2.4) of the text, and σ and κ that govern the shape of the GED function in (A.1). Modal shift is formalized by Proposition A1.

Proposition A1: *Suppose that selection is as defined by (2.4) in the text and the latent wage density is as in assumption A1 with $k < 1$. If the selection function reaches an asymptote p at $y_p > y_m$, then:*

(i) *Expressions (2.7) and (2.8) of the text are satisfied at a value for y that is the solution to:*

$$y = y_m + \sigma^{\frac{1}{1-k}}\left[\psi_{\pi,y}\right]^{\frac{k}{1-k}} \tag{A.2a}$$

where $\psi_{\pi,y} = \frac{\xi_{\pi,y}}{y}$.

(ii) *The mode of the conditional density in city s, denoted by $y_{m,s}$, is the value of y* that solves:

$$y_{m,s} = min\left\{y_p\ ,\ y_m + \sigma^{\frac{1}{1-k}}\left[\psi_{\pi,y}\right]^{\frac{k}{1-k}}\right\} \tag{A.2b}$$

(iii) *For normalized data with σ = 1, this simplifies to:*

$$y_{m,s} = min\left\{y_p\ ,\ y_m +\ \left[\psi_{\pi,y}\right]^{\frac{k}{1-k}}\right\} \tag{A.2c}$$

Proof: Taking the derivative of f_o in (A.1) with respect to the value for *y*,

$$f_0{}'(y) = -\ f_0(y)\left|\frac{y-y_m}{\sigma}\right|^{\frac{1}{\kappa}-1}\frac{1}{\sigma} \tag{A.3}$$

Recall from (2.7) in the text that the mode of the conditional density function must satisfy

$$\frac{\pi_s'(y)}{\pi_s(y)} = -\frac{f_0'(y)}{f_0(y)} \tag{A.4}$$

Substituting (A.3) into (A.4),

$$\frac{\pi_s'(y)}{\pi_s(y)} = \ \left|\frac{y-y_m}{\sigma}\right|^{\frac{1}{\kappa}-1}\frac{1}{\sigma} \tag{A.5}$$

Rearranging (A.5), the change in the mode in response to selection, $y_{m,s} - y_m$, is the value of *y* that solves

$$y_{m,s} - y_m = \sigma^{\frac{1}{1-k}}\left[\psi_{\pi,y}\right]^{\frac{k}{1-k}}. \tag{A.6a}$$

where $\psi_{\pi,y} = \frac{\pi_s'(y)}{\pi_s(y)}$ as in the text, or equivalently,

$$y_{m,s} - y_m = \sigma^{\frac{1}{1-k}}\left[\frac{\xi_{\pi,y}}{y_{m,1}}\right]^{\frac{k}{1-k}}. \tag{A.6b}$$

since $\xi_{\pi,y} =\ y\psi_{\pi,y}$.

If instead the GED shape parameter κ exceeds 1, the mode does not shift in response to selection. This condition and its proof are as follows.

Proposition A2: *For the GED defined in Assumption A1, if κ > 1 then the mode does not shift in response to a continuous, differentiable selection process, and the return to city size at the mode is unaffected by selection.*

Proof: Suppose a GED distribution with $\kappa > 1$ and let μ denote the mode for the latent distribution. With $\kappa > 1$, the GED sharpens to a non-differentiable point at the mode with infinite slope for y close to μ. Our goal is to show that the mode of the conditional density is μ, recalling that $f_s(y) = \frac{1}{c} f_0(y)\pi_s(y)$.

Because the density is not differentiable at μ, we cannot use first order conditions to locate the mode. Instead, note that density must decline upon moving away from the mode from above and below:

$$\lim_{y \downarrow \mu} f_s'(y) < 0 \tag{A.7}$$

$$\lim_{y \uparrow \mu} f_s'(y) > 0 \tag{A.8}$$

Focus first on (A.7). Substituting for f_s and applying the chain rule, (A.7) is equivalent to,

$$\lim_{y \downarrow \mu} \frac{f_0(y)\pi_s'(y)}{c} < \lim_{y \downarrow \mu} - \frac{f_0'(y)\pi_s(y)}{c} \tag{A.9}$$

Notice next that for $y > \mu$, the derivative of the GED in expression (2.8) is given by,

$$f_0'(y) = -\sigma^{-\frac{1}{k}}(y-\mu)^{\frac{1}{k}-1} f_0(y) \tag{A.10}$$

Substituting into (A.9) and rearranging yields,

$$\frac{\pi_s'(\mu)}{\pi_s(\mu)} < \left[\sigma^{-\frac{1}{k}}\right] \lim_{y \downarrow \mu} (y-\mu)^{\frac{1}{k}-1} = \infty \ . \tag{A.11}$$

Notice that the righthand side of (A.11) goes to infinity at μ because for all $k > 1$, the limit term rises to ∞ as y approaches μ. The condition described by (A.11) is therefore satisfied for all selection functions with finite slope at the latent density mode μ. This confirms that the conditional density has a negative slope as one moves to the right of the latent density mode, μ. Consider also the inequality in (A.8). By definition, the slope of f_0 is positive for $y < \mu$ and by assumption, $\pi_s'(y) > 0$ until the selection function reaches an asymptote p as in expression (2.4). These conditions ensure that the conditional density f_s has positive slope left of μ. Together, these arguments confirm that (A.7) and (A.8) are both satisfied at μ and, for that reason, the mode does not shift in response to selection.

Notice now that if y_p is sufficiently large so that the elasticity condition determines the conditional mode, then the mode shifts to the right by an amount governed by $\sigma^{\frac{1}{1-k}}\left[\psi_{\pi,y}\right]^{\frac{k}{1-k}}$, where σ and k determine the degree to which the mode is sharply defined. If instead, selection is random, as with the case associated with Proposition 2, then $\psi_{\pi,y} = 0$ and (A.2c) confirms that the mode does not shift. Also, for $\psi_{\pi,y} > 0$ and a given shape parameter k, as the scale parameter σ of the latent density increases, the distribution of y becomes more spread out and (A.2b) indicates that modal shift increases.