Monopsony and Employer Mis-optimization Explain Why Wages Bunch at Round Numbers

Arindrajit Dube*, Alan Manning†, and Suresh Naidu‡§

November 10, 2019

Abstract

We show that administrative wage data, including from online labor markets, exhibit considerable bunching at round numbers that cannot be explained by rounding of responses in survey data. We consider two hypotheses—worker left-digit bias and employer optimization frictions—and derive tests to distinguish between the two. Symmetry of the missing mass distribution around the round number suggests that optimization frictions are more important. We show that a more monopsonistic market requires less optimization frictions to rationalize the bunching in the data. The extent of monopsony power implied by our estimated separations elasticities, which are in line with other recent studies, are consistent with a sizable amount of non-optimal bunching, with only modest losses in profits.. We provide experimental validation of these results from an online labor market, where rewards are also highly bunched at round numbers. By randomizing wages for an identical task, our online experiment provides an independent estimate of the extent of employer market power, and fails to find evidence of any discontinuity in the labor supply function as predicted by workers' left-digit bias. Overall, the extent and form of round-number bunching suggests that "behavioral firms" can systematically misprice labor without being driven out of the market in the presence of monopsony power.

*University of Massachusetts Amherst, NBER, and IZA; adube@econs.umass.edu.

†London School of Economics; A.Manning@lse.ac.uk.

‡Columbia University, and NBER; sn2430@columbia.edu.

§We thank Sandra Baquie, Doruk Cengiz, Len Goff, Jeff Jacobs, and Jon Zytnick for excellent research assistance. We are grateful for helpful comments from Matt Backus, Ellora Derenoncourt, David Green, Ilyana Kuziemko, Ioana Marinescu, Jim Rebitzer and from seminar participants at Microsoft Research, Columbia University, Boston University, Harvard University, University of Toronto, MIT Sloan, and the 2018 Society of Labor Economists conference.

1 Introduction

Behavioral economics has documented a wide variety of deviations from perfect rationality among consumers and workers. For example, in the product market, prices are more frequently observed to end in 99 cents than can be explained by chance, and a sizable literature has confirmed that this reflects firms taking advantage of customers' left digit bias (e.g Levy et al. 2011). However, it is typically assumed that deviations from firm optimization are unlikely to survive, as competition among firms drives firms that fail to maximize profits out of business. Therefore, explanations for pricing and wage anomalies typically rely on human behavioral biases. In this paper, we show that when it comes to a quantitatively important anomaly in wage setting–bunching at round numbers—it is driven by behavioral biases of firms and not workers. We also show that round number bunching survives because of monopsony power in the labor market, which blunts the cost of mispricing. A moderate amount of monopsony power, well within the range obtained in both our data and other recent high-quality US evidence, is sufficient to explain the substantial bunching we document in this paper.

We begin by documenting bunching in the hourly wage distribution at "round" numbers. We use data from both administrative sources and an online labor market to confirm that there is true bunching of wages at round numbers, and it is not simply an artifact of survey reporting. We begin by providing the first (to our knowledge) credible evidence on the extent to which wages are bunched at round numbers in high quality, representative data on hourly wages from Unemployment Insurance records from the three largest U.S. states (Minnesota,Washington, and Oregon) that collect information on hours.[1] We further assess the extent of bunching in online labor markets, using a near universe of posted rewards on the online platform Amazon Mechanical Turk (MTurk).

We compare the size of the bunches in the administrative data to those in the Current

[1]We were able to use individual-level matched worker-firm data from Oregon, and micro-aggregated data of job counts by detailed wage bins in Minnesota and Washington.

Population Survey (CPS), where round number bunching is prevalent. For example, in the CPS data for 2016, a wage of $10.00 is about 50 times more likely to be observed than either $9.90 or $10.10. Figure 1 shows that the hourly wage distribution from the CPS outgoing rotation group (ORG) data between 2010 and 2016 has a visually striking modal spike at $10.00 (top panel). The middle panel of the figure shows that the share of wages ending in round numbers is remarkably stable over the past 35 years, between 30-40% of observations. The bottom panel of the figure also shows that since 2002, the modal wage has been exactly $10.00 in at least 30 states, reaching a peak of 48 in 2008. Of course, survey data is likely to overstate the degree of round-number bunching. We also use a unique CPS supplement, which matches respondents' wage information with those from the employers to correct for reporting error in the CPS, and show that the measurement error-corrected CPS replicates the degree of bunching found in administrative data in the states where we observe both. It seems highly unlikely that such bunching at $10.00 is present in the distribution of underlying marginal products of workers.

We first attempt to account for bunching by building and experimentally testing an imperfectly competitive model with workers' left-digit bias, paralleling product price studies from the marketing literature that have documented pervasive "left-digit bias" where agents ignore lower-order digits in price. First, we design and implement an experiment (N=5,017) on an online platform (MTurk). We randomly vary rewards above and below 10 cents for the same task to estimate the labor supply function facing an online employer. Like the administrative data, the task reward distribution on MTurk exhibits considerable bunching. However, our experimentally estimated labor supply function shows no evidence of a discontinuity as would be predicted by worker left-digit bias. We further use observational job posting data from MTurk and show there is no discontinuity in the duration of vacant tasks posted at $1.00.

Next, we use administrative matched worker-firm data on hourly wages from Oregon to show that there is no discontinuity in the separation elasticity at $10.00 conditional on past worker wage history and other characteristics. These results suggest that our experimental findings on lack of left-digit bias are likely to have external validity beyond on-line labor markets. In addition, our analysis using the matched worker-firm produces observational residual labor supply elasticities in the 3 to 5 range, which is very similar to those estimated in similar contexts in other recent work (Sokolova and Sorensen (2018); Caldwell and Oehlsen (2018)), and suggests a moderate degree of monopsony power in the U.S. low-wage labor market.Finally, we provide evidence on lack of worker left-digit bias using stated-preference experiments where workers at a large retailer (Wal-Mart) were asked whether they would choose to quite their current job for another, where the hypothetical wage was randomized. While a higher offered wage makes it more likely that the workers leave their current position, we found no discontinuity at round numbers.

We then extend the model to allow imperfect firm optimization in the form of employer preferences for round wages, parameterized by the fraction of profits foregone in order to pay a round number. We recover estimates of monopsony, employer misoptimization, and left-digit bias from the distribution of missing mass around $10.00 relative to a smooth latent density and conclude that employer mis-optimization accounts for much of the observed bunching.[2]

In the administrative data from Oregon, we further show that small, low-wage employers in sectors such as construction are most likely to pay round numbers, consistent with bunching being a result of small, potentially unsophisticated, employer pay-setting. Using our estimates of monopsony power, we calculate that employer misoptimization in US labor markets costs mis-optimizing employers between 2% to 8% of profits, but is an order of magnitude smaller in online labor markets. The reason is that the monopsony power on Amazon Mechanical Turk is so high that even very substantial mispricing (as indicated by

[2] While other configurations are logically possible, they do not easily explain why wages are bunched at round numbers. For example, if employers had a left-digit bias, any heaping would likely occur at $9.99 and not at $10.00, which is not true in reality. Similarly, if workers tended to round off wages to the nearest dollar, this would not encourage employers to set pay exactly at $10.00. In contrast, both workers' left-digit bias and employers' tendency to round off wages provide possible explanations for a bunching at $10.00/hour.

a high degree of bunching) does not lead employers to lose very much in profits.

Our paper is related to a small but growing literature on behavioral firms, which documents a number of ways firms fail to maximize profits (DellaVigna and Gentzkow 2017, Goldfarb and Xiao 2011, Hortacsu and Puller 2008, Bloom and Van Reenen 2007, Cho and Rust 2010). A large literature has discussed cognitive biases in processing price information, but little of this has discussed applications to wage determination. Behavioral labor economics has extensively documented other deviations from the standard model (e.g. time-inconsistency and fairness, see Babcock et al. (2012) for an overview) for worker behavior. Worker behavioral phenomena have been replicated even in online spot labor markets (Chen and Horton (2016) , Della Vigna and Pope (2016)). Our paper is the first to document the absence of left-digit bias among workers, as well as the first to systematically quantify the extent of firm mis-optimization in the labor market.

In our account of wage-bunching, it is important to assume that firms have some labor market power. In this, we follow work in behavioral industrial organization that explores how firms choose prices when facing behavioral consumers in imperfectly competitive markets.[3] A recent and fast-growing literature has argued that, far from requiring explicit collusion (as in professional sports) or restrictive non-compete contracts (Starr, Bishara and Prescott 2016, Krueger and Ashenfelter 2017) or being confined to particular institutional environments (e.g. Naidu 2010, Naidu, Nyarko and Wang 2016), a degree of monopsony is in fact pervasive in modern labor markets (Manning 2011).[4] We show that moderate amounts of monopsony in the labor market can provide a parsimonious explanation of anomalies in the wage distribution, such as patterns of wage-bunching at arbitrary numbers.[5]

[3]See Heidhues and Kőszegi (2018) for a survey and Gabaix and Laibson 2006 for an early example. Theoretical models to explain bunching in prices also assume firms have some market power: e.g., Basu (1997) has a single monopolist supplying each good, Basu (2006) has oligopolistic competition, and Heidhues and Kőszegi (2008) use a Salop differentiated products model.

[4]See Naidu et al. (2018) for a more recent survey.

[5]Hall and Krueger (2012) show that wage posting is much more frequent in low wage labor markets than bargaining. Their data shows that more than 75% of jobs paying an hourly wage of around \$10 were ones where employers made take-it-or-leave-it offers without any scope for bargaining. We also find that the

The plan of the paper is as follows. In section 2, we provide evidence on bunching at round numbers using administrative data as well as data from the CPS corrected for measurement error, and benchmark these against the raw CPS results. We recover the source of the bunched observations by comparing the observed distribution to an estimated smooth latent wage distribution. In section 3, we develop a model of bunching that nests worker left-digit bias and firm optimization frictions as special cases. Section 4 presents findings from the online experiment, administrative data, and stated-preference experiment–recovering both estimates of the degree of monopsony as well as the extent of worker-left digit bias. Section 5 extends the model to account for firm misoptimization. Section 6 recovers the degree of mis-optimization and monopsony from the bunching estimates under a variety of assumptions about the degree of heterogeneity in both, and documents the characteristics of firms that mis-optimize. Section 7 concludes.

2 Bunching of wages at round numbers

There is little existing evidence on bunching of wages. One possible reason is that hourly wage data in the Current Population Survey comes from self-reported wage data, where it is impossible to distinguish the rounding of wages by respondents from true bunching of wages at round numbers. Documenting the existence of wage-bunching requires the use of other higher-quality data.

2.1 Administrative hourly wage data from selected states

Earnings data from administrative sources such as the Social Security Administration or Unemployment Insurance (UI) payroll tax records is high quality, but most do not contain information about hours. However, 4 states (Minnesota, Washington, Oregon, and Rhode

---

bunching at the $10/hour wage in the Hall and Krueger data is almost entirely driven by jobs with such take-it-or-leave-it offers. Along with our evidence from MTurk, where there is no scope for bargaining, this makes it unlikely that employers offer round number wages as a signal for bargaining.

Island) have UI systems that collect detailed information on hours, allowing the measurement of hourly wages. We have obtained individual-level matched employer-employee data from Oregon, as well as micro-aggregated hourly wage data from Minnesota and Washington. The UI payroll records cover over 95% of all wage and salary civilian employment. Hourly wages are constructed by dividing quarterly earnings by total hours worked in the quarter. The micro-aggregated data are state-wide counts of employment (and hours) by nominal $0.05 bins between $0.05 and $35.00, along with a count of employment (and hours) above $35.00. The counts exclude NAICS 6241 and 814, home-health and household sectors, which were identified by the state data administrators as having substantial reporting errors.

Figure 2 shows the distribution of hourly wages in MN, OR and WA (we report the distributions separately in the Appendix). The histogram reports normalized counts in $0.10 (nominal) wage bins, averaged over 2003q1 to 2007q4. We focus on this period because in later years the nominal minimum wages often reach close to $10.00, making it difficult to reliably estimating a latent wage density in this range as we do in our analysis later in this paper. The counts in each bin are normalized by dividing by total employment. The wages are clearly bunched at round numbers, with the modal wage at the $10.00 bin representing more than 1.5 percent of overall employment. This suggests that observed wage bunching is not solely an artifact of measurement error, and is a feature of the "true" wage distribution. Further, the histogram reveals spikes at the MN, OR, and WA minimum wages in this period, suggesting that the hourly wage measure is accurate.

In Online Appendix C, we show very similar degree of bunching in a measurement-error corrected CPS, using the 1977 CPS Supplement that recorded wages from firms as well as workers. While the degree of bunching in the raw CPS falls with the measurement error correction, it remains significant, and indeed comparable to the administrative data.

2.2 Task rewards in an online market: Amazon Mechanical Turk

Amazon MTurk is an online task market, where "requesters" (employers) post small online Human Intelligence Tasks (HITs) to be completed by "Turkers" (workers).[6] Psychologists, political scientists, and economists have used MTurk to implement surveys and survey experiments (e.g. Kuziemko et al. (2015)). Labor economists have used MTurk and other online labor markets to test theories of labor markets, and have managed to reproduce many behavioral properties in lab experiments on MTurk (Shaw et al. 2011).

We obtained the universe of MTurk requesters from Panos Ipeirotis at NYU. We then used the Application Programming Interface developed by Ipeirotis to download the near universe of HITs from MTurk from May 2014 to February 2016, resulting in a sample of over 350,000 HIT batches. We have data on reward, time allotted, description, requester ID, first time seen and last time seen (which we use to estimate duration of the HIT request before it is taken by a worker). The data are described more fully in Online Appendix E and in Dube et al. (2018).

Figure 3 shows that there is considerable bunching at round numbers in the MTurk reward distribution. The modal wage is 30 cents, with the next modes at 5 cents, 50 cents, 10 cents, 40 cents, and at $1.00. This is remarkable, as this is a spot labor market that has almost no regulations, suggesting the analogous bunching in real world labor markets is not driven by unobserved institutional constraints, including long-term implicit or explicit contracts. Nor is there an opportunity for bargaining, so the rewards posted are generally the rewards paid, although there may be unobserved bonus payments.

3 A model of round-number bunching in the labor market

This section presents a model of bunching in the labor market which builds on features in the price-bunching literature (e.g. Basu 1997, Basu 2006) and the optimization friction

[6]The sub header of MTurk is "Artificial Artificial Intelligence", and it owes its name to a 19th century "automated" chess playing machine that actually contained a "Turk" person in it.

literature (e.g. Chetty 2012).

Suppose there are many workers differing in their marginal product p, assumed to have density $k(p)$ and CDF $K(p)$—assume labor is supplied inelastically to the market as a whole. We assume there is only one "round number" wage in the vicinity of the part of the productivity distribution we consider—denote this by w_0. We do not here attempt to micro-found w_0. There are various functions of w_j that could deliver w_0. For example we could set $w_0 = w_j - \mod(w_j, 10^h)$, where $\mod(w, 10^h)$ denotes the remainder when w is divided by 10^h and h is the highest digit of w. Or we could impose the formulation in Basu (1997), where agents form expectations about the non-leftmost digits. In contrast to Basu (1997), which delivers a strict step function, the discrete choice formulation allows supply to be increasing even at non-round numbers, as well as relaxing the assumption that each good is provided by a single monopolist (Basu (2006) considers a Bertrand variant of a similar model, showing that .99 cents can be supported as a Bertrand equilibrium with a number of homogeneous firms). We also extend the formulation of digit bias from Lacetera, Pope and Sydnor (2012) by allowing utility to depend on the true wage w as well as the leading digit.[7] We consider two reasons why w_0 might be chosen—left-digit bias on the part of workers, and mis-optimization on the part of employers in the form of paying round numbered wages.

We model the left-digit bias of workers in the following way. Assume that, for workers with marginal product, p, the supply of workers to a firm that pays wage w is given by:

$$l(w,p) = \frac{\left[we^{\gamma \mathbb{1}_{w \geq w_0}}\right]^{\eta}}{C} k\left(p\right) \tag{1}$$

where $C \equiv \sum_{j=1}^{M} \left[w_j e^{\gamma \mathbb{1}_{w_j \geq w_0}}\right]^{\eta}$. We assume that there are a sufficiently large number of firms that C is treated as exogenous by each individual firm. If $\gamma > 0$ then there is a discontinuity at w_0: γ is the percentage increase in labor supply that comes from the

[7] However we do not parameterize the extent of "left-digitness" as Lacetera, Pope and Sydnor (2012) do. We are implicitly assuming "full inattention" to non-leading digits.

left-digit bias of workers so the size of γ is a natural measure of the extent of left-digit bias. Left digit bias has been documented in a wide variety of markets, used to explain prevalence of product prices that end in 9 or 99, and is a natural candidate explanation for bunching in the wage distribution.[8] Our model of labor supply to individual firms can be micro-founded using a multinomial logit model—see Card et al. (2018) for an application to the labor market.[9] Our baseline model assumes some imperfect competition in the labor market but perfect competition is a special case as $\eta \to \infty$.

Denote by $l^*(w,p) = \frac{w^\eta}{C} k(p)$ the "nominal" labor supply curve facing the firm, without any worker left-digit bias. Define $\rho(w,p) = (p-w)l^*(w,p)$. Here $\rho(w,p)$ is, in the language of Chetty (2012), the "nominal model" that parameterizes profits in the absence of left-digit bias. Optimizing wages in the nominal model would yield a smooth "primitive" profit function of productivity given by $\pi(p_j) = (\frac{p_j}{1+\eta})^{1+\eta}$, but the presence of worker biases induces discontinuities in true profits at round numbers. In deciding on the optimal wage for employers one simply needs to compare the profits to be made by maximizing the nominal model and paying the round number. Consider the wage that maximizes the nominal model. Given the isoelastic form of the labor supply curve to the individual firm, this can simply be shown to be:

$$w^*(p) = \frac{\eta p}{1+\eta} \tag{2}$$

[8]For example, Levy et al. (2011) show that 65% of prices in their sample of supermarket prices end in 9 (33.4% of internet prices), and prices ending in 9 are 24% less likely to change than prices ending in other numbers. Snir et al. (2012) also document asymmetries in price increases vs. price decreases in supermarket scanner data, consistent with consumer left-digit bias. A number of field and lab experiments document that randomizing prices ending in 9 results in higher product demand (Anderson and Simester 2003, Thomas and Morwitz 2005, Manning and Sprott 2009). Pope, Pope and Sydnor (2015) show that final negotiated housing prices exhibit significant bunching at numbers divisible by $50,000, suggesting that round number focal points can matter even in high stakes environments. Lacetera, Pope and Sydnor (2012) show that car prices discontinuously fall when odometers go through round numbers such as 10,000. Allen et al. (2016) document bunching at round numbers in marathon times, and interpret this as reference-dependent utility. Backus, Blake and Tadelis (2015) show that posted prices ending in round numbers on eBay are also a signal of willingness to bargain down.

[9]Matejka and McKay (2015) provide foundations for discrete choice that incorporates inattention, and see Gabaix (2017) for applications of inattention to a wide variety of behavioral phenomena, including left-digit bias.

which reflects a mark-down on the marginal product with the size of the mark-down determined by the extent of imperfect competition in the labor market. If the labor market is perfectly competitive, $\eta = \infty$, wages are equal to marginal product. We will refer to the wage that maximizes the nominal model as the latent wage. The firm with job productivity p will pay the round number wage as opposed to the latent wage if:

$$e^{\eta\gamma\mathbb{1}_{w*(p)<w_0}} > \frac{\rho\left(w^*\left(p\right), p\right)}{\rho\left(w_0, p\right)} \tag{3}$$

Taking logs, we obtain that a firm will pay the round number if $p \in [p_*, \frac{(1+\eta)}{\eta}w_0]$, where p_*solves

$$\eta\gamma = \ln\rho\left(w\left(p_*\right), p_*\right) - \ln\rho\left(w_0, p_*\right) \tag{4}$$

This implies that all firms with latent wages between w_0 and $w(p_*)$ will pay w_0. Thus, bunching will be larger the greater both η and γ are. We now turn towards identifying estimates of η and γ from experimental and observational data.

4 Experimental and Observational Evidence on Worker Left-Digit Bias

4.1 Experimental Evidence From MTurk

We begin with providing direct evidence on the parameters of interest based on an experiment on an online labor market. The advantage of this approach is the high degree of internal validity, though a disadvantage is that one is inevitably unsure about the external validity of the estimates. For example, one might expect that these "gig economy" labor markets are very competitive because they are lightly regulated and there are large numbers of workers and employers with little long-term contracting. However, we show that a standard measure of monopsony, the labor supply elasticity facing the firm, is quite low, implying considerable inefficiencies in these types of "crowdsourcing" labor markets, which are finding increased use by large employers (for example Google, AOL, Netflix, and Unilever all subcontract with crowdsourcing platforms akin to MTurk) around the world (Kingsley, Gray and Suri 2015).

The use of Amazon Turk by researchers in computer science (particularly the subfield of human computation), psychology, political science and economics has increased in recent years. However, little of this research has considered the market structure of Amazon Turk (although see Kingsley, Gray and Suri (2015) for complementary evidence of requester market power on MTurk) or indeed any online labor market. In their original paper on labor economics on Amazon Turk, Horton, Rand and Zeckhauser (2011) implement a variant of the experiment we conduct below, making take it or leave it offers to workers with random wages in order to trace out the labor supply curve. However, while they label this an estimate of labor supply to the *market*, it is in fact a labor supply to the requester that they are tracing out, as the MTurk worker has the full list of alternative MTurk jobs to choose from. [10] We designed an experiment to test for worker left-digit bias.[11] We randomize wages for a census image classification task to estimate discontinuous labor supply elasticities at round numbers (in particular at 10 cents, to test for left-digit bias). We choose 10 cents because it is the lowest round number, allowing us to maximize the power of the experiment to detect left-digit bias. We also aim to replicate the upward sloping labor supply functions to a given task estimated in Horton, Rand and Zeckhauser (2011). We posted a total of 5,500 unique HITS on MTurk tasks for $0.10 that includes a brief survey and a screening task, where respondents view a digital image of a historical slave census schedule from 1850 or 1860, and answer whether they see markings in the "fugitives" column (for details on the 1850 slave census, see Dittmar and Naidu (2016)).

[10] In a companion paper, Dube, Jacobs, Naidu and Suri (2018) compile labor supply elasticities implicit in the results from a number of previous crowdsourcing compensation experiments on MTurk. We find they are uniformly small, with a precision-weighted elasticity of 0.14, and a similarly low non-experimental labor supply elasticity (< 0.2) estimated on the scraped MTurk data.

[11] Pre-registered as AEA RCT ID AEARCTR-0001349.

This is close to the maximum number of unique respondents obtainable on MTurk within a month-long experiment. Respondents are offered a choice of completing an additional set of classification tasks for a specific wage. Appendix Figure E.1 shows the screens as seen by participants with (1) the consent form, (2) the initial screening questions and demographic information sheet, and (3) the coding task content.

We refer to the initial screening part as stage-1. Those who complete stage-1 and indicate that the primary reason for participation is "money" or "skills" (as opposed to "fun") are then offered an additional task of completing either 6 or 12 such image classifications (chosen randomly) for a specific (randomized) wage, w, which we refer to as the stage-2 offer. If they accept the stage-2 offer, they are provided either 6 images (task type A) or 12 images (task type B) to classify, and are paid the wage w. These 5,500 HITs will remain posted until completed, or for 3 months, which ever is shorter. Any single individual on MTurk (identified by their MTurk ID) will be allowed to only do one of the HITs. We aim to assess the left-digit bias in wage perceptions experimentally by randomizing the offered wages for HITs on MTurk to vary between \$0.05 and \$0.15, and assessing whether there is a jump in the acceptance probability between \$0.09 and \$0.10 as would be predicted by a left-digit bias.[12]

[12]There are a few anomalies in the data relative to our design. The first was that a small number (17) of individuals were able to get around our javascript mechanism for preventing the same person from doing multiple HITs. In the worst cases, one worker was able to do 118 HITs, while 3 others were able to do more than 10. The second is that 9 individuals were entering responses to images they had not been assigned. We drop these HITs from the sample, which costs us 316 observations. None of the substantive results change, although the nominal labor supply effect is slightly more precise when those observations are included. We also drop 3 observations where participants were below the age of 16 or did not give the number of hours they spent on MTurk. Finally, we underestimated the time it would take for all of our HITs to be completed, and thus some (roughly 11%) of our observations occur after the pre-registration plan specified data collection would be complete. We construct an indicator variable for these observations and include it in all specifications discussed in the text (the additional specifications in Online Appendix E omit this variable).

4.2 Specifications

While our model entails a sharp discontinuity in the level of labor supplied at a round number (a "notch") we do not impose this in all our specifications, and allow for either a kink or a notch, and also control for the overall shape of the labor supply curve in a variety of ways. We estimate the following 3 specifications, all of which were included in the pre-analysis plan. We deviate slightly from our pre-analysis plan by including controls and using logit rather than linear probability to better match our model, and we will estimate analogous specifications in the observational Oregon data, as well as the worker stated preference experiment below. We show the exact specifications from the pre-analysis plan in Online Appendix E.

First, we estimate a logit regression of an indicator for accepting a task on log wages, essentially following the specification entailed by our model:

$$Pr(Accept_i) = \beta_0 + \eta_1 log(w_i) + \beta_1 T_i + \beta_2 X_i + \epsilon_i \tag{5}$$

Here T is a dummy indicating the size of the task. We add individual covariates X_i for precision; point estimates remain unchanged when controls are excluded (shown in Online Appendix E). Our main test from this specification is that the slope (semi-elasticity) $\eta_1 > 0$: labor supply curves (to the requester) are upward sloping. We will also report the elasticity $\eta = \frac{\eta_1}{E[Accept]}$ in every specification where we estimate it.

Our first test for left-digit bias is based on a logit regression allowing for a jump in the labor supply at \$0.10, but constraining the slope to the the same on both sides:

$$Pr(Accept_i) = \beta_{0A} + \eta_{1A} log(w_i) + \gamma_{1A} \mathbb{1}\left\{w_i \geq 0.1\right\}_i + \beta_{1A} T_i + \beta_2 X_i + \epsilon_i \tag{6}$$

Here left-digit bias is rejected if $\gamma_{A1} = 0$. This specification corresponds closely to the theoretical model with constant labor supply semi-elasticity η_{1A}, and with $\gamma = e^{\gamma_{1A}}$ measuring the extent of left-digit bias.

Our second specification allows for heterogeneous slopes in labor supply above and below \$0.10 using a knotted spline, where the knots are at \$0.09 and \$0.10:

$$\begin{aligned} Pr(Accept_i) = \beta_{0B} + \eta_{1B} log(w_i) + \gamma_{2B} \times (log(w_i) - log(0.09)) \times \mathbb{1}\{w_i \geq 0.09\}_i \\ + \gamma_{3B} \times (log(w_i) - log(0.10)) \times \mathbb{1}\{w_i \geq 0.1\}_i + \beta_{2B} T_i + \beta_2 X_i + \epsilon_i \end{aligned} \tag{7}$$

Our main test here is that the slope between \$0.09 and \$0.10 (i.e., $\eta_{1B} + \gamma_{2B}$) is greater than the average of the slopes below \$0.09 and above \$0.10 $\left(\frac{1}{2} \times \eta_{1B} + \frac{1}{2} \times (\eta_{1B} + \gamma_{2B} + \gamma_{3B})\right)$; or equivalently to test: $\gamma_{2B} > \gamma_{3B}$.

Finally, our most flexible specification estimates:

$$Pr(Accept_i) = \sum_{k \in S} \delta_k \mathbb{1}\{w_i = k\}_i + \gamma \beta_{3B} T + \beta_2 X_i + \epsilon_i \tag{8}$$

And then calculates the following statistics:

$$\delta_{jump} = (\delta_{0.1} - \delta_{0.09})$$

$$\beta_{local} = (\delta_{0.1} - \delta_{0.09}) - \frac{\left(\sum_{k=.08, k \neq 0.1}^{0.12} \delta_k - \delta_{k-0.01}\right)}{4}$$

$$\beta_{global} = (\delta_{0.1} - \delta_{0.09}) - \frac{1}{10}(\delta_{0.15} - \delta_{0.05})$$

The β_{local} estimate provides us with a comparison of the jump between \$0.09 and \$0.10 to other localized changes in acceptance probability from \$0.01 increases. In contrast, β_{global} provides us with a comparison of the jump with the full global (linear) average labor supply response from varying the wage between \$0.05 and \$0.15. The object $\frac{1}{10}(\delta_{0.15} - \delta_{0.05})$ will also be used to estimate the overall labor supply response and elasticity facing the person posting a task on MTurk.

A left-digit bias might not only affect willingness to accept a task, but also may affect a worker's performance. For example, if workers are driven by reputational concerns or exhibit reciprocity, and they perceive \$0.10 to be discontinuously more attractive than \$0.09, we may expect a jump in performance at that threshold. To assess this, we will also estimate the same statistics, but with the error rate for the two known images (i.e., equal to 0, 0.5, or 1) as the outcome instead of $Accept_i$.

4.3 Experimental results

Our distribution of wages was chosen to generate power for detecting a discontinuity at 10 cents, as can be seen in the wage distribution in Figure 4. The binned scatterplot in Figure 4 shows the basic pattern of a shallow slope (in levels) with no discontinuity at 10. Table 1 below shows the key experimental results from the specifications above, which uses log wages as the main independent variable. Column 1 reports the estimates using a log wage term only; the elasticity, η, is 0.083. The elasticity is statistically distinguishable from zero at the 1 percent level, consistent with an upward sloping labor supply function facing requesters on MTurk. However, the magnitude is quite small, suggesting a sizable amount of monopsony power in online labor markets. When we restrict attention only to "sophisticated" MTurkers (column 5), the elasticity is only somewhat larger at 0.132, still surprisingly small.

While we find a considerable degree of wage-setting power in online labor markets, we do not find any evidence of left-digit bias for workers. Column 2 estimates equation 6 and tests for a jump at \$0.10 assuming common slopes above and below \$0.10. Column 3 corresponds to equation 7 and allows for slopes to vary on both sides of \$0.10. Finally, column 4, following the flexible specification in equation 8, estimates coefficients for each 1-cent dummy in the regression and compares the change between \$0.09 and \$0.10 to either local or global changes. In all of these cases, the estimates are close to zero in magnitude, and not statistically significant. We can rule out even small differences in the probability of acceptance between $0.09 and $0.10. When we limit our sample to sophisticated MTurkers, we do not find any left-digit bias either. None of the estimates for discontinuity in the labor supply function are statistically significant or sizable in columns 6, 7 or 8.

Column 2's specification corresponds closely to the theoretical model, where we can recover γ by exponentiating the coefficient on the dummy for a reward greater than or equal to $0.10. The point estimate for γ is 0.99, while the 95 percent confidence interval of (0.972, 1.029) is concentrated around one.

We also estimate parallel logit regressions using task quality as the outcome, which is defined as the probability of getting at least 1 out of two pre-tagged images correct. In Appendix Table E.3, we find that no evidence that task performance rises discontinuously at the $0.10 threshold. We also find little impact of the reward on task performance for the range of rewards offered; the most localized comparison yields estimates very close to zero.

We interpret the evidence as strongly pointing away from any left-digit bias on the workers' side. Moreover, it also suggests that locally, there is not very much impact of rewards on task performance: therefore, the primary benefit of providing a slightly higher reward is occurring through increased labor supply and not through performance. Summarizing to this point, while there is considerable bunching at round numbers in the MTurk reward distribution, including at $0.10, there is no indication of worker-side left-digit bias in labor supply or in performance quality. This finding is counter to the analogous explanation for the product market, where a number of experiments have found that demand for products increases when prices ending in 9 are posted (e.g. Anderson and Simester 2003). At the same time, we find considerable amount of wage-setting power in this online labor market: labor is fairly inelastically supplied to online employers, with an estimated elasticity η generally between 0.1 and 0.2.

In Online Appendix D, we present complementary evidence from scraped MTurk data for the $0.51 to $1.49 range, to show that similar patterns are obtained at the even more salient round number of $1.00. By estimating how long a job stays posted before being filled as a function of the reward posted (and controlling for hour of first posting, requester and task keyword fixed effects), we can recover another estimate of the labor supply curve facing an employer. In the fixed-effects estimator, the implied labor-supply elasticity under a constant offer arrival rate assumption (so that variation in durations are reflecting heterogeneity in tastes for the job rather than heterogeneity in search outcomes) is quite small, between .5 and 1, although larger than our experimental estimates. We also show that tasks with rewards greater than $1.00 do not discontinuously fall in the time to fulfillment, consistent with our experimental findings at $0.10. Together, the observational and experimental evidence suggest that, at least on Amazon Turk, there is plenty of monopsony, and little left-digit bias, at both the $0.10 and $1.00 thresholds.

4.4 Observational Evidence From Oregon

While the experimental evidence strongly points away from worker left-digit bias in the online labor markets, here we additionally use observation evidence from "real world" labor market to assess whether workers' labor supply exhibits sharp breaks (notch or kink).

Using matched employer-employee data from Oregon, we estimate the responsiveness of separations to wages, and for a discontinuity at $10.00/hour wage. The separation elasticity is widely used in the monopsony literature to measure the responsiveness of the supply of labor to a firm to the firm's posted wage (see Manning (2011)). A key concern in estimating this elasticity is that workers earning different wages may differ in other dimensions; as we discuss below, we use a variety of worker characteristics including past wages to account for such heterogeneity. Using the 2003-2007 period, we consider all workers who were hired within this period, had at least 2 quarters of tenure, and had a least an average of 5 hours/week of work experience in previous quarter (excludes the bottom 5% of observations). These restrictions workers who are very marginally attached to the workforce. We then compare how separation rate at time t varies by wage w_{it} among workers who were hired exactly in the same quarter, at the same starting hourly wage (within 10 cents), were in the same category of weekly hours in the prior quarter.[13]. Namely, we include as baseline controls the vector $\mathbf{X_{it}}$ which includes fully saturated interactions between (1) calendar time in quarter, (2) quarter of hire, (3) previous quarter's hours category. In a second version, we further interact all of these with 2-digit industry indicators and a 6-categories of firm size indicators.[14] In yet a third version, we instead interact the baseline controls with the firm fixed effect. This third (most saturated) version compares very similar workers hired within the same firm at the same time and at the same original wage working similar hours, but who happen to be earning somewhat different wages at the present time.

We estimate analogues of our experimental specifications above, regressing separations S_{it} on log wages, and an indicator for earning more than \$10.00.

$$S_{it} = \beta_0 + \eta_1 log(w_{it}) + \gamma_{1A} \mathbb{1}\left\{w_i \geq 10.00\right\}_i + \Lambda \mathbf{X}_{it} + \epsilon_{it} \tag{9}$$

This specification tests for a jump in the labor supply at \$10, but constraining the slope to the the same on both sides. So left-digit bias is rejected if $\gamma_{1A} = 0$. Similar to the case of MTurk experimental data, a second specification allows for heterogeneous slopes in labor supply above and below \$10 using a knotted spline, where the knots are at \$9.90 and \$10.00:

$$\begin{aligned} S_{it} = \beta_{0B} + \eta_{1B} log(w_{it}) + \gamma_{2B} \times (log(w_{it}) - log(9.90)) \times \mathbb{1}\left\{w_{it} \geq 9.90\right\}_i \\ + \gamma_{3B} \times (log(w_{it}) - log(10.00)) \times \mathbb{1}\left\{w_{it} \geq 10.00\right\}_i + \Lambda \mathbf{X}_{it} + \epsilon_{it} \end{aligned} \tag{10}$$

[13]The break points for last quarter's weekly hours were 10, 20, 35, 45 and 55
[14]The break points for firm size are 50, 100, 500, 1000, 5000

Here we test whether the slope between \$9.90 and \$10.00 (i.e., $\eta_{1B} + \gamma_{2B}$) is greater than the average of the slopes below \$9.90 and above \$10.10 $\left(\frac{1}{2} \times \eta_{1B} + \frac{1}{2} \times (\eta_{1B} + \gamma_{2B} + \gamma_{3B})\right)$; or equivalently to test $\beta_{spline} = \gamma_{2B} - \gamma_{3B} > 0$.

Finally, our most flexible specification including 10-cent dummies spanning 9.50 to 10.40 (where the 9.50 dummy includes wages between 9.50 and 9.59):

$$S_{it} = \sum_{k \in S} \delta_k \mathbb{1}\left\{w_i = k\right\}_i + \Lambda \mathbf{X}_{it} + \epsilon_{it} \tag{11}$$

And then calculates the following statistics:

$$\delta_{jump} = (\delta_{10.00} - \delta_{9.90})$$

$$\beta_{local} = (\delta_{10.00} - \delta_{9.90}) - \frac{\left(\sum_{k=9.80,k\neq 10.00}^{10.20} \delta_k - \delta_{k-0.10}\right)}{4}$$

$$\beta_{global} = (\delta_{10.00} - \delta_{9.90}) - \frac{1}{10}\left(\delta_{10.40} - \delta_{9.50}\right)$$

The β_{local} estimate provides us with a comparison of the jump between \$9.90 and \$10.00 to other localized (10 cent) changes in separation probability from \$0.10 increases. In contrast, β_{global} provides us with a comparison of the jump with the full global (log-linear) average separation response from varying the wage between \$9.50 and \$10.40.

Columns 1-6 of TableOnline Appendix D show estimates using the jump specification of equation 9, with the baseline controls, additional firm size and sector controls, and additional firm FE controls. Columns 1,3, and 5 report the semi-elasticity for separations without any jumps. The semi-elasticities range between -0.215 and -0.310. Given the mean quarterly separation rate of 0.13 in our estimation sample, the implied separation elasticities are between -1.7 and -2.4. In turn, if we use the Manning steady state assumption of firm labor supply elasticity being twice the separation elasticity, the labor supply elasticities range between 3.3 and 4.8. These are consistent with a moderate amount of monopsony power in the low-wage labor market in Oregon over this period.[15]

Next, when we add indicator for \$10/hour or more, we do not see any indication that the separation rate changes discontinuously at \$10. The coefficients are small and not statistically distinguishable from zero. In columns 7 and 8 we report the results from equations (10) and (11); specifically we report β_{spline}, β_{local},and β_{global} and associated standard errors. Neither the spline, local or global comparisons indicate any discontinuous change in the separation rate around \$10. Finally, in Figure 5, we plot the coefficient on each 10 cent dummies between \$9.50 and \$10.40 associated with equation (11). Consistent with the findings in the table, there is a general negative relationship between wage and separation, with a similar implied separations elasticity as the baseline specification above. However, there is no indication of a discontinuous drop in the separation rate going from \$9.90 to \$10.00. Again, we can use column (2) as the estimate most closely related to our theoretical model. The 95 percent confidence interval for the estimate of γ is (0.986, 1.050), and similarly to the online labor market, it is concentrated around one.

4.5 Stated Preference Experiments on Worker Mobility

A final piece of evidence on worker left-digit bias comes from stated preference experiments we conducted to see how responsive job mobility is to the outside offered wage. To leverage firm-level wage variation, we focus on workers at a single large retailer: using Facebook we reached a sample of current or recent employees at Wal-Mart, the largest private employer

[15]One can imagine that even conditional on these individual level covariates (such as starting wage, hours of work, firm), wage differences across workers can reflect other forms of heterogeneity such as learning about match quality or other time varying productivity differences. Bassier et al. (2019) uses the data from Oregon but uses a different strategy that focuses on wage policy differences across firms: namely it regresses the separation rate on firm effect to isolate the part of wage differences that are more plausibly due to differences in firm-level wages after accounting for individual fixed effects. Their estimates of the labor supply elasticity (using a broader set of Oregon workers) for the 2003-2007 period is around 3.6, which is fairly similar to what we find here. Here, however, our key interest is in detecting left-digit bias among workers, by testing for discontinuties in the worker-level separation elasticity around \$10.00, as opposed to the causal interpretation of the firm-level slopes around \$10.00, and this exercise requries using individual and not firm-level wage variation.

in the U.S.. After asking about their current wage and working conditions, we presented them with 3 possible jobs, where we randomly varied wages (and job amenities) around their current job values, and asked them if they would leave their current employment if offered such a job [16] The randomized wage offer allows us to estimate a recruit elasticity among workers in a specific firm (Walmart) and allows us to test whether the willingness to stay at the current job falls discontinuously when the offered wage crosses a round number (integer). In our sample of 312 workers, the 10th percentile was \$9.75 while the 90th percentile was \$17.00 dollars. The gap between the randomized offered wages and current wages was \$0.07 with a standard deviation of \$1.80.

We estimated an analogous set of regressions as in our MTurk and Oregon samples. Since this sample has multiple round number wage thresholds, here we construct a running variable d_{it} which is the distance from the nearest round number, w_R. However, to estimate comparable regressions to the MTurk and Oregon samples (and obtain a recruitment elasticity), we need a log wage variable; we create a (recentered) wage variable $\tilde{w}$ that centers d_{it} around the sample median round wage, \$12: $\tilde{w}_{it} = d_{it} + 12$. We also include fixed effects for each \$1 intervals such as \$9.50 to \$10.49, \$10.50 to \$11.49, etc., to ensure the identifying variation is only coming from within the interval around the round number. Finally, because each respondent was given up to 3 randomized wage offers, and because randomization was around the respondent's current wage, we also include a fixed effect for each respondent, to ensure all wage variation is from the randomization.

In the baseline specification, we regress an indicator for remaining on the job, R_{it} on log wages, and an indicator for earning more than \$10.00.

$$R_{it} = \beta_0 + \eta_1 log(\tilde{w}_{it}) + \gamma_{1A} \mathbb{1}\left\{w_i \geq w_R\right\}_i + \Lambda \mathbf{X}_{it} + \epsilon_{it} \tag{12}$$

Here, the controls $\mathbf{X_{it}}$ include the \$1 interval and respondent fixed effects. We run

[16] The work conditions included commute time, assessment of co-workers and supervisors, scheduling, and other characteristics, shown in the survey excerpt in Appendix FigureE.2.

analogous regressions to equations (10) and (11) above, and estimate analogous test statistics (i.e., jump, spline, local and global changes). Note that in the equivalent of equation (11), we estimate the remaining probabilities by \$0.10 intervals as before.

Column 1 of Table 3 first reports the elasticity of remaining with respect to the log of the randomized offered wage, $log(w_{it})$. The coefficient of -0.886 is statistically significant at the 1 percent level and has an economically meaningful magnitude. At the same time, the elasticity of remaining at Wal-Mart with respect to the outside offered wage is -0.886/0.54 = -1.64, which suggests only a modest impact of a higher wage on the ability to entice a current worker to move–consistent with the importance of other job-based amenities and/or mobility cost. In column 2, we report the estimate but using the recentered wage $\tilde{w}_{it}$ along with interval fixed effects. The coefficient of -0.982 is similar in magnitude but of lower precision due to suing only within interval variation. Column (3) estimates the jump model, and finds no evidence of a jump in the remain probability at the round number; the coefficient -0.026 (SE=0.036) can rule out $\hat{\gamma}$ outside of (0.91,1.05). None of the other tests (spline, local or global) in columns 4 and 5 suggest any discontinuities, though the estimates from the spline specification are imprecise. However, the local and global jump estimates are more precise, and the corresponding Figure 6 confirms that while the probability of staying at the current job is declining with the offered outside wage, there is no indication of any discontinuity at the round number.

5 Incorporating employer optimization frictions

Given the lack of evidence for worker left-digit bias, we now consider the other possible explanation for bunching: employer mis-optimization. We extend the model to allow employers to "benefit" by paying a round number, despite lowered profits.[17] While consistent with employers preferring to pay round numbers, it could reflect internal fairness

[17] It would be equivalent to assume that firms suffer an effective loss from not paying a round number.

constraints or administrative costs internal to the firm. These could be transactions costs involved in dealing with round numbers, cognitive costs of managers, or administrative costs facing a bureaucracy. δ is a simple way to capture satisficing behavior by firms willing to use a simple heuristic (choose nearest round number) instead of bearing the costs of locating at the exact profit-maximizing wage. These costs may be substantial, as evidenced by the prevalence of pay-setting consultancies[18] and pervasive use of round-numbers in publicly stated wage-policies of large firms.[19]

The presence of δ modifies the profit function to be:

$$\pi(w, p) = (p - w)l(w, p)e^{\delta \mathbf{1}_{w=w_0}} \tag{13}$$

where δ is the percentage "gain" in profits from paying the round number.[20] This specification parallels that in Chetty (2012), who restricts optimization frictions to be constant fractions of optimal consumer expenditure (in the nominal model), except applied to the employer's choice of wage for a job rather than a consumer's choice of a consumption-leisure bundle. In the taxable income model, optimization frictions parameterize the lack of responsiveness to tax incentives, while in our model they parameterize the willingness to forgo profits in order to pay a round number.

Given (1) and (13), profits from paying a wage w to a workers with marginal product p can be written as:

$$\pi(w, p) = (p - w)\frac{w^{\eta}}{C} e^{\eta\gamma \mathbb{1}_{w \geq w_0}} e^{\delta \mathbf{1}_{w=w_0}} k\left(p\right) = (p - w)l^{*}(w, p)e^{\eta\gamma \mathbb{1}_{w \geq w_0}} e^{\delta \mathbf{1}_{w=w_0}} \tag{14}$$

[18] Companies and consortia such as ADP, The Survey Group, The Mayflower Group, Paychex, and Payscale all sell consultancy services for pay setting, informed by the shared payroll data of member companies.

[19] The National Employment Law Project (2016) documents a large number of voluntary wage policies by employers. McDonald's, T.J. Maxx, The Gap, and Walmart all voluntarily adopted a $10.00 base wage in 2015/2016, and many other firms have company wage policies that mandate round numbers from $9.00 (Target) to $18.00 (Hello Alfred).

[20] While we do not microfound why employers may have preferences for paying a particular round number, this may reflect inattention among wage-setters. For example, Matějka (2015) shows that rationally inattentive monopoly sellers will choose a discrete number of prices even when the profit-maximizing price is continuous.

which can be written in terms of the nominal model ρ as:

$$e^{\eta\gamma\mathbb{1}_{w*(p)<w_0}}e^{\delta} > \frac{\rho\left(w^*\left(p\right),p\right)}{\rho\left(w_0,p\right)} \tag{15}$$

Taking logs, we obtain that a firm will pay the round number if

$$\delta + \eta\gamma\mathbb{1}_{w*(p)<w_0} > \ln\rho\left(w*\left(p\right),p\right) - \ln\rho\left(w_0,p\right) \tag{16}$$

This shows that bunching is more likely the greater is the left-digit bias of workers and the optimization cost for employers. The optimization bias is symmetric whether the latent wage is above or below the round number. But left-digit bias is asymmetric because it only has an impact if the latent wage is below the round number. The right-hand side of (16) can be approximated using the following second-order Taylor series expansion of $\rho\left(w_0,p\right)$ about $w^*\left(p\right)$[21]:

$$\ln\rho\left(w_0,p\right) \simeq \ln\rho\left(w^*,p\right) + \frac{\partial\ln\rho\left(w^*,p\right)}{\partial w}\left[w_0 - w^*\right] + \frac{1}{2}\frac{\partial^2\ln\rho\left(w^*,p\right)}{\partial w^2}\left[w_0 - w^*\right]^2 \tag{17}$$

The first-order term is zero by the definition of the latent wage (Akerlof and Yellen (1985) use this idea to explain price and wage rigidity). Using the definition of the nominal model, the second derivative can be written as:

$$\frac{\partial^2\ln\rho\left(w,p\right)}{\partial w^2} = -\frac{1}{\left(p-w\right)^2} - \frac{\eta}{w^2} \tag{18}$$

Using (2) this can be written as:

$$\frac{\partial^2\ln\rho\left(w^*,p\right)}{\partial w^2} = -\frac{\eta\left(1+\eta\right)}{w^{*2}} \tag{19}$$

[21]One can use the actual profit function, instead of the approximation, but the difference is small for the parameters we use, and the approximation has a clearer intuition.

where it is convenient to invert (2) and express in terms of the latent wage because wages are observed but marginal products are not. Substituting (19) into (17) and then into (16) leads to the following expression for whether a firm pays the round number:

$$\frac{1}{2}\left[\frac{w_0 - w^*}{w^*}\right]^2 \equiv \frac{\omega^2}{2} \leq \frac{\delta + \eta\gamma\mathbb{1}_{w*<w_0}}{\eta\left(1+\eta\right)} \tag{20}$$

The left-hand side of (20) implies that the size of the loss in nominal profits from bunching is increasing in the square of the proportional distance of the latent wage from the round number (ω). The right-hand side tells us that, for a given latent wage, whether a firm will bunch depends on the extent of left-digit bias as measured by γ (only relevant for wages below the round number), the extent of optimization frictions as measured by δ and the degree of competition in the labor market as measured by η. The extent of labor market competition matters because the loss in profits from a sub-optimal wage are greater the more competitive is the labor market. Define:

$$z_0 = \frac{\delta + \eta\gamma}{\eta\left(1+\eta\right)}, z_1 = \frac{\delta}{\eta\left(1+\eta\right)} \tag{21}$$

Assume, for the moment, that there is some potential variation in (δ, γ, η) across firms which is independent of the latent wage and leads to a CDF for z_0 of $\Lambda_0^z(z)$ and a CDF for z_1 of $\Lambda_1^z(z)$. From (21) it must be the case that $\Lambda_0^z(z) \leq \Lambda_1^z(z)$ with equality if there is no left-digit bias. The way in which we use this is the following—suppose the fraction of firms who bunch is denoted by $\phi(\omega^*) = \phi\left(\frac{w_0 - w^*}{w^*}\right)$, where ω^* is the proportionate gap between the latent optimal wage under the nominal model,w^*, and the round number w_0, with $\phi(\omega_*)$ defined similarly as $\phi(\frac{w_* - w}{w_*})$. Then (20) implies that we will have, for $\omega < 0$, :

$$\phi\left(\omega_*\right) = 1 - \Lambda_0^z\left[\frac{\omega_*^2}{2}\right] \tag{22}$$

and for $\omega > 0$:

$$\phi(\omega^*) = 1 - \Lambda_1^z \left[\frac{\omega^{*2}}{2}\right] \tag{23}$$

In the next section, we empirical recover estimates of the left-hand sides of (22) and (23). The results in this section imply that these estimates of the source of the missing mass in the wage distribution can be used to nonparametrically identify the distributions of z_0 and z_1, Λ_0 and Λ_1. However, these estimates alone do not allow us to nonparametrically identify the distribution of (δ, γ, η), the underlying economic parameters of interest, and below we estimate the degree of monopsony and employer misoptimization under a variety of parametric assumptions.

5.1 Estimating the origin of the missing mass

The excess mass in the wage distribution at a bunch that has been documented in the previous sections must come from somewhere in the latent wage distribution that would result from the "nominal model" without any bunching (in the terminology of Chetty (2012)).[22] This section describes how we estimate the origin of this "missing mass." To do so, we follow the now standard approach in the bunching literature of fitting a flexible polynomial to the observed distribution, excluding a range around the threshold, and using the fitted values to form the counterfactual at the threshold (see Kleven 2016 for a discussion).

We focus on the bunching at the most round number ($10.00 in the wage data, $1.00

[22]Following the literature, our procedure assumes that the missing mass is originating entirely from the surrounding basin. In principle, it is possible that the missing mass is originating from latent non-employment—i.e., jobs that would not exist under the nominal model in the absence of bunching. However, the extent to which some of the excess jobs at $10 is coming from latent non-employment, one would need to assume either that (1) these jobs have latent productivity exactly at $10.00 so that employers are indifferent between entering and not entering, or (2) they have productivity greater than $10 but have a fixed cost of not paying exactly $10 that is independent of the size of the profits from paying different wages under the nominal model. Both of these assumptions strike us as implausible. As an empirical matter, if some of the excess mass at $10 are originating from latent non-employment, the estimated missing mass around $10 would be smaller in magnitude than the excess mass at $10. However, our estimated missing mass from the surrounding basin is, indeed, able to account for the size of the excess mass—which suggests that latent non-employment is unlikely to be an important contributor to the excess mass in our case.

in the MTurk rewards data). We ignore the secondary bunches; this will attenuate our estimate of the extent of bunching, as we will ignore the attraction that other round numbers exert on the distribution.

We use bin-level counts of wages c_w in, say, \$0.10 bins, and define $p_w = \frac{c_w}{\sum_{j=0}^{\infty} c_j}$ as the normalized count or probability mass for each bin. We then estimate:

$$p_w = \sum_{j=w_0-\Delta w}^{w_0+\Delta w} \beta_j \mathbb{1}_{w=j} + \sum_{i=0}^{K} \alpha_i w^i + \epsilon_w \tag{24}$$

In this expression j sums over 10 cent wage bins (we use 1 cent bins in the MTurk data), and the $g(w) \equiv \sum_{i=0}^{K} \alpha_i w^i$ terms are a K^{th}order polynomial, while β_j terms are coefficients on dummies for bins in the excluded range around w_0, between $w_L = w_0 - \Delta w$ and $w_H = w_0 + \Delta w$. β_{w_0} is the excess bunching (EB) at w_0. In addition, $\sum_{j=w_0-\Delta w}^{w_0-10} \beta_j$ is the missing mass strictly below w_0 (MMB), while $\sum_{j=w_0+10}^{w_0+\Delta w} \beta_j$ is the missing mass strictly above w_0 (MMA). However, it might be likely that the the number of workers is increasing in w; so our test for left-digit bias tests whether $\sum_{j=w_0-\Delta w}^{w_0-10} \frac{\beta_j}{g(j)} = \sum_{j=w_0+10}^{w_0+\Delta w} \frac{\beta_j}{g(j)}$. In other words, we test whether the share of workers would have been paid the latent wage j but are instead paying w_0 is different when considering bunchers from above versus below.

Since Δw is unknown, we use an iterative procedure similar to Kleven and Waseem (2013). Starting with $\Delta w = 10$, we estimate equation 24 and calculate the excess bunching EB and compare it with the missing mass $MM = MMA + MMB$. If the missing mass is smaller in magnitude than the excess mass, we increase Δw and re-estimate equation (24). We do this until we find a Δw such that the excess and missing masses are equalized. Since Δw is itself estimated, we estimate its standard error using a bootstrapping procedure suggested by Chetty (2012) and Kleven (2016). In particular, we resample (with replacement) the errors $\hat{\epsilon}_w$ from equation (24) and add these back to the fitted $\hat{p}_w$ to form a new distribution $\tilde{p}_w$, and estimate regression (24) using this new outcome. We repeat this 500 times to derive the standard error for Δw. The estimate of Δw and its standard error will be useful later for the estimation of other parameters of interest.

In Figure 7 we show the estimates for the administrative data from MN, OR, and WA, using polynomial order $K = 6$. For visual ease, we plot the kernel-smoothed $\hat{\beta}_j$ for the missing mass. Even leaving out the prominent spike at \$10.00, the wage distribution is not smooth, and has relatively more mass at multiples of 5, 10 and 25 cents. For this reason, it is easier to detect the shape of the missing mass by looking at the kernel-smoothed $\hat{\beta}_j$. Moreover, we show the excess and missing mass relative to the counterfactual $\widehat{p_w^C} = \sum_{i=0}^{6} \alpha_i w^i$. There is clear bunching at \$10.00 in the administrative data, consistent with evidence from the histogram above. We find that the excess bunching can be accounted for by missing mass spanning $\Delta w = \$0.80$; we can also divide Δw by w_0 and normalize the width as $\omega = \frac{w_H - w_0}{w_0} = 0.08$. Visually, the missing mass is coming from both below and above \$10.00, which is relevant when considering alternative explanations.

These estimates are also reported in Table 4, column 1. The bunch at \$10.00 is statistically significant, with a coefficient of 0.010 and standard error of 0.002. In addition, the size of the missing mass from above and below w_0 are quantitatively very close, at -0.006 and -0.007 respectively. The t-statistic for the null hypothesis that the missing mass (relative to latent) are equal for bunchers from above and below is 0.338. This provides strong evidence against worker left-digit bias, which would have implied an asymmetry in the missing masses. The width of the missing mass interval is $\omega = 0.08$, with a standard error of 0.027. In other words, employers who are bunching appear to be paying as much as 8% above or below the wage that maximizes profits under the nominal model.

In column 2, we use the CPS data limited to MN, OR, and WA only. We find a substantially larger estimate for the excess mass, around 0.043. In column 3, we report estimates using the re-weighted CPS counts for MN, OR, and WA adjusted for rounding due to reporting error using the 1977 supplement (CPS-MEC). The CPS estimate of bunching adjusted for measurement error is much closer to the administrative data, with an estimated magnitude of 0.014; while it is still somewhat larger, we note that the estimate from the administrative data is within the 95 percent confidence interval of the CPS-MEC estimate.

In column 4, we use the raw CPS data for all states and find the excess mass estimate of 0.041. Therefore, while some of the gap between the all-state CPS and the MN-OR-WA administrative data estimates is due to the differences in samples (MN, OR, and WA versus all states), most of it is due to rounding error of respondents in the CPS. The use of the CPS supplement substantially reduces the discrepancy, which is re-assuring. At the same time, we note that the estimates for ω using the CPS (0.07) are remarkably close to those using the administrative data (0.08). The graphical analogue of column 3 is in Figure 8.

Since the counterfactual involves fitting a smooth distribution using a polynomial in the estimation range, in Table 5 we assess the robustness of our estimates to alternative polynomial orders between 2 and 6. Both the size of the bunch, and the width of the interval with missing mass, ω, are highly robust to the choice of polynomials. For example, using the pooled administrative data, the bunching β_0 is always 0.01, and ω is always 0.08 for all polynomial orders K.

One concern with bunching methods in cross sectional data is that the estimation of missing mass requires parametric extrapolation of the wage distribution around \$10. In our case, however, the bunching is at a nominal number (\$10) that sits on a different part of the real wage distribution in each of the 20 quarters of our sample. As an alternative, instead of collapsing the data into a single cross section, we use quarterly cross sectional data and fit a polynomial in the real wage $w_r = w/P_t$ where P_t is the price index in year t relative to 2003.

$$p_{w_r} = \sum_{j=w_0-\Delta w}^{w_0+\Delta w} \beta_j \mathbb{1}_{w_r \times P_t = j} + \sum_{i=0}^{K} \alpha_i w_r^i + \epsilon_{w_r} \tag{25}$$

We again iterate estimating this equation until $MM = MMA + MMB$ to recover Δw. If the real wage distribution is assumed to be stable during this period (i.e. the $\alpha_i$are constant over time), then in principle the latent wage distribution within the bunching interval can be identified nonparametrically, because each w_r bin falls outside of the bunching interval in at least some periods. More precisely, suppose there were only two periods, and $(w_0 - \Delta w)/P_{T_1} \geq (w_0 + \Delta w)/P_{T_0}$, for some T_1 and T_0. In this case β_j is identified from the mass at $w_r \times P_{T_1}$ controlling for a flexible function of w_r which is effectively identified from the real wage distribution in T_0 as well as the mass at $w_r \times P_{T_0}$ conditional on the real wage density in T_1. This specification is an example of a "difference in bunching" approach that compares the same part of the real wage distribution across years (Kleven (2016)), and addresses criticisms of bunching estimators being dependent on parametric assumptions about the shape of the latent distribution (Blomquist and Newey, 2017). To show that this assumption of non-overlapping bunching intervals is satisfied for at least some portion of our data, Appendix Figure A.2 shows that the bunching interval around the nominal \$10.00 mode in 2007 does not overlap with that from the 2003 real wage distribution, allowing for estimation of the latent (real) density around the nominal \$10.00 mode using variation in the price level over time. In column (8) of Table 5 we show that estimates with the repeated cross section and real wage polynomials are virtually identical to our baseline estimates, providing reassurance that our estimates are not being driven by parametric assumptions about the latent distribution within the bunching interval.

The main conclusions from this section are that the missing mass seems to be drawn symmetrically from around the bunch and from quite a broad range. As the next section shows, these facts are informative about possible explanations for bunching and the nature of labor markets.

6 Recovering optimization frictions from bunching estimates

The first result of our framework above is that worker left-digit bias implies that the degree of bunching is asymmetric, in that missing mass will come more from below the round number than above. Thus, finding symmetry in the origin of the missing mass implies that we can approximate ω^* and ω_* with the harmonic mean of the two, which we denote $\omega \equiv \frac{w - w_0}{w_0}$, and is exactly the proportional width of the basin of attraction in Table 4 .

This further implies that $\Lambda_0 = \Lambda_1$ and allows us to accept the hypothesis that $\gamma = 0$. The intuition for this is that left-digit bias implies that firms with a latent wage 5 cents below the round number have a higher incentive to bunch than those with a latent wage 5 cents above. We fail to reject symmetry of the missing mass in Table 4 and so we proceed holding $\gamma = 0$.

Under the assumption that bunchers have latent wages near the round number, the presence of missing mass greater than w_0 also rules out many imperfect competition stories that do not require monopsony in the labor market. If the labor market were perfectly competitive, then no worker could be *underpaid*, even though misoptimizing firms could still *overpay* workers. Explanations involving product market rents or other sources of profit for firms cannot explain why firms systematically can pay below the marginal product of workers; only labor market power can account for this. Similarly, however, the presence of missing mass below w_0 rules out pure employer collusion around a focal wage of w_0, as the pure collusion case would imply that all the missing mass was coming from *above* w_0.

Taking $\gamma = 0$ as given, our estimates of the proportion of firms who bunch for each latent wage identifies the CDF of $z_1 = \frac{\delta}{\eta(1+\eta)}$, but does not allow us to identify the distributions of δ and η separately. This section describes how one can make further assumptions to identify these separate components. First, note that if there is perfect competition in labor markets ($\eta = \infty$) or no optimization frictions ($\delta = 0$), we have that $z_1 = 0$ in which case there would be no bunches in the wage distribution. The existence of bunches implies that we can reject the joint hypothesis of perfect competition for all firms and no optimization frictions for all firms. But there is a trade-off between the extent of labor market competition and optimization friction that can be used to rationalize the data on bunches. To see this note that if the labor market is more competitive i.e. η is higher, a higher degree of optimization friction is required to explain a given level of bunching. Similarly, if optimization frictions are higher i.e. a higher δ, then a higher degree of labor market competition is required to explain a given level of bunching.

To estimate η and δ separately from $\phi(\omega)$, we need to make assumptions about the joint distribution. A natural first place to start is to assume a single value of η and a single value of δ. In this case, the missing mass takes the form of a flat basin of attraction around the whole number bunch with all latent wages inside the basin bunching and none outside. If there is no left-digit bias ($\gamma = 0$) (because of the symmetry in the missing mass), then ω , η and δ must satisfy:

$$\frac{2\delta}{\eta\,(1+\eta)} = \omega^2 \tag{26}$$

This expression shows that, armed with an empirical estimate of ω, we can draw a locus in δ-η, showing the values of δ and η that can together rationalize a given ω. For a given size of the basin, a higher value of optimization frictions (higher δ) implies a more competitive labor market (a higher η). [23]

But our estimates of the "missing mass" do not suggest a basin with this shape. At all latent wages, there seem to be some employers who bunch and others who do not. To rationalize this requires a non-degenerate distribution of δ and/or η . We make a variety of different assumptions on these distributions in order to investigate the robustness of our results.

We always assume that the distributions of η and δ are independent with cumulative distributions $H(\eta)$ and $G(\delta)$. At least one of these distributions must be non-degenerate because, by the argument above, if they both have a single value for all firms one would observe an area around the bunch where all firms bunch so the missing mass would be 100% - this is not what the data look like. Our estimates imply that there are always some firms who do not bunch, however close is their latent wage to the bunch. We rationalize this as being some fraction of employers who are always optimizers i.e. have $\delta = 0$.

[23] Andrews, Gentzkow and Shapiro (2017) make a similar point in a different context, arguing that differing percentages of people with optimization frictions can substantively affect other parameter estimates using the example of DellaVigna, List and Malmendier (2012).

We first make the simplest parametric assumptions that are consistent with the data: we assume that η is constant, and δ has a 2-point distribution with δ=0 with probability $\underline{G}$ and $\delta = \delta^*$ with probability $1 - \underline{G}$, so that $E[\delta|\delta > 0] = \delta^*$. Below, we will extend this formulation to consider other possible shapes for the distribution $G(\delta|\delta > 0)$, keeping a mass point at $G(0) = \underline{G}$.

This then implies the missing mass at w is given by:

$$\phi(\omega) = [1 - \underline{G}]\, I \left[\omega^2 < \frac{2\delta^*}{\eta\,(1+\eta)}\right]$$

In this model, the share of jobs with a latent wage close to the bunch that continue to pay a non-round w identifies $\underline{G}$, and the width of the basin of attraction in the distribution identifies $\frac{\delta^*}{\eta(1+\eta)}$. The width of the basin was estimated, together with its standard error, in the estimation of the missing mass where, relative to the bunch, it was denoted by $\frac{\Delta w}{w_0}$. Under assumptions about δ^* we can recover a corresponding estimate of η and vice versa.

What do plausible values of optimization error imply about the likely labor supply elasticities for bunchers? To answer this question, we report bounds using "economic standard errors" similar to Chetty (2012). We calculate estimates of η assuming δ^* equal to 0.01, 0.05 or 0.1 in rows A, B, and C of Table 6 respectively. The implied labor supply elasticity η varies between 1.337 and 5.112 when we vary δ^* between 0.01 and 0.1. Even assuming a substantial amount of mis-optimization (around 10% of profits) suggests a labor supply elasticity facing a firm of less than 5.5, and we can rule out markdowns smaller than 6 percent. If we assume, instead, a 1% loss in profits due to optimization friction, the 90 percent confidence bounds rule out $\eta > 3.1$ and markdowns smaller than 25 percent. While our estimate for the labor supply elasticity are not highly precise, the extent of bunching at $10.00 suggests considerable wage setting power on firms' part even for a sizable amount of optimization frictions, δ.

The admissible values of δ, η can also be seen in Figure 9. Here we plot the δ^*, η locus for the sample mean of estimated bunching, ω, as well as for the 90 percent confidence interval around it. We can see visually that as we consider higher values of δ^*, the range of admissible $\eta's$ increases and becomes larger in value. However, even for sizable δ^*'s the implied values of the labor supply elasticity are often modest, implying at least a moderate amount of monopsony power. Our estimates are plausible given the recent literature: Caldwell and Oehlsen (2018) find an experimental labor-supply elasticity for Uber drivers at between 4 and 5, similar to the estimates in Dube et al. (2015), while Kline et al. (2017) estimate a labor-supply elasticity facing the firm of 2.7, using patent decisions as an instrument for firm productivity, both of which would be well within the range of η implied by our estimates together with a δ^* less than 0.05.

We examine robustness of the estimates to alternative specifications of the latent distribution of wages in Table 7. Columns 1 and 2 add indicator variables for "secondary" modes, to capture the bunching induced at 50 cent and 25 cent bins. Columns 3 and 4 specify the latent distribution as a Fourier polynomial, in order to allow the specification to pick up periodicity in the latent distribution that even a high-dimensional polynomial may miss. Columns 5 and 6 of Table 7 explore changing the degree of the polynomial used to fit the main estimates in Table 6. Column 5 uses a quadratic and column 6 uses a quartic, and our results stay very similar to our main estimates in Table 6.

6.1 Alternative assumptions on heterogeneity

While assuming a single value of non-zero δ and a constant elasticity η may seem restrictive, it is a restriction partially made for empirical reasons as our estimate of the missing mass at each latent wage is not very precise and we will also be unable to distinguish heterogeneous elasticities in our experimental design. Nonetheless, there is a concern that different assumptions about the distribution of δ and η might be observationally indistinguishable but have very different implications for the extent of optimization frictions and monopsony power in the data. This section briefly describes a number of robustness exercises that vary the possible heterogeneity in δ and η, with details relegated to Appendix Online Appendix A .

While it is not possible to identify arbitrary nonparametric distributions of δ and η, as robustness checks we consider polar cases allowing each to be unrestricted one at a time, and then finally a semi-parametric deconvolution approach that allows for an unrestricted, non-parametric distribution $H(\eta)$, along with a flexible, parametric distribution $G(\delta)$. First, we continue to assume a constant η but allow δ to be have an arbitrary distribution $G(\delta|\delta > 0)$ while continuing to fix the probability that $\delta = 0$ at $\underline{G}$. Second, at the opposite pole, we allow each firm to have its own labor supply elasticity η, while each firm either mis-optimizes profits by a fixed fraction δ^* or not at all.[24] Finally, we continue to allow arbitrary heterogeneity in η but only restrict $G(\delta)$ to have a continuous lognormal distribution, with prespecified variances of .1 and 1.

We quantitatively show robustness of our main estimates to these four alternate specifications in Table 8**.** Column 1 shows the implied $E[\delta|\delta > 0]$ and $\bar{\delta}$ when an arbitrary distribution of δ is allowed. The implied η for $E[\delta|\delta > 0] = 0.01$ is 1.67 instead of 1.33 in the baseline estimates from Table 6. Similarly, in column 2 we see the estimates under the 2-point distribution for δ and an arbitrary distribution for η. The mean η of 1.56 in this case is quite close to column 1. The implied bounds are somewhat larger, with a 1% loss in profits for those bunching (i.e., $E(\delta|\delta > 0) = 0.01$) generating 95% confidence intervals that rule out estimates of 5.4 or greater. Under 5% loss in profits, we get elasticities in columns 1 and 2 that are close to 4 , somewhat larger than the comparable baseline estimate of 3.5, but with similarly close to 20 percent wage markdown. Therefore, allowing for heterogeneity in either δ or η only modestly increases the estimated mean η as compared to our baseline

[24] This exercise is in the spirit of Saez (2010) who estimates taxable income elasticities using bunching in income at kinks and thresholds in the tax code (Kleven 2016). Kleven and Waseem (2013) use incomplete bunching to estimate optimization frictions, similar to our exercise in this paper; however, in our case optimization frictions produce bunching while in Kleven and Waseem (2013) they prevent it. This has been applied to estimating the implicit welfare losses due to various non-tax kinks, such as gender norms of relative male earnings (Bertrand, Kamenica and Pan 2015) as well as biases due to behavioral constraints (Allen et al. 2016).

estimates.

In columns 3 and 4 we report our estimates allowing for an arbitrary distribution for η, along with a lognormal conditional distribution for δ. These estimates are obtained using a deconvolution estimator to recover the distribution of a difference in random variables, described in more detail in the Appendix. As in columns 1 and 2, we consider the case where $E(\delta|\delta > 0) = 0.01$ or 0.05, but now allow the standard deviation σ_δ to vary. In column 3 we take the case where δ is fairly concentrated around the mean with $\sigma_\delta = 0.1$. Here the estimated $E(\eta)$ is equal to 2.5, which is larger than the analogous baseline estimates in columns 1 and 2 allowing for an arbitrary distributions for δ and η, respectively. In column 4, we allow δ to be much more dispersed, with $\sigma_\delta = 1$. In this case the estimated $E(\eta)$ falls somewhat to 2. With $E(\delta|\delta > 0) = 0.05$, we get $E[\eta] = 6$ and 4.6 under $\sigma_\delta = 0.1$ and $\sigma_\delta = 1$, respectively, and we are able to rule out markdowns less than 5 percent easily. Encouragingly, for a given mean value of optimization friction, $E[\delta|\delta > 0]$, allowing for heterogeneity in δ and η together only modestly affects the estimated mean η as compared to our baseline estimates.

Our conclusion from this investigation is that our qualitative finding of significant monopsony power remains robust to a wide range of assumptions made about the distribution of δ and η. The robustness to the forms of modeling heterogeneity is noteworthy in light of Blomquist and Newey (2017), which suggests bunching estimators often rely on strong distributional assumptions to obtain identification.

6.2 Heterogeneous δ by Firms - Evidence from Oregon Administrative Data

We can now examine what type of firms are more likely to bunch exactly at $10.00 to provide additional validation for the hypothesis that this reflects employer mis-optimization. If heterogeneity in δ is indeed driving variation in bunching, then firms that bunch should exhibit plausibly "behavioral" characteristics, for example a lack of administrative sophistication or modern management practices. While we do not have extensive firm-level data on human resource practices, we do have a variety of firm characteristics from the matched employer-employee data from Oregon. In Figure 10 we show the firm-specific determinants of bunching, where we regress the extent of bunching (ratio of excess mass to latent wage density) on a variety of firm level characteristics, including firm size, deciles of hourly wage firm effects[25] , industry, and part-time (less than 20 hours) share of employment. While bunching is present in every subgroup, with no coefficient less than 1, we find that the firms most likely to bunch are very small firms, those in the construction sector , and those in the bottom of the firm-effect distribution for hourly wages. This is consistent with these firms being relatively "unsophisticated", more likely to pay in cash and less likely to have standardized pay practices (e.g. automatic inflation escalation) that would eliminate employer discretion in wage setting; this results in bunching of wages at round numbers. The public sector also exhibits some unusual bunching, which could reflect low cost-mininimization pressure on public wage-setting, but may also indicate some preference of public sector workers or unions for benefits over wages. Starting wages are also more likely to bunch; this is consistent with firms using other rules (such as common percentage increase in wages) that erode bunching at round numbers over time.

In Appendix Table C.1 we examine heterogeneity in η by worker characteristics, holding fixed δ and using measurement error corrected CPS data. The estimates are consistent with plausible heterogeneity in residual labor supply elasticities: women have lower estimated η while new workers have higher values, but the extent of heterogeneity is generally limited.

[25]These firm effects are estimated using the Abowd et al. (1999) decomposition, with the Oregon-specific hourly wage decomposition results detailed more fully in Bassier et al. (2019).

7 Conclusion

Significantly more U.S. workers are paid exactly round numbers than would be predicted by a smooth distribution of marginal productivity. This fact is documented in administrative data, mitigating any issues due to measurement error, and is present even in Amazon MTurk, an online spot labor market, where there are no regulatory constraints nor long-term contracts. We integrate imperfect labor market competition with left-digit bias by workers and a general employer preference for round-number wages to evaluate the source of left-digit bias. Using administrative wage data, we reject a role for worker left-digit bias using the symmetry of the missing mass around round numbers. We also reject the left-digit bias hypothesis using a high-powered, preregistered experiment conducted on MTurk: despite considerable monopsony power (in a putatively thick market), there is no discontinuity in labor supply or quality of work at 10 cents relative to 9. Observational evidence from administrative hourly wage data from Oregon similarly confirms the lack of any discontinuity in labor supply facing firms at $10/hour.

This evidence shows that the extent of round-number bunching can be explained by a combination of a plausible degree of monopsony together with a small degree of employer mis-optimization. We show that when there is sizable market power, it requires only a modest extent of optimization error to rationalize substantial bunching in wages. With optimization error less than 5% of profits, the observed degree of bunching in administrative data can be rationalized with a firm-specific labor supply elasticity less than 2.5; at 1% of profits lost from round-number bias of employers, the implied labor-supply elasticity is between .8 and 1.5, depending on the extent and shape of heterogeneity assumed.

This research suggests that bunching in the wage distribution may not be merely a curiosity. Spikes at arbitrary wages suggest a failure of labor-market arbitrage due to employer mis-optimization and market power. Given the prevalence of round numbers in the wage distribution, it suggests that market power may be ubiquitous in labor markets as well as product markets. Moreover, our evidence suggests that when there is market power, we can expect employers to exhibit a variety of deviations from optimizing behavior, including adoption of heuristics such as paying round number wages.

References

Abowd, John M, Francis Kramarz, and David N Margolis. 1999. "High wage workers and high wage firms," *Econometrica*, 67(2): 251–333.

Akerlof, George A and Janet L Yellen. 1985. "A near-rational model of the business cycle, with wage and price inertia," *The Quarterly Journal of Economics*, 100(Supplement): 823–838.

Allen, Eric J, Patricia M Dechow, Devin G Pope, and George Wu. 2016. "Reference-dependent preferences: Evidence from marathon runners," *Management Science*.

Anderson, Eric T and Duncan I Simester. 2003. "Effects of $9 price endings on retail sales: Evidence from field experiments," *Quantitative marketing and Economics*, 1(1): 93–110.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M Shapiro. 2017. "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *The Quarterly Journal of Economics*: 1553–1592.

Babcock, Linda, William J Congdon, Lawrence F Katz, and Sendhil Mullainathan. 2012. "Notes on behavioral economics and labor market policy," *IZA Journal of Labor Policy*, 1(1), p. 2.

Backus, Matthew, Tom Blake, and Steven Tadelis. 2015. "Cheap talk, round numbers, and the economics of negotiation," NBER Working Paper No. 21285.

Bassier, Ihsaan, Arindrajit Dube, and Suresh Naidu. 2019. "Monopsony In Movers."

Basu, Kaushik. 1997. "Why are so many goods priced to end in nine? And why this practice hurts the producers," *Economics Letters*, 54(1): 41–44.

——— 2006. "Consumer cognition and pricing in the nines in oligopolistic markets," *Journal of Economics & Management Strategy*, 15(1): 125–141.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender identity and relative income within households," *The Quarterly Journal of Economics*, 130(2): 571–614.

Blomquist, Sören and Whitney Newey. 2017. "The bunching estimator cannot identify the taxable income elasticity,"Technical report, National Bureau of Economic Research.

Bloom, Nicholas and John Van Reenen. 2007. "Measuring and explaining management practices across firms and countries," *The Quarterly Journal of Economics*, 122(4): 1351–1408.

Caldwell, Sydnee and Emily Oehlsen. 2018. "Monopsony and the Gender Wage Gap: Experimental Evidence from the Gig Economy," *Massachusetts Institute of Technology Working Paper*.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 36(S1): S13–S70.

Chen, Daniel L and John J Horton. 2016. "Are Online Labor Markets Spot Markets for Tasks? A Field Experiment on the Behavioral Response to Wage Cuts," *Information Systems Research*, 27(2): 403–423.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2017. "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*.

Chetty, Raj. 2012. "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply," *Econometrica*, 80(3): 969–1018.

Cho, Sungjin and John Rust. 2010. "The flat rental puzzle," *The Review of Economic Studies*, 77(2): 560–594.

Delaigle, A and Irène Gijbels. 2004. “Bootstrap bandwidth selection in kernel density estimation from a contaminated sample,” *Annals of the Institute of Statistical Mathematics*, 56(1): 19–47.

Della Vigna, Stefano and Devin Pope. 2016. “What motivates effort? Evidence and expert forecasts,” *The Review of Economic Studies*.

DellaVigna, Stefano and Matthew Gentzkow. 2017. “Uniform Pricing in US Retail Chains,”Technical report, Working Paper.

DellaVigna, Stefano, John A List, and Ulrike Malmendier. 2012. “Testing for altruism and social pressure in charitable giving,” *The quarterly journal of economics*, 127(1): 1–56.

Dittmar, Jeremiah and Suresh Naidu. 2016. “Contested Property: Fugitive Slaves in the Antebellum US South,”Technical report, Working paper.

Dube, Arindrajit, Laura M Giuliano, and Jonathan S Leonard. 2015. “Fairness and frictions: The impact of unequal raises on quit behavior.”

Dube, Arindrajit, Jeff Jacobs, Suresh Naidu, and Siddarth Suri. 2018. “Monopsony Power In Crowdsourcing Markets,” Unpublished manuscript.

Gabaix, Xavier. 2017. “Behavioral Inattention,” NBER Working Paper No. 21285.

Gabaix, Xavier and David Laibson. 2006. “Shrouded attributes, consumer myopia, and information suppression in competitive markets,” *The Quarterly Journal of Economics*, 121(2): 505–540.

Goldfarb, Avi and Mo Xiao. 2011. “Who thinks about the competition? Managerial ability and strategic entry in US local telephone markets,” *The American Economic Review*, 101(7): 3130–3161.

Hall, Robert E and Alan B Krueger. 2012. "Evidence on the incidence of wage posting, wage bargaining, and on-the-job search," *American Economic Journal: Macroeconomics*, 4(4): 56–67.

Heidhues, Paul and Botond Kőszegi. 2008. "Competition and price variation when consumers are loss averse," *The American Economic Review*, 98(4): 1245–1268.

——— 2018. "Behavioral Industrial Organization."

Hortacsu, Ali and Steven L Puller. 2008. "Understanding strategic bidding in multi-unit auctions: a case study of the Texas electricity spot market," *The RAND Journal of Economics*, 39(1): 86–114.

Horton, John J, David G Rand, and Richard J Zeckhauser. 2011. "The online laboratory: Conducting experiments in a real labor market," *Experimental Economics*, 14(3): 399–425.

Kingsley, Sara Constance, Mary L Gray, and Siddharth Suri. 2015. "Accounting for market frictions and power asymmetries in online labor markets," *Policy & Internet*, 7(4): 383–400.

Kleven, Henrik J and Mazhar Waseem. 2013. "Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan," *The Quarterly Journal of Economics*, 128(2): 669–723.

Kleven, Henrik Jacobsen. 2016. "Bunching," *Annual Review of Economics*, 8: 435–464.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar. 2017. "Who Profits from Patents? Rent-sharing at Innovative Firms," Mimeo.

Krueger, Alan B. and Orley Ashenfelter. 2017. "Theory and Evidence on Employer Collusion in the Franchise Sector," Unpublished manuscript.

Kuziemko, Ilyana, Michael I Norton, Emmanuel Saez, and Stefanie Stantcheva. 2015. "How elastic are preferences for redistribution? Evidence from randomized survey experiments," *The American Economic Review*, 105(4): 1478–1508.

Lacetera, Nicola, Devin G Pope, and Justin R Sydnor. 2012. "Heuristic thinking and limited attention in the car market," *The American Economic Review*, 102(5): 2206–2236.

Levy, Daniel, Dongwon Lee, Haipeng Chen, Robert J Kauffman, and Mark Bergen. 2011. "Price points and price rigidity," *Review of Economics and Statistics*, 93(4): 1417–1431.

Manning, Alan. 2011. "Imperfect competition in the labor market," *Handbook of labor economics*, 4: 973–1041.

Manning, Kenneth C and David E Sprott. 2009. "Price endings, left-digit effects, and choice," *Journal of Consumer Research*, 36(2): 328–335.

Matějka, Filip. 2015. "Rationally inattentive seller: Sales and discrete pricing," *The Review of Economic Studies*, 83(3): 1125–1155.

Matejka, Filip and Alisdair McKay. 2015. "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 105(1): 272–98.

Naidu, Suresh. 2010. "Recruitment restrictions and labor markets: Evidence from the postbellum US South," *Journal of Labor Economics*, 28(2): 413–445.

Naidu, Suresh, Yaw Nyarko, and Shing-Yi Wang. 2016. "Monopsony power in migrant labor markets: evidence from the United Arab Emirates," *Journal of Political Economy*, 124(6): 1735–1792.

Naidu, Suresh, Eric A Posner, and E Glen Weyl. 2018. "Antitrust Remedies for Labor Market Power," *Harvard Law Review*.

National Employment Law Project. "Fight for 15 : *FourYears*,62 Billion," http://www.nelp.org/content/uploads/Fight-for-15-Four-Years-62-Billion-in-Raises.pdf, Accessed: 2017-09-30.

Pope, Devin G, Jaren C Pope, and Justin R Sydnor. 2015. "Focal points and bargaining in housing markets," *Games and Economic Behavior*, 93: 89–107.

Rebitzer, James B and Lowell J Taylor. 1995. "Do Labor Markets Provide Enough Short-Hour Jobs? An Analysis of Work Hours and Work Incentives," *Economic Inquiry*, 33(2): 257–273.

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

Shapiro, Carl and Joseph E Stiglitz. 1984. "Equilibrium unemployment as a worker discipline device," *The American Economic Review*, 74(3): 433–444.

Shaw, Aaron D, John J Horton, and Daniel L Chen. 2011. "Designing incentives for inexpert human raters," in *Proceedings of the ACM 2011 conference on Computer supported cooperative work*: 275–284, ACM.

Snir, Avichai, Daniel Levy, Alex Gotler, and Haipeng Allan Chen. 2012. "Not all price endings are created equal: price points and asymmetric price rigidity."

Sokolova, Anna and Todd Sorensen. 2018. "Monopsony in Labor Markets: A Meta-Analysis."

Solow, Robert M. 1979. "Another possible source of wage stickiness," *Journal of macroeconomics*, 1(1): 79–82.

Starr, Evan P, Norman Bishara, and JJ Prescott. 2016. "Noncompetes in the US labor force."

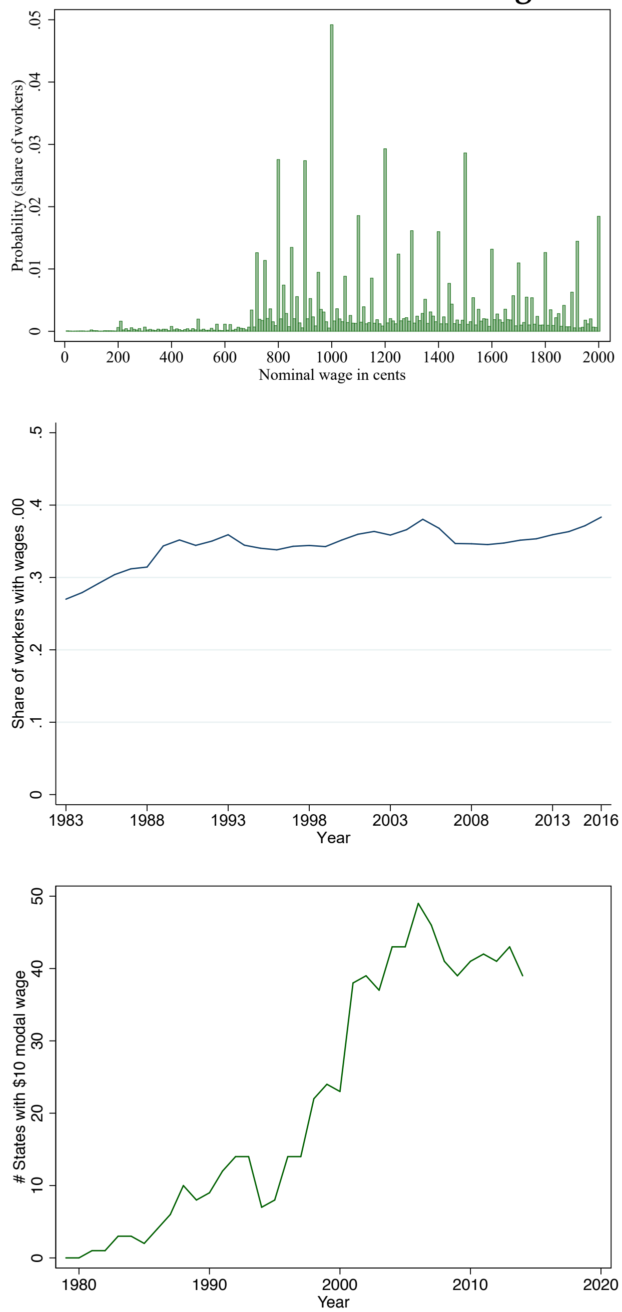
Stefanski, Leonard A and Raymond J Carroll. 1990. "Deconvolving kernel density estimators," *Statistics*, 21(2): 169–184.

Thomas, Manoj and Vicki Morwitz. 2005. “Penny wise and pound foolish: the left-digit effect in price cognition,” *Journal of Consumer Research*, 32(1): 54–64.

Wang, Xiao-Feng and Bin Wang. 2011. “Deconvolution Estimation in Measurement Error Models: The R Package Decon,” *Journal of Statistical Software*, 39(10), p. i10.

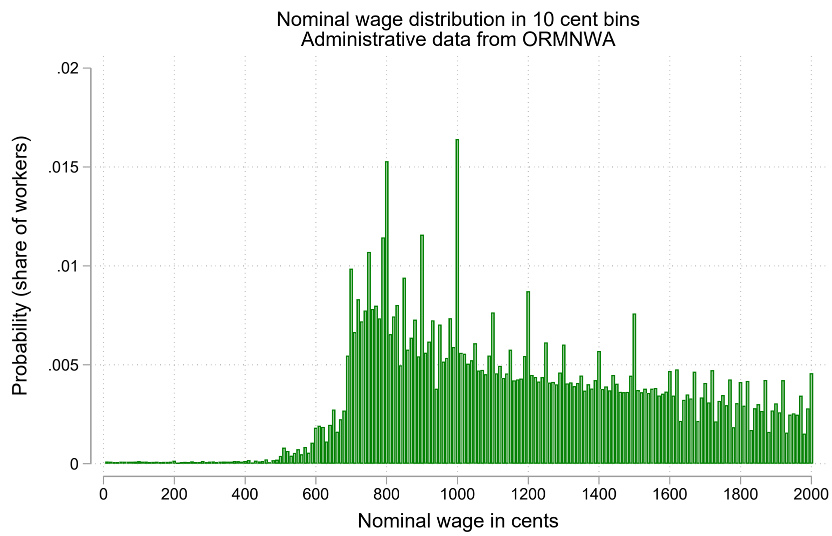
Webber, Douglas A. 2016. “Firm-Level Monopsony and the Gender Pay Gap,” *Industrial Relations: A Journal of Economy and Society*, 55(2): 323–345.

Figure 1: **Prevalence of Round Nominal Wages in the CPS**



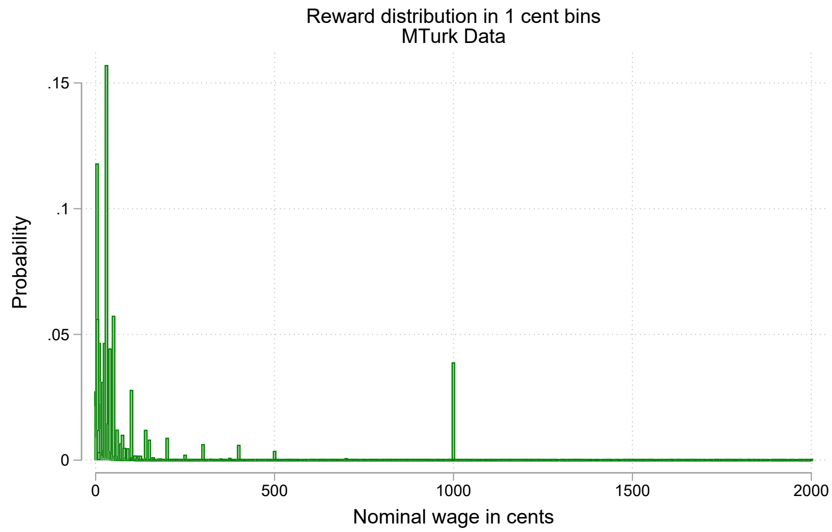
Notes. The top figure shows the CPS hourly nominal wage distribution, pooled between 2010 and 2016, in 10 cent bins. The middle figure shows the fraction of hourly wages in the CPS that end in .00 from 2003 through 2016. The bottom figure shows the fraction of states with $10.00 modal wages in the CPS. We exclude imputed wages.

Figure 2: **Histogram of Hourly Wages In Pooled Administrative Payroll Data from Minnesota, Oregon, and Washington, 2003-2007**



Notes. The figure shows a histogram of hourly wages in $0.10 (nominal) wage bins, averaged over 2003q1 to 2007q4, using pooled administrative Unemployment Insurance payroll records from the states of Oregon, Minnesota and Washington. Hourly wages are constructed by dividing quarterly earnings by the total number of hours worked in the quarter. The counts in each bin are normalized by dividing by total employment in that state, averaged over the sample period. The UI payroll records cover over 95% of all wage and salary civilian employment in the states. The counts here exclude NAICS 6241 and 814, home-health and household sectors, which were identified by the state data administrators as having substantial reporting errors.

Figure 3: **Bunching in Task Rewards in Online Labor Markets - MTurk**

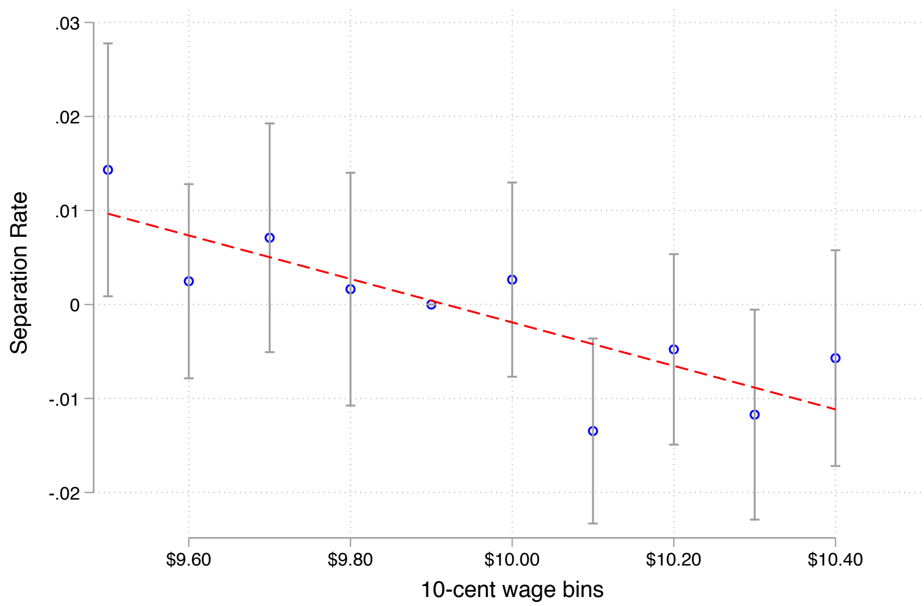


Notes. The figure shows a histogram of posted rewards by \$0.01 (nominal) bins scraped from MTurk. The sample represents all posted rewards on MTurk between May 01, 2014 and September 3, 2016.

Figure 4: **Distribution of Randomized Rewards in the MTurk Experiment, and Resulting Probability of Task Acceptance**

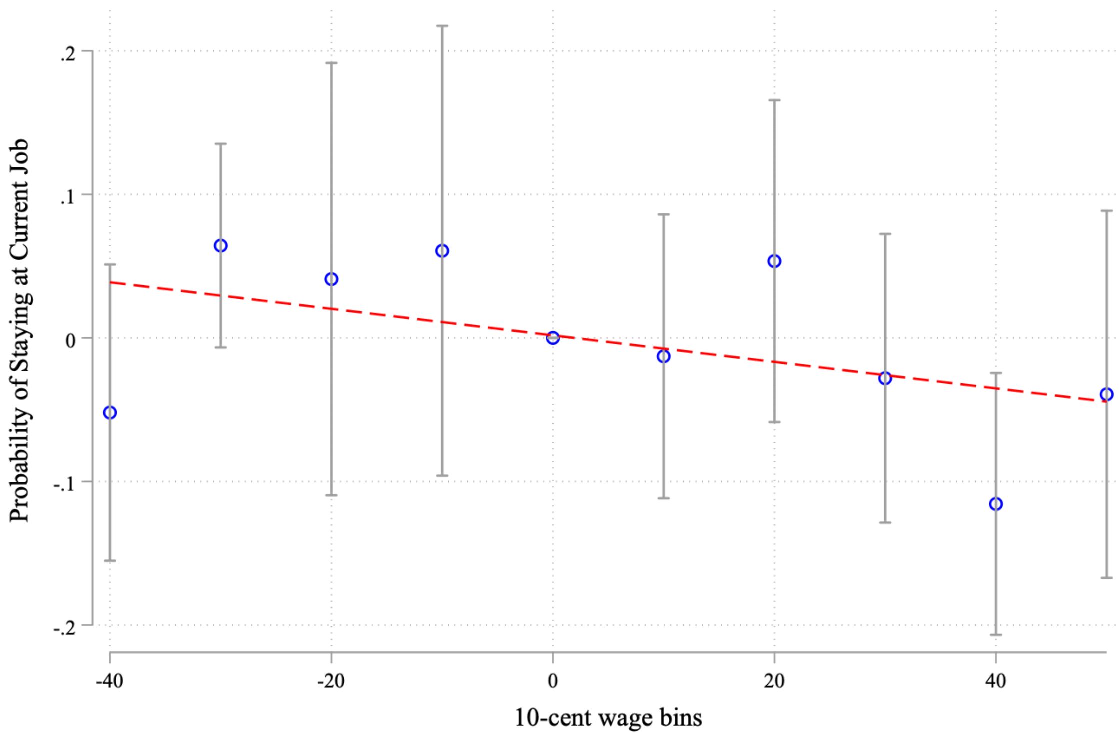
Notes. The figure shows the raw probability of accepting the bonus task as a function of the wage. Dots are scaled proportionally to the number of observations.

Figure 5: Oregon separation response by wage bin



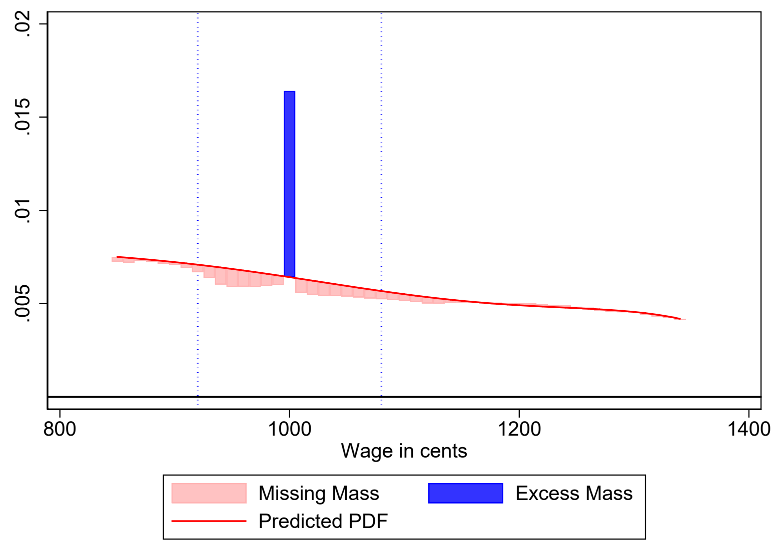
Notes: Estimated from quarterly matched worker-establishment data from Oregon 2003-2007. Separations with blue circles are coefficient on $0.10 wage bin dummies relative to $9.90 (the bin right below $10.00), and control for fully saturated interactions of current quarter, quarter of hire, starting wage bin ($0.10), 5-categories of weekly hours, and firm fixed effects. Robust standard errors are clustered at the 1-cent wage bin level.

Figure 6: Wal-Mart worker stated job mobility preference by offered wage bin



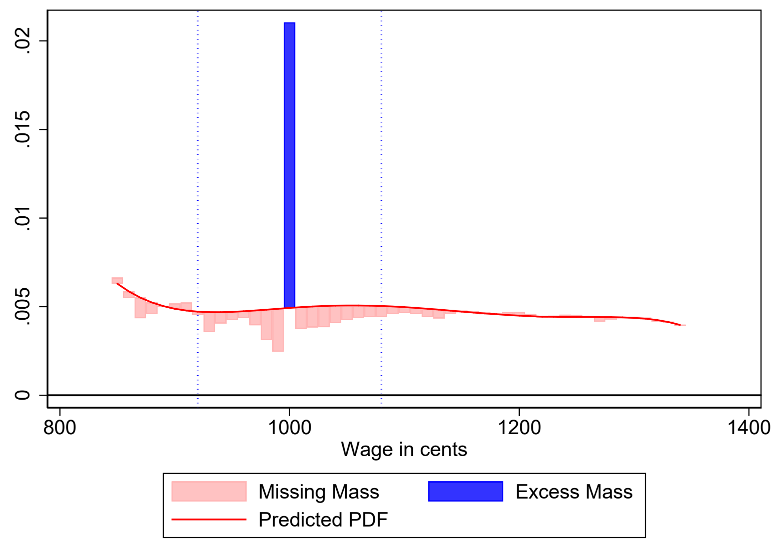
Notes: Estimated from responses of Wal-Mart workers in stated preference experiments. Stay probabilities with blue circles are coefficient on \$0.10 wage bin dummies relative to d_{it}= -\$0.10 (the bin right below round number), and control for respondent fixed effect and \$1 interval fixed effect. Robust standard errors are clustered at the respondent level..

Figure 7: **Excess Bunching and Missing Mass Around $10.00 Using Administrative Data on Hourly Wages (MN, OR, WA)**



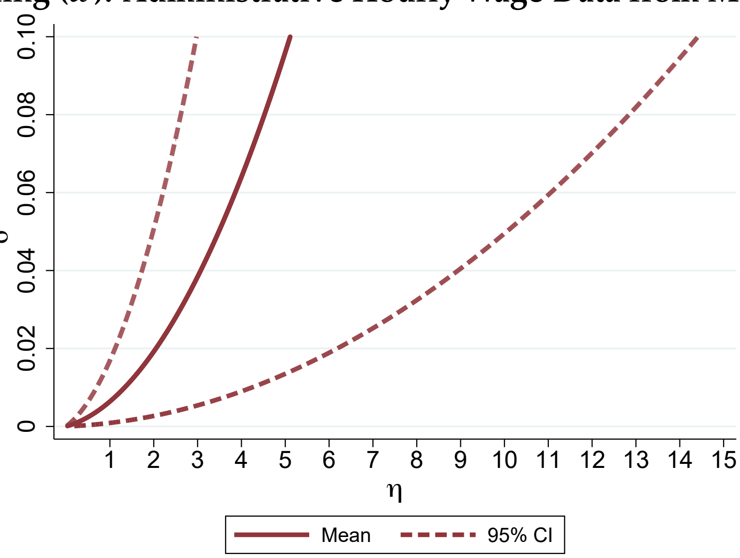
Notes. The reported estimates of excess bunching at $10.00, and missing mass in the interval around $10.00 as compared to the smoothed predicted probability density function, using administrative hourly wage counts from OR, MN and WA, aggregated by $0.10 bins, over the 2003q1-2007q4 period. The darker shaded blue bar at $10.00 represents the excess mass, while the lighter red shaded region represents the missing mass. The dotted lines represent the estimated interval from which the missing mass is drawn. The predicted PDF is estimated using a sixth order polynomial, with dummies for each $0.10 bin in the interval from which the missing mass is drawn. The width of the interval is chosen by iteratively expanding the interval until the missing and excess masses are equal, as described in the text.

Figure 8: **Excess Bunching and Missing Mass Around $10.00 Using Measurement Error Corrected CPS Data**



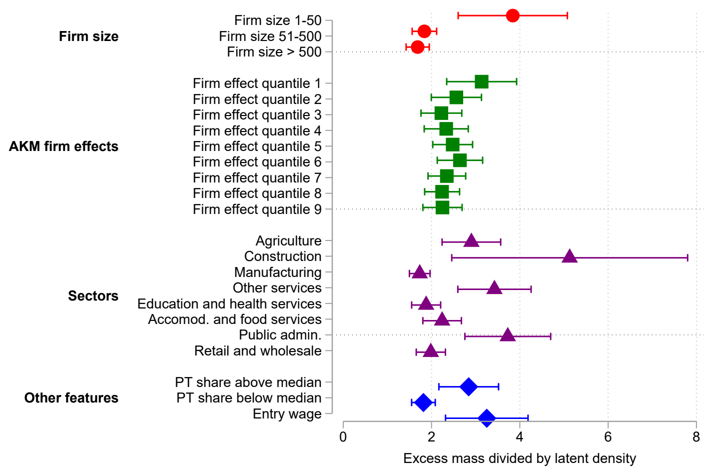
Notes. The reported estimates of excess bunching at $10.00, and missing mass in the interval around $10.00 as compared to the smoothed predicted probability density function, using CPS data corrected for measurement error using the 1977 administrative supplement. The darker shaded blue bar at $10.00 represents the excess mass, while the lighter red shaded region represents the missing mass. The dotted lines represent the estimated interval from which the missing mass is drawn. The predicted PDF is estimated using a sixth order polynomial, with dummies for each $0.10 bin in the interval from which the missing mass is drawn. The width of the interval is chosen by iteratively expanding the interval until the missing and excess masses are equal, as described in the text.

Figure 9: **Relationship Between Labor Supply Elasticity (η) , Optimization Frictions (δ) and Size of Bunching (ω): Administrative Hourly Wage Data from MN, OR, and WA**



Notes. The solid, red, upward sloping line shows the locus of the labor supply elasticity η and optimization frictions $\delta^* = E[\delta|\delta > 0]$ consistent with the extent of bunching ω estimated using the administrative hourly wage data from MN, OR, and WA between 2003q1-2007q4, as described in equation 26 in the paper. The dashed lines are the 95 percent confidence intervals estimated using 500 bootstrap replicates.

Figure 10: Heterogeneity in bunching by Oregon firm characteristics



Notes: AKM firm effects refer to deciles of firm effects estimated via Abowd et al. (1999), PT refers to part-time workers (less than 20 hours).The omitted firm size category is over 500 workers; the omitted sector is Retail. 95 percent confidence intervals are based on standard errors clustered at the firm level.

Table 1: **Task Acceptance Probability by Offered Task Reward on MTurk**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Wage	0.068***	0.081**	0.094**		0.111***	0.137**	0.194***	
	(0.025)	(0.036)	(0.042)		(0.040)	(0.059)	(0.063)	
Jump at 10		-0.008				-0.017		
		(0.016)				(0.027)		
Spline			-0.066				-0.104	
			(0.157)				(0.261)	
Local				0.002				0.036
				(0.022)				(0.044)
Global				-0.005				-0.010
				(0.015)				(0.025)
η	0.083***	0.098**	0.114**		0.132***	0.162**	0.230***	
	(0.030)	(0.044)	(0.051)		(0.048)	(0.070)	(0.075)	
Sample	Pooled	Pooled	Pooled	Pooled	Sophist.	Sophist.	Sophist.	Sophist.
Sample Size	5017	5017	5017	5017	1618	1618	1618	1618

Notes. The reported estimates are logit regressions of task acceptance probabilties on log wages, controlling for number of images done in the task (6 or 12), age, gender, weekly hours worked on MTurk, country (India/US/other), reason for MTurk work, and an indicator for HIT accepted after pre-registered close date. Column 1 reports specification 1 that estimates the labor-supply elasticity, without a discontinuity. Column 2 estimates specification 2, which tests for a jump in the probability of acceptance at 10 cents. Column 3 estimates a knotted spline in log wages, with a knot at 10 cents, and reports the difference in elasticities above and below 10 cents. Column 4 estimates specification 4, including indicator variables for every wage and testing whether the different in acceptance probabilities between 10 and 9 cents is different from the average difference between 12 and 8 (local) or the average difference between 5 and 15 (global). Columns 5-8 repeat 1-4, but restrict the sample to "sophisticates": Turkers who respond that they work more than 10 hours a week and their primary motivation is money. Robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$

Table 2: Oregon Separation response around $10.00/hour jobs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log wage	-0.215**	-0.456*	-0.310***	-0.470**	-0.271**	-0.252	-0.286	
	(0.075)	(0.243)	(0.076)	(0.203)	(0.116)	(0.187)	(0.203)	
Jump at 10.00		0.016		0.010		-0.001		
		(0.017)		(0.013)		(0.007)		
Spline							0.001	
							(0.002)	
Local								0.002
								(0.007)
Global								(0.005)
								(0.005)
Obs	668831	668831	272735	272735	68760	68760	68760	91982
Baseline controls	Y	Y	Y	Y	Y	Y	Y	Y
Sector and firm size			Y	Y				
Firm FE					Y	Y	Y	Y

Notes. Baseline controls include saturated interactions between current quarter, quarter of hire, 0.10 starting wage bin, 6-part weekly lagged hours category. The outcome is all separations. Exact 10-dollar bin defined as [10.00, 10.09]. Robust standard errors are clustered at the 1-cent wage bin level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table 3: Wal-Mart worker job mobility response to offered wage around a round number

	(1)	(2)	(3)	(4)	(5)
Log wage	-0.886***	-0.982	-0.501	2.634	
	(0.156)	(0.657)	(0.997)	(2.225)	
Jump at Round Number			-0.026		
			(0.036)		
Spline				-10.930	
				(7.222)	
Local					0.001
					(0.068)
Global					(0.052)
					(0.052)
Obs	936	927	927	927	927
Wage Relative to Round Number	N	Y	Y	Y	Y

Notes. Estimated from responses of Wal-Mart workers in stated preference experiments. Regressions control for respondent fixed effect and $1 interval fixed effect. Robust standard errors are clustered at the respondent level.$

* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

Table 4: **Estimates for Excess Bunching, Missing Mass, and Interval around Threshold**

	(1)	(2)	(3)	(4)	(5)
Value of w_0	\$10.00	\$10.00	\$10.00	\$10.00	\$10.00
Excess mass at w_0	0.010	0.034	0.014	0.041	0.016
	(0.002)	(0.008)	(0.004)	(0.009)	(0.004)
Total missing mass	-0.013	-0.043	-0.016	-0.033	-0.019
	(0.009)	(0.034)	(0.017)	(0.047)	(0.022)
Missing mass below	-0.007	-0.024	-0.009	-0.019	-0.009
	(0.007)	(0.023)	(0.010)	(0.029)	(0.014)
Missing mass above	-0.006	-0.019	-0.007	-0.014	-0.010
	(0.008)	(0.029)	(0.013)	(0.036)	(0.013)
Test of equality of missing shares of latent $<$ and $>$ 0:					
t-statistic	-0.338	-0.097	-0.065	-0.151	0.049
Bunching = $\frac{Actual\ mass}{Latent\ density}$	2.555	6.547	4.172	8.394	4.269
	(0.343)	(6.476)	(1.772)	(7.405)	(2.178)
w_L	\$9.20	\$9.30	\$9.30	\$9.30	\$9.20
w_H	\$10.80	\$10.70	\$10.70	\$10.70	\$10.80
$\omega = \frac{(w\_H - w\_0)}{w\_0}$	0.080	0.070	0.070	0.070	0.080
	(0.026)	(0.033)	(0.034)	(0.033)	(0.036)
Data:	Admin OR & MN & WA	CPS-Raw OR & MN & WA	CPS-MEC OR & MN & WA	CPS-Raw	CPS-MEC

Notes. The table reports estimates of excess bunching at threshold w_0, missing mass in the interval around w_0 as compared to the smoothed predicted probability density function, and the interval (ω_L, ω_H) from which the missing mass is drawn. It also reports the t-statistic for the null hypothesis that the size of the missing mass to the left of w_0 is equal to the size of the missing mass to the right. The predicted PDF is estimated using a sixth order polynomial, with dummies for each bin in the interval from which the missing mass is drawn. The width of the interval is chosen by iteratively expanding the interval until the missing and excess masses are equal, as described in the text. In columns 1-3, estimates are shown for bunching at \$10.00 from pooled MN, OR, and WA using the administrative hourly wage counts, the raw CPS data, and measurement error corrected CPS (CPS-MEC) over the 2003q1-2007q4 period. In column 4, estimates are shown for all states using the raw CPS data. Bootstrap standard errors based on 500 draws are in parentheses.

Table 5: **Robustness of Estimates for Excess Bunching, Missing Mass, and Interval Around Threshold**

	Dum. for \$0.5 (1)	Dum. for \$0.25 & \$0.5 (2)	Poly. of degree 4 (3)	Poly. of degree 7 (4)	Fourier, degree 3 (5)	Fourier, degree 6 (6)	Real wage poly. (7)
Value of w_0	\$10.00	\$10.00	\$10.00	\$10.00	\$10.00	\$10.00	\$10.00
Excess mass at w_0	0.010	0.010	0.010	0.010	0.010	0.009	0.010
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Total missing mass	-0.010	-0.011	-0.012	-0.013	-0.009	-0.017	-0.009
	(0.005)	(0.005)	(0.007)	(0.011)	(0.004)	(0.006)	(0.003)
Missing mass below	-0.007	-0.007	-0.006	-0.006	-0.007	-0.007	-0.004
	(0.005)	(0.005)	(0.005)	(0.007)	(0.004)	(0.005)	(0.003)
Missing mass above	-0.004	-0.004	-0.006	-0.006	-0.002	-0.009	-0.005
	(0.004)	(0.004)	(0.004)	(0.008)	(0.004)	(0.004)	(0.003)
Test of equality of missing shares of latent $<$ and $>$ w_0:							
t-statistic	-0.325	-0.341	-0.812	-0.595	-0.875	-0.830	-1.216
Bunching = $\frac{Actual\ mass}{Latent\ density}$	2.658	2.625	2.594	2.566	2.643	2.233	2.664
	(0.297)	(0.298)	(0.293)	(0.342)	(0.238)	(0.285)	(0.238)
w_L	\$9.40	\$9.40	\$9.20	\$9.20	\$9.30	\$9.40	\$9.20
w_H	\$10.60	\$10.60	\$10.80	\$10.80	\$10.70	\$10.60	\$10.80
$\omega = \frac{(w\_H w\_0)}{w\_0}$	0.060	0.060	0.080	0.080	0.070	0.060	0.080
	(0.023)	(0.023)	(0.025)	(0.026)	(0.038)	(0.025)	(0.026)

Notes. The table reports estimates of excess bunching at the threshold w_0 as compared to a smoothed predicted probability density function, and the interval (ω_L, ω_H) from which the missing mass is drawn. All columns use the pooled MN, OR, and WA administrative hourly wage data. The predicted PDF is estimated using a K-th order polynomial or values of K between 2 and 6 as indicated, with dummies for each bin in the interval from which the missing mass is drawn. The width of the interval is chosen by iteratively expanding the interval until the missing and excess masses are equal, as described in the text. Columns 1 and 2 include indicator variables for wages that are divisible by 50 cents and 25 cents, respectively. Columns 3 and 4 vary the order of the polynomial used to estimate the latent wage. Columns 5 and 6 represent the latent wage with a 3 and 6 degree Fourier polynomial, respectively. Column 7 estimates the predicted PDF using a sixth order polynomial of real wage bins, as opposed to the nominal ones. Bootstrap standard errors based on 500 draws are in parentheses.

Table 6: **Bounds for Labor Supply Elasticity in Administrative Data**

	(1)	(2)	(3)	(4)	(5)
A. δ^{*} = 0.01					
$\overline{\delta}$	0.001	0.004	0.002	0.003	0.002
η	1.337	1.581	1.581	1.581	1.337
90% CI	[0.417, 2.050]	[0.417, 4.525]	[0.472, 9.512]	[0.417, 4.525]	[0.417, 9.512]
95% CI	[0.417, 2.871]	[0.417, 4.525]	[0.417, 9.512]	[0.417, 4.525]	[0.417, 9.512]
B. δ^{*} = 0.05					
$\overline{\delta}$	0.005	0.020	0.011	0.017	0.010
η	3.484	4.045	4.045	4.045	3.484
90% CI	[1.291, 5.112]	[1.291, 10.692]	[1.429, 21.866]	[1.291, 10.692]	[1.291, 21.866]
95% CI	[1.291, 6.970]	[1.291, 10.692]	[1.291, 21.866]	[1.291, 10.692]	[1.291, 21.866]
C. δ^{*} = 0.1					
$\overline{\delta}$	0.010	0.041	0.022	0.034	0.019
η	5.112	5.908	5.908	5.908	5.112
90% CI	[1.983, 7.421]	[1.983, 15.319]	[2.182, 31.127]	[1.983, 15.319]	[1.983, 31.127]
95% CI	[1.983, 10.053]	[1.983, 15.319]	[1.983, 31.127]	[1.983, 15.319]	[1.983, 31.127]
$G(0)= \underline{G}$	0.896	0.592	0.785	0.662	0.807
Data:	Admin OR & MN & WA	CPS-Raw OR & MN & WA	CPS-MEC OR & MN & WA	CPS-Raw	CPS-MEC

Notes. The table reports point estimates and associated 95 percent confidence intervals for labor supply elasticities, η, and markdown values associated with different values of optimization friction δ. All columns use the pooled MN, OR, and WA administrative hourly wage data. In columns 1, 2 and 3, we use hypothesized values of δ of 0.01, 0.05 and 0.1 respectively. The labor supply elasticity, η, and the markdown are estimated using the estimated extent of bunching, ω, and the hypothesized δ, using equations 26 and 2 in the paper. The 95 percent confidence intervals in square brackets are estimated using 500 boostrap draws.

Table 7: **Bounds for Labor Supply Elasticity in Administrative Data — Robustness to Specifications of Latent Wage**

	Dum. for \$0.5 (1)	Dum. for \$0.25 & \$0.5 (2)	Poly. of degree 4 (3)	Poly. of degree 7 (4)	Fourier, degree 3 (5)	Fourier, degree 6 (6)	Real wage poly. (7)
A. $E(\delta \mid \delta > 0)$= 0.01							
$\bar{\delta}$	0.001	0.001	0.001	0.001	0.001	0.002	0.001
η	1.909	1.909	1.337	1.337	1.581	1.909	1.337
90% CI	[0.472, 2.050]	[0.472, 2.050]	[0.417, 2.050]	[0.417, 2.050]	[0.300, 2.871]	[0.538, 2.871]	[0.417, 2.050]
95% CI	[0.417, 2.050]	[0.417, 2.050]	[0.417, 2.050]	[0.417, 2.871]	[0.247, 2.871]	[0.417, 4.525]	[0.417, 2.871]
	0.006	0.006	0.005	0.005	0.005	0.008	0.004
B. $E(\delta \mid \delta > 0)$= 0.05	4.794	4.794	3.484	3.484	4.045	4.794	3.484
$\bar{\delta}$	[1.429, 5.112]	[1.429, 5.112]	[1.291, 5.112]	[1.291, 5.112]	[0.984, 6.970]	[1.593, 6.970]	[1.291, 5.112]
η	[1.291, 5.112]	[1.291, 5.112]	[1.291, 5.112]	[1.291, 6.970]	[0.839, 6.970]	[1.291, 10.692]	[1.291, 6.970]
90% CI							
95% CI	0.889	0.883	0.906	0.899	0.908	0.835	0.919

Notes. The table reports point estimates and associated 95 percent confidence intervals for labor supply elasticities, η, and markdown values associated with hypothesized δ = 0.01 and δ = 0.05. All columns use the pooled MN, OR and WA administrative hourly wage counts. The first two columns control for bunching at wage levels whose modulus with respect to \$1 is \$0.5, and \$0.5 or \$0.25, respectively. Column 3 uses a quadratic polynomial to estimate the wage distribution, whereas column 4 uses a quartic. In columns 5 and 6, instead of polynomials, Fourier transformations of degree 3 and 6 are employed. Column 7 estimates the predicted PDF using a sixth order polynomial of real wage bins, as opposed to the nominal ones. In row A, we hypothesize δ = 0.01; whereas it is δ = 0.05 in row B. The labor supply elasticity, η, and markdown values are estimated using the estimated extent of bunching, ω, and the hypothesized δ, using equation 26 and 2 in the paper. The 95 percent confidence intervals in square brackets are estimated using 500 boostrap draws.

Table 8: **Bounds for Labor Supply Elasticity in Offline Labor Market - Heterogeneous δ and η**

	Heterogeneous δ	Heterogeneous η	Heterogeneous δ & η, σ_δ = 0.1	Heterogeneous δ & η, σ_δ = 1
A. $E(\delta \| \delta > 0)$= 0.01				
$\bar{\delta}$	0.001	0.001	0.001	0.001
η	1.668	1.559	2.469	1.849
90% CI	[0.969, 4.394]	[0.917, 4.650]	[1.056, 5.446]	[0.758, 4.163]
95% CI	[0.845, 4.816]	[0.823, 5.328]	[0.905, 6.589]	[0.649, 5.062]
Markdown	0.375	0.391	0.288	0.351
90% CI	[0.185, 0.508]	[0.177, 0.522]	[0.155, 0.486]	[0.194, 0.569]
95% CI	[0.172, 0.542]	[0.158, 0.548]	[0.132, 0.525]	[0.165, 0.606]
B. $E(\delta \| \delta > 0)$= 0.05				
$\bar{\delta}$	0.006	0.006	0.006	0.006
η	4.244	3.991	6.036	4.616
90% CI	[2.629, 10.397]	[2.503, 10.965]	[2.808, 12.739]	[2.108, 9.833]
95% CI	[2.337, 11.346]	[2.284, 12.453]	[2.469, 15.445]	[1.837, 11.894]
Markdown	0.191	0.200	0.142	0.178
90% CI	[0.088, 0.276]	[0.084, 0.285]	[0.073, 0.263]	[0.092, 0.322]
95% CI	[0.081, 0.300]	[0.074, 0.304]	[0.061, 0.288]	[0.078, 0.352]
$G(0)=\underline{G}$	0.880	0.880	0.880	0.880
Data:	Admin OR & MN & WA	Admin OR & MN & WA	Admin OR & MN & WA	Admin OR & MN & WA

Notes. The table reports point estimates and associated 95 percent confidence intervals for labor supply elasticities, η, and markdown values associated with hypothesized δ=0.01 and δ=0.05. All columns use the pooled MN, OR, and WA administrative hourly wage counts. Heterogeneous δ and η are allowed in columns 1 and 2, using equations 27 and 28, respectively. Columns 3 and 4 allow heterogeneous δ and η, and assume a conditional lognormal distribution of δ, using a deconvolution estimator based on equation 29. The third column assumes a relatively concentrated distribution of δ (σ_δ = 0.1); whereas the fourth column assumes a rather dispersed distribution (σ_δ = 1). In row A, we hypothesize δ = 0.01; whereas it is δ = 0.05 in row B. The 90 and 95 percent confidence intervals in square brackets in columns 1 and 2 (3 and 4) are estimated using 500 (1000) boostrap draws.

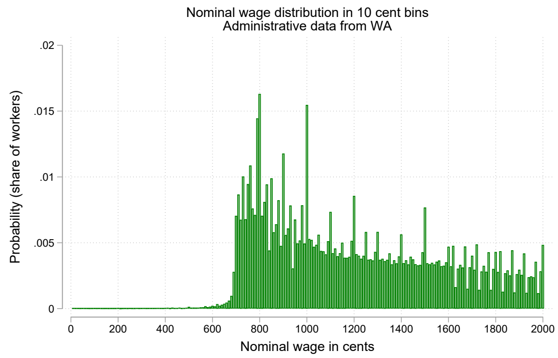
Online Appendix A

Additional Figures

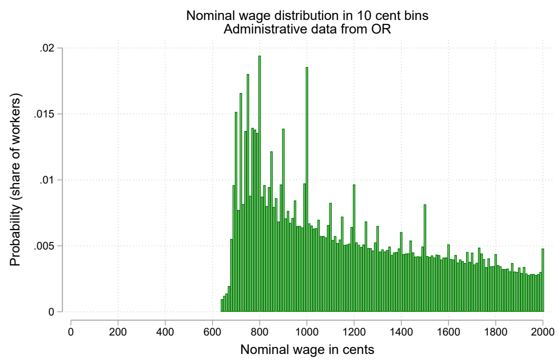
Appendix Figure A.1 plots the histograms of hourly wages in (nominal) $0.10 bins using administrative data separately for the states of Minnesota (panel A), Oregon (Panel B) and Washington (panel C). All are based on hourly wage data from UI records from 2003-2007. Hourly wages are constructed by dividing quarterly earnings by the total number of hours worked in the quarter. The counts are normalized by dividing by total employment in that state, averaged over the sample period. The figure shows very clear bunching at multiples of $1 in both states, especially at $10. Appendix Figure A.2 plots the overlaid histograms of hourly wages, pooled across MN, OR, and WA, in real $0.10 bins from 2003q4 and 2007q4, and shows that the nominal bunching at $10.00 occurs at different places in the real wage distribution in 2003 and 2007.

Figure A.1: **Histograms of Hourly Wages In Administrative Payroll Data from Minnesota, Oregon, and Washington, 2003-2007**

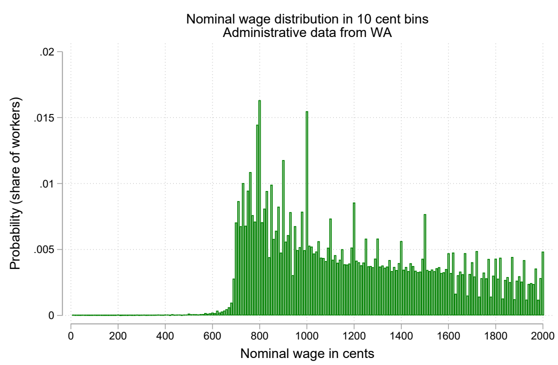
Panel A: Minnesota



Panel B: Oregon

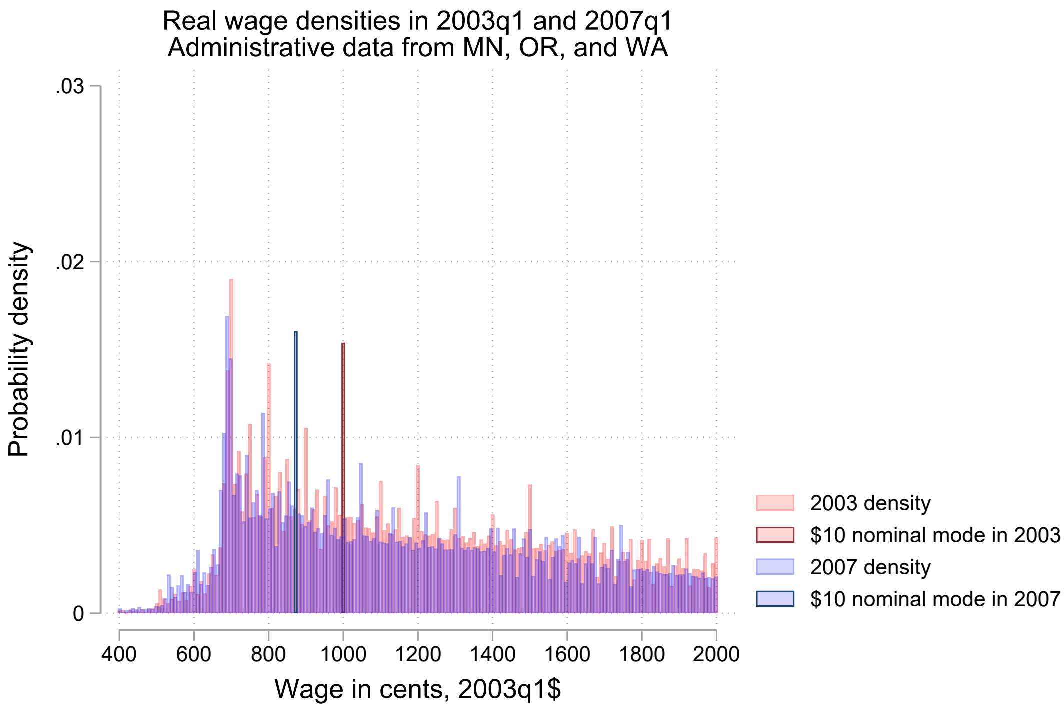


Panel C: Washington



Notes. The figure shows histograms of hourly wages in $0.10 (nominal) wage bins, averaged over 2003q1 to 2007q4, using administrative Unemployment Insurance payroll records from the states of Minnesota (Panel A) , Oregon (Panel B), and Washington (Panel C). Hourly wages are constructed by dividing quarterly earnings by the total number of hours worked in the quarter. The counts in each bin are normalized by dividing by total employment in that state, averaged over the sample period. The UI payroll records cover over 95% of all wage and salary civilian employment in the states. The counts here exclude NAICS 6241 and 814, home-health and household sectors, which were identified by the state data administrators as having substantial reporting errors.

Figure A.2: **Histograms of Real Hourly Wages In Administrative Payroll Data from Minnesota, Oregon, and Washington, 2003-2007**



Notes. The figure shows a histogram of hourly wages in $0.10 real wage bins (2003q1 dollars) for 2003q1 and 2007q1, using pooled administrative Unemployment Insurance payroll records from the states of Minnesota and Washington. The nominal $10 bin is outlined in dark for each year—highlighting the fact that this nominal mode is at substantially different part of the real wage distributions in these two periods. Hourly wages are constructed by dividing quarterly earnings by the total number of hours worked in the quarter. The counts in each bin are normalized by dividing by total employment in that state for that quarter. The UI payroll records cover over 95% of all wage and salary civilian employment in the states. The counts here exclude NAICS 6241 and 814, home-health and household sectors, which were identified by the state data administrators as having substantial reporting errors.

Online Appendix B

Allowing Flexible Heterogeneity in η and δ

In this Appendix, we provide details on the derivations of the robustness checks in section 6.1. We also show the estimated CDFs for the distributions of δ and η under the different distributional assumptions.

For the first exercise, we continue to assume a constant η but allow δ to have an arbitrary distribution $G(\delta|\delta > 0)$ while continuing to fix the probability that $\delta = 0$ at $\underline{G}$. In this case, for a given value of η the non-missing mass at ω would equal:

$$\phi\left(\omega\right) = 1 - \hat{G}(\eta(1+\eta)\frac{\omega^2}{2})$$

This expression implicitly defines a distribution $\hat{G}(\delta)$:

$$\hat{G}(\delta) = 1 - \phi\left(\sqrt{\frac{2\delta}{\eta(1+\eta)}}\right) \tag{27}$$

Note that this implies that $\delta \in [0, \delta_{max}]$ where $\delta_{max} = \frac{\omega^2}{2}\eta(1+\eta)$ where ω is the radius of the basin of attraction. We then fix $E(\delta|\delta > 0)$ at a particular value, similar to what we do with δ^*, and then can identify both an arbitrary shape of $\hat{G}(\delta)$ as well as η. Figure B.3 shows the distribution along with values of η from equation (27) in the MN-OR-WA administrative data. As can be seen, a higher η implies a first-order stochastic dominating distribution of δ; thus average δ is higher for higher η. This CDF also suggests our 2-point distribution is not too extreme an assumption: the non-zero δ are confined to about 20% of the distribution, and are bounded above by 0.11, suggesting that most firms are not foregoing more than 10% of profits in order to pay a round number.

A natural question is how our estimates could differ if, instead of a constant η and flexibly heterogeneous δ, we assume a heterogeneous η with an arbitrary distribution $H(\eta)$, along with some specified distribution $G(\delta)$. The simplest variant of this is to consider a two-point distribution (where δ is either 0 or δ^*) as in our baseline case above. In this variant of the model each firm is allowed to have its own labor supply elasticity, and each firm either mis-optimizes profits by a fixed fraction δ^* or not at all. In this case, solving for the positive value of η, the missing mass at ω should be equal to:

$$\phi(\omega) = [1 - \underline{G}]\, H\left(\frac{1}{2}\left(\sqrt{1 + \frac{8\delta^*}{\omega^2}} - 1\right)\right)$$

Since we can identify $\underline{G} = G(0) = 1 - \lim_{\omega \to 0^+} \hat{\phi}(\omega)$, for a particular δ^* we can empirically estimate the distribution of labor supply elasticities as follows:

$$\hat{H}(\eta) = \frac{\hat{\phi}\left(\sqrt{\left(\frac{8\delta^*}{(2\eta+1)^2 - 1}\right)}\right)}{1 - \underline{G}} \tag{28}$$

We can use $\hat{H}(\eta)$ to estimate the mean $E(\hat{\eta})$ for a given value of δ^*:

$$E(\hat{\eta}) = \int_0^\infty \eta d\hat{H}(\eta)$$

Note that under these assumptions, η is bounded from below at $\eta_{min} = \frac{1}{2}\sqrt{1 + \frac{8\delta^*}{\omega^2}} - 1$. In other words, the lower bound of η from the third method is equal to the constant estimate of η from our baseline, both of which come from the marginal bunching condition at the boundary of the interval ω. While we can only recover the distribution of η conditional on $\delta > 0$ (i.e. those that choose to bunch), we can make some additional observations about the parameters for non-bunchers. In particular, we can rule out the possibility that some of the the non-bunchers have $\delta > 0$ while being in a perfectly competitive labor market with $\eta = \infty$. This is because in our model those firms would be unable to attract workers from those firms with $\delta = 0$ and $\eta = \infty$. The gradual reduction in the missing mass $\phi(\omega)$ that occurs from moving away from $\omega = 0$ is entirely due to heterogeneity in $\eta's$. It rules out, for instance, that such a gradual reduction is generated by heterogeneity in $\delta's$ in contrast to the second method. As a result, the third method is likely to provide the largest estimates of the labor supply elasticity.

In parallel fashion to the previous case, we graphically show the implied distribution of η with a 2-point distribution for δ in Figure B.4. This figure shows the distribution of η implied by different values of δ from the MN-OR-WA administrative data. As can be seen, a higher η implies a first-order stochastic dominating distribution of η, thus average η is higher for higher δ.

Finally, we can extend this framework to allow for $G(\delta)$ to have a more flexible parametric form (with known parameters) than the 2-point distribution. We rely on recently developed methods in non-parametric deconvolution of densities to estimate the implicit $H(\eta)$. If we condition on $\delta > 0$,we can take logs of equation 20 (again maintaining that $\gamma = 0$) we get

$$2\ln(\omega) = ln(2) - \ln(\eta(1+\eta)) + \ln(\delta) = ln(2) - \ln(\eta(1+\eta)) + E[\ln(\delta)|\delta > 0] + \ln(\delta_{res}) \quad (29)$$

Here $\ln(\delta_{res}) \sim N(0, \sigma_\delta^2)$, and we fix $E[\ln(\delta)|\delta > 0] = \ln(E(\delta|\delta > 0)) + \frac{1}{2}\sigma_\delta^2$. We can use the fact that the cumulative distribution function of $2\ln(\omega)$ is given by $1 - \phi\left(\frac{1}{2}\exp(2\ln(\omega)\right))$ to numerically obtain a density for $2\ln(\omega)$. This then becomes a well-known deconvolution problem, as the density of $-\ln(\eta(1+\eta))$ is the deconvolution of the density of $2\ln(\omega)$ by the Normal density we have imposed on $\ln(\delta_{res})$. We can then recover the distribution of η,$H(\eta)$, from the estimated density of $-\ln(\eta(1+\eta)) + E[\ln(\delta)|\delta > 0]$.

We now illustrate how Fourier transforms recover the distribution $H(\eta)$. Consider the general case of when the observed signal (W) is the sum of the true signal (X) and noise (U). (In our case $W = 2\ln(\omega) - E[\ln(\delta)|\delta > 0]$ and $U = \ln(\delta_{res})$.)

$$W = X + U$$

Manipulation of characteristic functions implies that the density of W is $f_W(x) = (f_X * f_U)(x) = \int f_X(x-y) f_U(y) dy$ where $*$ is the convolution operator. Let W_j be the observed sample from W.

Taking the Fourier transform (denoted by $\sim$) , we get that $\tilde{f_W} = \int f_W(x)e^{itx}dx = \tilde{f_X} \times \tilde{f_U}$. To recover the distribution of X, in principle it is enough to take the inverse Fourier transform of $\frac{\tilde{f_W}}{\tilde{f_U}}$. This produces a "naive" estimator $\widehat{f_X} = \frac{1}{2\pi}\int e^{-itx}\frac{\sum_{j=1}^{N}\frac{e^{itWj}}{N}}{\phi(t)}dt$, but unfortunately this is not guaranteed to converge to a well-behaved density function. To obtain such a density, some smoothing is needed, suggesting the following deconvolution estimator:

$$\widehat{f_X} = \frac{1}{2\pi}\int e^{-itx}K(th)\frac{\sum_{j=1}^{N}\frac{e^{itW_j}}{N}}{\phi(t)}dt$$

where K is a suitably chosen kernel function (whose Fourier transform is bounded and compactly supported). The finite sample properties of this estimator depend on the choice of f_U. If $\tilde{f_U}$ decays quickly (exponentially) with t (e.g. U is normal), then convergence occurs much more slowly than if $\tilde{f_U}$ decays slowly (i.e. polynomially) with t (e.g. U is Laplacian). Note that once we recover the density for $X = \ln(\eta(1+\eta))$, we can easily recover the density for η.

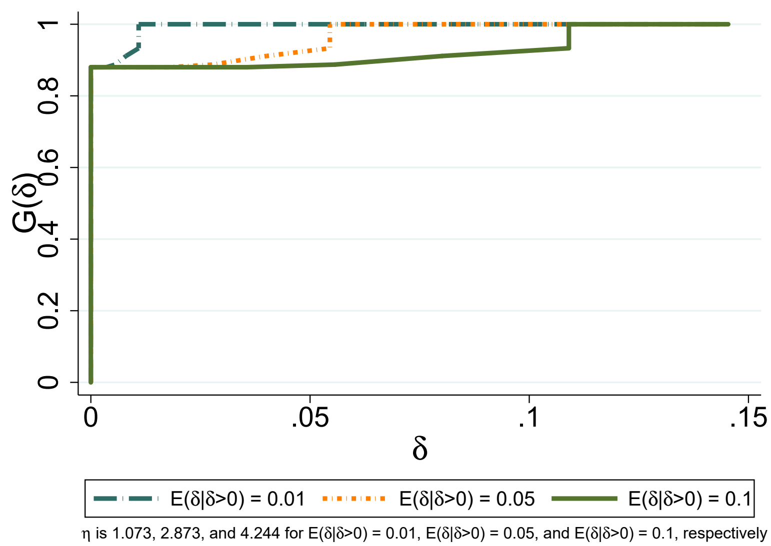
For normal $U = \ln(\delta_{res})$, Delaigle and Gijbels (2004) suggest a kernel of the form:

$$K(x) = 48\frac{cos(x)}{\pi x^4}(1-\frac{15}{x^2}) - 144\frac{sin(x)}{\pi x^5}(1-\frac{5}{x^2})$$

We use the Stefanski and Carroll (1990) deconvolution kernel estimator. This estimator also requires a choice of bandwidth which is a function of sample size. Delaigle and Gijbels (2004) suggest a bootstrap-based bandwidth that minimizes the mean-integral squared error, which is implemented by Wang and Wang (2011) in the R package `decon`, and we use that method here, taking the bandwidth that minimizes the mean-squared error over 1,000 bootstrap samples.

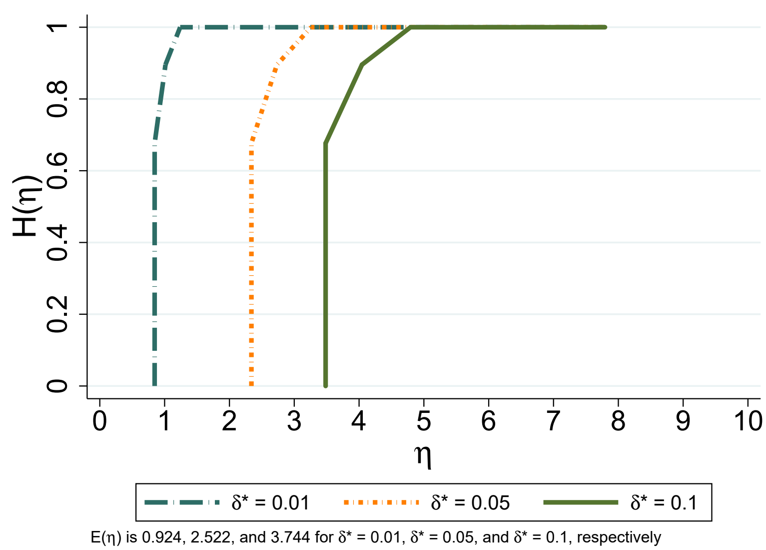
In Figure B.5, we show the distribution of η using the deconvolution estimator, assuming a lognormal distribution of δ. In the first panel, we estimate $H(\eta)$ assuming the standard deviation $\sigma_{ln(\delta)} = 0.1$, which is highly concentrated around the mean. In the second panel, we instead assume $\sigma_{ln(\delta)} = 1$. This is quite dispersed: among those with a non-zero optimization friction, δ around 16% have a value of δ exceeding 1, and around 31% have a value exceeding 0.5. As a result, we think the range between 0.1 and 1 to represent a plausible bound for the dispersion in δ. As before, we see a higher $E[\delta|\delta > 0]$ leads to first-order stochastic dominance of $H(\eta)$. For both cases with high- and low-dispersion of δ, the distribution $H(\eta)$ is fairly similar, though increase in $\sigma_{ln(\delta)}$ tends to shift $H(\eta)$ up somewhat, producing a smaller $E(\eta)$.

Figure B.3: **Implied Distribution of δ Under Constant η**



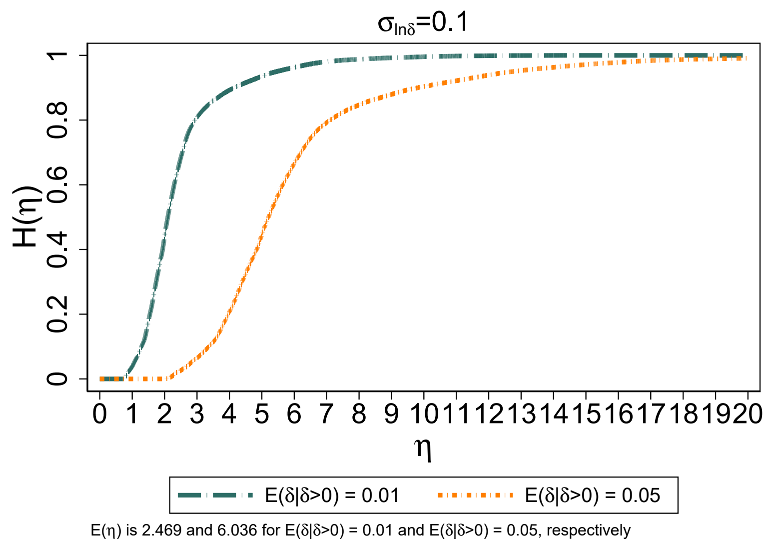
Notes. The figure plots the cumulative distributions $G(\delta)$ based on equation 27, for alternative values of $E(\delta|\delta > 0)$. The elasticity η is assumed to be a constant. The estimates use administrative hourly wage data from MN, OR, and WA.

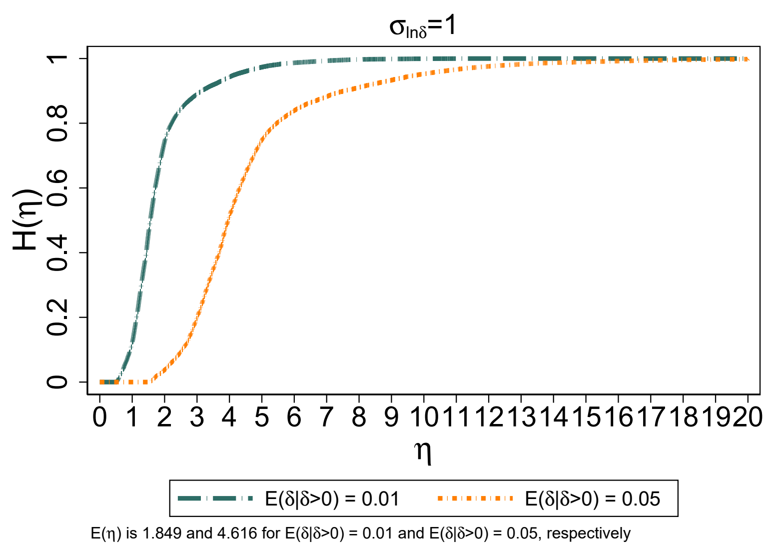
Figure B.4: **Implied Distribution of η with a 2-point Distribution of δ**



Notes. The figure plots the cumulative distributions $H(\eta)$ based on equation 28, for alternative values of $\delta^* = E(\delta|\delta > 0)$. δ is assumed to follow a 2-point distribution with $\delta = 0$ with probability $\underline{G}$ and $\delta = \delta^*$ with probability $1 - \underline{G}$. The estimates use administrative hourly wage data from MN, OR, and WA.

Figure B.5: **Implied Distribution of η using a Deconvolution Estimator where δ has a Conditional Lognormal Distribution**





Notes. The figure plots the cumulative distributions $H(\eta)$ using a deconvolution estimator based on equation 29, for alternative values of $E(\delta|\delta > 0)$. The procedure allows for an arbitrary smooth distribution of η, while assuming δ is lognormally distributed (conditional on being non-zero) with a standard deviation σ_δ. The top panel assumes a relatively concentrated distribution of δ with $\sigma_\delta = 0.1$; in contrast, the bottom panel assumes a rather dispersed distribution with $\sigma_\delta = 1$. The estimates use administrative hourly wage data from MN, OR, and WA.

Online Appendix C

Bunching in Hourly Wage Data from Current Population Survey and Supplement

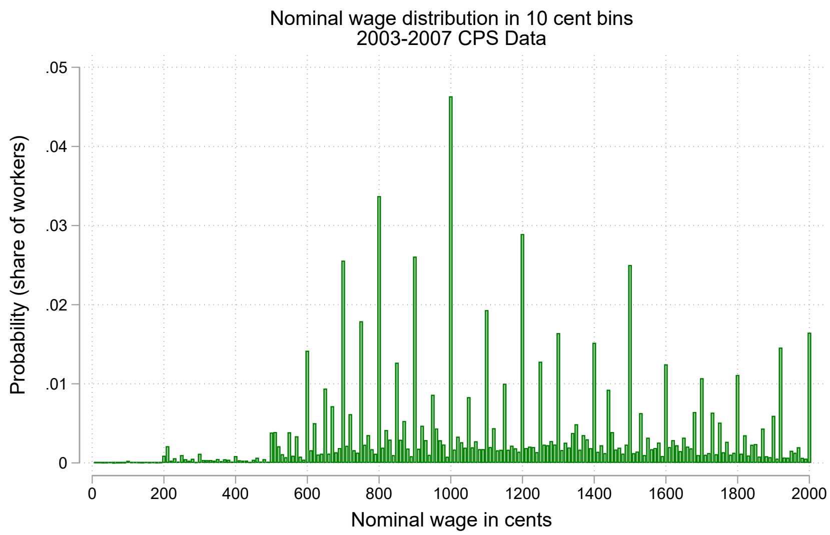
For comparison, we next show an analogous histogram of hourly nominal wage data using the national CPS data. In Figure C.6, we plot the nominal wage distribution in U.S. in 2003 to 2007 in $0.10 bins. There are notable spikes in the wage distribution at $10, $7.20 (the bin with the federal minimum wage), $12, $15, along with other whole numbers. At the same time, the spike at $10.00 is substantially larger than in the administrative data (exceeding 0.045), indicating rounding error in reporting may be a serious issue in using the CPS to accurately characterize the size of the bunching.

We also use a 1977 CPS supplement, which matches employer and employee reported hourly wages, to correct for possible reporting errors in the CPS data. We re-weight wages by the relative incidence of employer versus employee reporting, based on the two ending digits in cents (e.g., 01, 02, ... , 98, 99). As can be seen in Figure C.7, the measurement error correction produces some reduction in the extent of visible bunching, which nonetheless continues to be substantial. For comparison, the probability mass at $10.00 is around 0.02, which is closer to the mass in the administrative data than in the raw CPS. This is re-assuring as it suggests that a variety of ways of correcting for respondent rounding produce estimates suggesting a similar and substantial amount of bunching in the wage distribution.

Online Appendix C.1 Heterogeneous η by Worker Characteristics-CPS

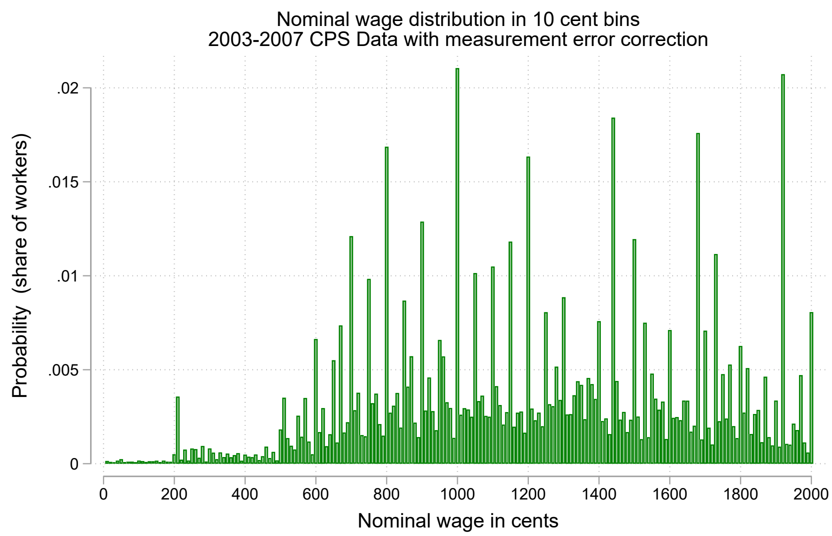
In Appendix Table C.1 we estimate the implied η for different δ^* under our baseline 2-point model across subgroups of the measurement corrected CPS data, as we do not have worker-level covariates for the administrative data. We examine young and old workers, as well as male and female separately. Consistent with other work suggesting that women are less mobile than men (Webber (2016); Manning (2011)), the estimated η for women is somewhat lower than that for men. We do not find any differences between older and younger workers. However, the extent of bunching is substantially larger for new hires consistent with bunching being a feature of initial wages posted, while workers with some degree of tenure are likelier to have heterogeneous raises that reduce the likelihood of being paid a round number. We find that among new hires the estimated η is somewhat higher than non-new hires. However, even for new hires—who arguably correspond most closely to the wage posting model—the implied η is only 1.58 if employers who are bunching are assumed to be losing 1% of profits from doing so, increasing to 4 when firms are allowed to lose up to 5% in profits.

Figure C.6: **Histogram of Hourly Wages in National CPS data, 2003-2007**



Notes. The figure shows a histogram of hourly wages by $0.10 (nominal) wage bins, averaged over 2003q1 and 2007q4, using CPS MORG files. Hourly wages are constructed by average weekly earnings by usual hours worked. The sample is restricted to those without imputed earnings. The counts here exclude NAICS 6241 and 814, home-health and household sectors. The histogram reports normalized counts in $0.10 (nominal) wage bins, averaged over 2003q1 and 2007q4. The counts in each bin are normalized by dividing by total employment, averaged over the sample period.

Figure C.7: **Wage Bunching in CPS data, 2003-2007, Corrected for Reporting Error Using 1977 CPS supplement**



Notes. The figure shows a histogram of hourly wages by $0.10 (nominal) wage bins, averaged over 2003q1 to 2007q4, using CPS MORG files, where individual observations were re-weighted to correct for overreporting of wages ending in particular two-digit cents using the 1977 CPS supplement. Hourly wages are constructed by dividing average weekly earnings by usual hours worked. The sample is restricted to those without imputed earnings. The counts here exclude NAICS 6241 and 814, home-health and household sectors. The histogram reports normalized counts in $0.10 (nominal) wage bins, averaged over 2003q1 and 2007q4. The counts in each bin are normalized by dividing by total employment, averaged over the sample period.

Table C.1: **Bounds for Labor Supply Elasticity in U.S. Labor Market — Heterogeneity by Demographic Groups**

	Male	Female	Age<30	Age≥30	Same job as last month	Different job from last month
Excess mass at w_0	0.018	0.015	0.030	0.012	0.014	0.028
	(0.003)	(0.004)	(0.006)	(0.003)	(0.003)	(0.006)
Total missing mass	-0.011	-0.012	-0.042	-0.012	-0.015	-0.023
	(0.009)	(0.007)	(0.013)	(0.006)	(0.007)	(0.012)
Bunching = $\frac{Actual\ mass}{Latent\ density}$	5.906	3.890	4.923	3.907	4.137	6.347
	(2.034)	(0.989)	(1.634)	(1.033)	(1.122)	(2.273)
A. $\delta^* = 0.01$						
$\overline{\delta}$	0.002	0.001	0.003	0.001	0.002	0.003
η	1.581	1.337	1.337	1.337	1.337	1.581
90% CI	[0.538, 4.525]	[0.618, 9.512]	[0.538, 4.525]	[0.618, 9.512]	[0.576, 9.512]	[0.538, 4.525]
95% CI	[0.472, 9.512]	[0.538, 9.512]	[0.472, 4.525]	[0.472, 9.512]	[0.472, 9.512]	[0.472, 4.525]
B. $\delta^* = 0.05$						
$\overline{\delta}$	0.009	0.005	0.014	0.007	0.008	0.013
η	4.045	3.484	3.484	3.484	3.484	4.045
90% CI	[1.593, 10.692]	[1.791, 21.866]	[1.593, 10.692]	[1.791, 21.866]	[1.687, 21.866]	[1.593, 10.692]
95% CI	[1.429, 21.866]	[1.593, 21.866]	[1.429, 10.692]	[1.429, 21.866]	[1.429, 21.866]	[1.429, 10.692]
$G(0)= \underline{G}$	0.820	0.895	0.713	0.863	0.834	0.750
Data:	CPS-MEC	CPS-MEC	CPS-MEC	CPS-MEC	CPS-MEC	CPS-MEC

Notes. The table reports point estimates and associated 95 percent confidence intervals for labor supply elasticities, η, and markdown values associated with hypothesized δ=0.01 and δ=0.05. All columns use the national measurement error corrected CPS data. The first two columns analyze by gender, the third and fourth by age, and the columns 5 and 6 by incumbency. In row A, we hypothesize $\delta = 0.01$; whereas it is $\delta = 0.05$ in row B. The labor supply elasticity, η, and markdown are estimated using the estimated extent of bunching, ω, and the hypothesized δ, using equations 26 and and 2 in the paper. The 95 percent confidence intervals in square brackets are estimated using 500 boostrap draws.

Online Appendix D

Testing Discontinuous Labor Supply on Amazon Mechanical Turk Observational Data

Our Amazon Mechanical Turk experiment focused on discontinuities at 10 cents, while our bunching estimator used the excess mass at $1.00. In this appendix we present evidence from observational data scraped from Amazon Mechanical Turk to show that there is also no evidence of a discontinuity in worker response to rewards at $1.00. Our primary source of data was collected by Panos Ipseiros between January 2014 and February 2016, and, in principle, kept track of all HITs posted in this period.

We keep the discussion of the data and estimation details brief, as interested readers can see details in Dube et al. (2018). Dube et al. (2018) combines a meta-analysis of experimental estimates of the elasticity of labor supply facing requesters on Amazon Mechanical Turk with Double-ML estimators applied to observational data.. That paper does not look at discontinuities in the labor supply at round numbers.

Following Dube et al. (2018) we use the observed duration of a batch posting as a measure of how attractive a given task is as a function of observed rewards and observed characteristics. We calculate the duration of the task as the difference between the first time it appears and the last time it appears, treating those that are present for the whole period as missing values. We convert the reward into cents. We are interested in the labor supply curve facing a requester. Unfortunately, we do not see individual Turkers in this data. Instead we calculate the time until the task disappears from our sample as a function of the wage. Tasks disappear once they are accepted. While tasks may disappear due to requesters canceling them rather than being filled, this is rare. Therefore, we take the time until the task disappears to be the duration of the posting—i.e., the time it takes for the task to be accepted by a Turker. The elasticity of this duration with respect to the wage will be equivalent to the elasticity of labor supply when offer arrival rates are constant and reservation wages have an exponential (constant hazard) distribution.

In order to handle unobserved heterogeneity, Dube et al. (2018) implement a double-machine-learning estimator proposed by Chernozhukov et al. (2017), which uses machine learning (we used random forests) to form predictions of log duration and log wage (using one half of the data), denoted $\ln(\widehat{duration_h})$ and $\ln(\widehat{wage_h})$,and then subtracts them from the actual variable values in the other sample, leaving residualized versions of both variables. The predictions use a large number of variables constructed from the metadata and textual descriptions of each task, and have high out-of-sample predictive power, and so the residuals are likely to reflect variation that, if not exogenous, are at least orthogonal to a very flexible and predictive function of all the other observable characteristics of a task. See Dube et al. (2018) for further details on implementation and estimation.

We then estimate regressions of the form:

$$\ln(duration_h) - \ln(\widehat{duration_h}) = \eta \times (\ln(wage_h) - \ln(\widehat{wage_h})) + \gamma \mathbf{1}_{\mathbf{w}>\mathbf{w_0}} + \epsilon$$

Results are shown in Table D.2. We restrict attention to windows of wages around our two most salient round numbers, 10 cents, where the window is 6 to 14 cents, and 1 dollar, where the window is \$0.80 to \$1.20. Across specifications, there is a clear negative relationship between wages/rewards and duration, with a coefficient on η similar in magnitude to the - 0.11 estimate obtained on the whole sample in Dube et al. (2018), and close to the experimental estimates reported there. We also show analogues of our experimental specifications from our pre-analysis plan. The first approach tests for a discontinuity by adding an indicator for rewards greater than or equal to 10 or 100 ("Jump at 10/100"). This level discontinuity is tested in specifications 3 and 4, and there is no evidence of log durations becoming discontinuously larger above either 10 cents or \$1.00. The second approach tests for a slope break at \$1.00 by estimating a knotted spline that allows the elasticity to vary between 6 and 9 cents, 9 and 10 cents, and then greater than 10 cents, or 81 and 95 cents, 95 cents and \$1.00, and then greater than \$1.00 up to \$1.20. The slope break specification is tested in specifications 5 and 6, where we report the change in slopes at 10 cents and \$1.00 ("Spline"). Again, there is no evidence of a change in the relationship between log duration and log reward between 9 and 10 cents, vs greater than 10 cents, or \$0.95 and \$1.00 versus greater than \$1.00.

Table D.2: **Duration of Task Posting by Log Reward and Jump at $1.00**

	(1)	(2)	(3)	(4)	(5)	(6)
Log Wage	-0.089***	-0.066***	-0.089***	-0.069***	-0.090***	-0.070***
	(0.024)	(0.014)	(0.024)	(0.015)	(0.025)	(0.015)
GEQ 10			0.014			
			(0.018)			
GEQ 100				0.027		
				(0.026)		
Spline 10					0.084	
					(0.225)	
Spline 100						0.693
						(0.700)
Double-ML	Y	Y	Y	Y	Y	Y
Window	6-14	80-120	6-14	80-120	6-14	80-120
Sample size	59,654	39,442	59,654	39,442	59,654	39,442

Notes. Sample is restricted to HIT batches with rewards between 51 and 149 cents. Columns 3, 4 and 8 estimate a specification testing for a discontinuity in the duration at $1.00, as in our pre-analysis plan, while columns 5 and 6 estimate the spline specification testing for a change in the slope of the log duration log reward relationship at $1.00, also from the pre-analysis plan. Significance levels are * 0.10, ** 0.05, *** 0.01.

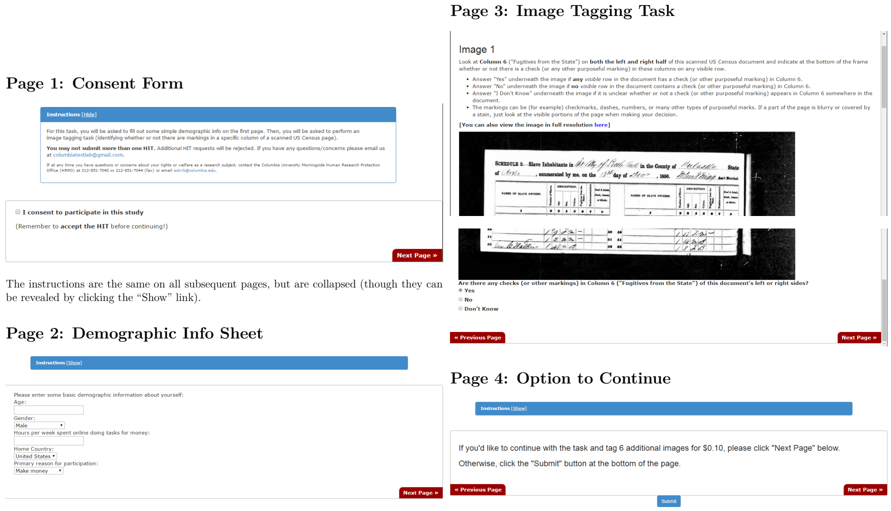
Online Appendix E

Additional Experimental Details and Specifications from Pre-analysis Plan

Figure E.1 shows screenshots from the experimental layout facing MTurk subjects. while Table E.3 shows specifications parallel to those from the main text, except with the number correct as the outcome, to measure responsiveness of subject effort to incentives. There is no evidence of any effect of higher rewards on the number of images labeled.

In Tables E.1 and E.2 we show specifications from our pre-analysis plan that parallel those in E.2 and E.3, respectively. These were linear probability specifications in the level of wages without any controls, instead of the logit specifications with log wages and controls we show in the main text. We also pool the two different task volumes. The initial focus of our experiment was to test for a discontinuity at 10 cents, which is unaffected by our changes in specification. While the elasticity is qualitatively very similar, the logit-log wage specification shown in the text is closer to our model, a variant of the model specified by Card et al. (2018), and improves precision on the elasticity estimate.

Figure E.1: **Online Labor Supply Experiment on MTurk**



Notes. The figure shows the screen shots for the consent form and tasks associated with the online labor supply experiment on MTurk.

Figure E.2: **Online Stated Preference Experiment for Wal-Mart Workers on Facebook**

	Current Job: Remodel Associate	Offered Job: Stocker, Backroom, & Receiving
Helpful Coworkers?	Often	Often
Work With Friends?	Some	Some
Supervisor Treats Everyone Fairly?	Sometimes	Often
Supervisor Treats You With Respect?	Sometimes	Often
Hours Per Week	20-40 hours	20-40 hours
Hourly Wage	$12.00	$9.00
Learn Transferable Skills?	No	Yes
Commute Time	15-30 minutes	0-15 minutes
Opportunities for Self-Expression?	Almost Always	Often
Control Over Your Hours?	No	No
Paid Time Off	1-10 days	1-10 days
Physically Demanding?	Yes	Yes

Imagine you are offered the job shown in the right column above (under "Offered Job"), which is compared to your current position at Walmart in the left column. Except for the highlighted characteristics, please assume the offered job is the same as your current position at Walmart, including on characteristics not listed in the table. You may scroll over the characteristics to see their definitions.

Please review the jobs and indicate below whether you would leave your current position at Walmart for the offered job, ask for a raise from your current position at Walmart, or stay at your current position at Walmart without asking for a raise.

Which action would you take?

- ○ Accept the offer and leave Walmart job
- ○ Ask for a raise at Walmart job
- ○ Stay at Walmart job without asking for a raise

Notes. The figure shows the screen shots for the stated preference experiment conducted on Facebook with Walmart. Items in the left column are characteristics of the worker's current job at Walmart, items in the right column are hypothetical characteristics of an offered job, with differences from current job highlighted in yellow.

Table E.1: **Preanalysis Specifications: Task Acceptance Probability by Offered Task Reward on MTurk**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Wage	0.004	0.008	0.004		0.001	0.003	-0.003		0.008*	0.013	0.011*	
	(0.003)	(0.006)	(0.004)		(0.004)	(0.008)	(0.005)		(0.004)	(0.009)	(0.006)	
Jump at 10			0.001				0.022				-0.021	
			(0.016)				(0.022)				(0.025)	
Spline		-0.002				0.193				-0.205		
		(0.156)				(0.206)				(0.236)		
Local				0.010				0.015				0.004
				(0.023)				(0.031)				(0.034)
Global				-0.000				0.011				-0.012
				(0.015)				(0.020)				(0.023)
η	0.052	0.090	0.050		0.015	0.035	-0.029		0.095*	0.157	0.140*	
	(0.033)	(0.071)	(0.048)		(0.042)	(0.087)	(0.062)		(0.051)	(0.116)	(0.073)	
Sample	Pooled	Pooled	Pooled	Pooled	6 HITs	6 HITs	6 HITs	6 HITs	12 HITs	12 HITs	12 HITs	12 HITs
Sample Size	5184	5184	5184	5184	2683	2683	2683	2683	2501	2501	2501	2501

Notes. The reported estimates are linear regressions of task acceptance probabilities on log wages, controlling for number of images. Column 1 reports specification 1 that estimates the labor-supply elasticity, without a discontinuity. Column 2 estimates specification 2, which tests for a jump in the probability of acceptance at 10 cents. Column 3 estimates a knotted spline in log wages, with a knot at 10 cents, and reports the difference in elasticities above and below 10 cents. Column 4 estimates specification 4, including indicator variables for every wage and testing whether the difference in acceptance probabilities between 10 and 9 cents is different from the average difference between 12 and 8 (local) or the average difference between 5 and 15 (global). Columns 5-8 repeat 1-4, but restrict the sample to "sophisticates": Turkers who respond that they work more than 10 hours a week and their primary motivation is money. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$

Table E.2: **Preanalysis Specifications: Task Correct Probability by Offered Task Reward on MTurk**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Wage	-0.001	0.001	-0.001		0.001	0.006*	0.001		-0.003**	-0.005	-0.003	
	(0.001)	(0.003)	(0.002)		(0.002)	(0.004)	(0.003)		(0.002)	(0.003)	(0.002)	
Jump at 10			-0.000				0.000				-0.002	
			(0.007)				(0.011)				(0.009)	
Spline		-0.012				-0.013				-0.008		
		(0.067)				(0.101)				(0.087)		
Local				0.003				-0.000				0.006
				(0.012)				(0.018)				(0.015)
Global				-0.003				-0.007				0.000
				(0.007)				(0.009)				(0.009)
η	-0.009	0.014	-0.008		0.008	0.060*	0.008		-0.029**	-0.047*	-0.026	
	(0.013)	(0.024)	(0.018)		(0.020)	(0.036)	(0.029)		(0.015)	(0.028)	(0.023)	
Sample	Pooled	Pooled	Pooled	Pooled	6 HITs	6 HITs	6 HITs	6 HITs	12 HITs	12 HITs	12 HITs	12 HITs
Sample Size	5184	5184	5184	5184	2683	2683	2683	2683	2501	2501	2501	2501

Notes. The reported estimates are linear regressions of task acceptance probabilities on log wages, controlling for number of images. Column 1 reports specification 1 that estimates the labor-supply elasticity, without a discontinuity. Column 2 estimates specification 2, which tests for a jump in the probability of acceptance at 10 cents. Column 3 estimates a knotted spline in log wages, with a knot at 10 cents, and reports the difference in elasticities above and below 10 cents. Column 4 estimates specification 4, including indicator variables for every wage and testing whether the difference in acceptance probabilities between 10 and 9 cents is different from the average difference between 12 and 8 (local) or the average difference between 5 and 15 (global). Columns 5-12 repeat 1-4, but restrict attention to subsamples based on the number of images given (randomized to be either 6 or 12). Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$

Table E.3: **Task Correct Probability by Offered Task Reward on MTurk**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Wage	-0.004	0.000	0.009		0.007	0.012	0.024	
	(0.009)	(0.014)	(0.014)		(0.017)	(0.025)	(0.027)	
Jump at 10		-0.002				-0.003		
		(0.006)				(0.010)		
Spline			-0.019				-0.025	
			(0.055)				(0.100)	
Local				-0.002				0.005
				(0.009)				(0.014)
Global				-0.003				-0.002
				(0.005)				(0.009)
η	-0.004	0.000	0.009		0.007	0.012	0.024	
	(0.009)	(0.014)	(0.015)		(0.017)	(0.026)	(0.027)	
Sample	Pooled	Pooled	Pooled	Pooled	Sophist.	Sophist.	Sophist.	Sophist.
Sample Size	3966	3966	3966	3925	1345	1345	1345	1334

Notes. The reported estimates are logit regressions of getting at least 1 out of 2 images correctly tagged on log wages (conditional on accepting the task), controlling for number of images done in the task (6 or 12), age, gender, weekly hours worked on MTurk, country (India/US/other), reason for MTurk, and an indicator for HIT accepted after pre-registered close date. Column 1 reports specification 1 that estimates the labor-supply elasticity, without a discontinuity. Column 2 estimates specification 2, which tests for a jump in the probability of acceptance at 10 cents. Column 3 estimates a knotted spline in log wages, with a knot at 10 cents, and reports the difference in elasticities above and below 10 cents. Column 4 estimates specification 4, including indicator variables for every wage and testing whether the different in acceptance probabilities between 10 and 9 cents is different from the average difference between 12 and 8 (local) or the average difference between 5 and 15 (global). Columns 5-12 repeat 1-4, but restrict attention to subsamples based on the number of images given (randomized to be either 6 or 12). Robust standard errors in parentheses.
* *$p < 0.10$, ** $p < 0.5$, *** $p < 0.01$*

Online Appendix F

Theoretical extension: An efficiency wage interpretation where effort depends on wage

In the main paper, we assume that the firm's ability to set wages comes from monopsony power. However, it may be recasted in terms of efficiency wages where wage affects productivity: there, too, the employer will set wages optimally such that the impact of a small change in wages around the optimum is approximately zero. In this section, we show a very similar logic applies in an efficiency wage model with identical observational implications as our monopsony model, with a re-interpretation of the labor supply elasticity η as capturing the rate at which the wage has to increase to ensure that the no-shirking condition holds when the firm wishes to hire more workers. Indeed, the observation that the costs of optimization errors are limited when wages are a choice variable was originally made by Akerlof and Yellen (1985) in the context of an efficiency wage model.

As in Shapiro and Stiglitz (1984), workers choose whether to work or shirk. Working entails an additional effort cost e. Following Rebitzer and Taylor (1995), we allow the detection of shirking, $D(l)$, to fall in the amount of employment $l(w)$.[26] Workers quit with an exogenous rate q. An unemployed worker receives benefit b and finds an offer at rate s. The discount rate is r. All wage offers are assumed to be worth accepting; once we characterize the wage setting mechanism, this implies a bound for the lowest productivity firm. Finally, generalizing both Rebitzer and Taylor (1995) and Shapiro and Stiglitz (1984),

[26] In Shapiro and Stiglitz (1984), the detection probability is exogenously set. This produces some predictions which are rather strong. For example, the model does not predict wages to vary with productivity, as the no shirking condition that pins down the optimal wage does not depend on firm productivity. The same is true for the Solow model, where the Solow condition is independent of firm productivity (see Solow 1979). As a result, those models cannot readily explain wage dispersion that is independent of skill distribution, which makes it less attractive to explain bunching. However, if we generalize the Shapiro-Stiglitz model to allow the detection probability to depend on the size of the workforce as in Rebitzer and Taylor (1995), this produces a link between productivity, firm size and wages. Going beyond Rebitzer and Taylor, we further generalize the model to allow for heterogeneity in firm productivity, which produces a non-degenerate equilbrium offer wage distribution.

we allow the wages offered by firms to vary; indeed our model will predict that higher productivity firms will pay higher wages—leading to equilibrium wage dispersion.

We can write the expected value of not shirking as:

$$V^N(w) = w - e + \frac{(1-q)V^N(w)}{1+r} + \frac{qV^U}{1+r}$$

The value of shirking can be written as:

$$V^S(w) = w + \frac{(1-q)(1-D)V^S(w)}{(1+r)} + \frac{(1-(1-q)(1-D))\,V^U}{(1+r)}$$

Finally, the value of being unemployed is:

$$V^U = b + \frac{sEV^N + (1-s)V^U}{(1+r)}$$

The (binding) no shirking condition, NSC, can be written as:

$$V^N(w) = V^S(w)$$

Plugging in the expressions above and simplifying we get the no-shirking condition:

$$w = \frac{r}{1+r}V^U + \frac{e(r+q)}{D(l)(1-q)}$$

We can further express V^U as a function of the expected value of an offer $V^N$and the probability of receiving an offer, s, as well as the unemployment benefit b. However, for our purposes, the key point is that this value is independent of the wage w and is taken to be exogenous by the firm in its wage setting. Since detection probability $D(l)$ is falling in l, we can now write:

$$D(l) = \frac{e\,(r+q)}{\left(w - e + \frac{1}{1+r}V^U\right)(1-q)}$$

This generates a relationship between l and w:

$$l(w) = D^{-1}\left(\frac{e\,(r+q)}{\left(w - e + \frac{1}{1+r}V^U\right)(1-q)}\right) = d\left(\frac{\left(w - e + \frac{1}{1+r}V^U\right)(1-q)}{e\,(r+q)}\right)$$

where $d(x) = D^{-1}(\frac{1}{x})$. Since $D'(x) < 0$, we have $d'(x) > 0$. This is analogous to the labor supply function facing the firm: to attract more workers who will work, one needs to pay a higher wage because detection is declining in employment, $D'(l) < 0$. Therefore, we can write the elasticity of the implicit labor supply function as:

$$\frac{l'(w)w}{l(w)} = \frac{d'(.)w}{d(.)} \times \frac{1-q}{e(r+q)}$$

.

If we assume a constant elasticity $d(x)$ function with elasticity ρ then the implicit "effective labor" supply elasticity is also constant:

$$\eta = \frac{l'(w)w}{l(w)} = \rho \times \frac{1-q}{e(r+q)}$$

The elasticity is falling in effort cost e, exogenous quit rate q, as well as the discount rate, r. It is also rising in the elasticity ρ, since a higher ρ means detection does not fall as rapidly with employment.

The implicit effective labor supply function is then:

$$l(w) = \frac{w^{\eta}}{C} = \frac{w^{\rho \times \frac{1-q}{e(r+q)}}}{C}$$

which is identical to the monopsony case analyzed in the main text. For a firm with productivity p_i, profit maximization implies setting the marginal cost of labor to the marginal revenue product of labor (p_i), i.e., $w_i = \frac{\eta}{1+\eta}p_i$. [27]

[27]We can also solve for $V^N = \frac{(E(w)-e)(1+r)}{r-b(1+r)} = \frac{\left(\frac{\eta}{1+\eta}E(p)-e\right)(1+r)}{r-b(1+r)}$. This implies we can write the equilibrium value of being unemployed as a function of the primitive parameters as follows: $V^U =$

Finally, we can augment this labor supply function to exhibit left-digit bias. Consider the case where for wage $w \geq w_0$, the wage is perceived to be to equal to $\tilde{w} = w + g$ while under w_0 it is perceived to be $\tilde{w} = w$. Now, the labor supply can be written as:

$l(w) = D^{-1}\left(\frac{e(r+q)}{\left(w-e+\frac{1}{1+r}V^U\right)(1-q)}\right) = d\left(\frac{\left(w-e+\frac{1}{1+r}V^U\right)(1-q)}{e(r+q)}\right)$ for $w < w_0$

$l(w) = D^{-1}\left(\frac{e(r+q)}{\left(w+g-e+\frac{1}{1+r}V^U\right)(1-q)}\right) = d\left(\frac{\left(w+g-e+\frac{1}{1+r}V^U\right)(1-q)}{e(r+q)}\right)$ for $w \geq w_0$

Note that under the condition that $d(x)$ has a constant elasticity, the implicit labor supply elasticity continues to be constant both below and above w_0. However, there is a discontinuous jump up in the $l(w)$ function at w_0. Therefore, we can always appropriately choose a γ such that this implicit labor supply function can be written as:

$$l(w_j, \gamma) = \frac{w^{\eta} \times \gamma^{\mathbb{1}_{w_j \geq w_0}}}{C} = \frac{w^{\rho \times \frac{1-q}{e(r+q)}} \times \gamma^{\mathbb{1}_{w_j \geq w_0}}}{C}$$

Facing this implicit labor supply condition, firms will optimize:

$$\Pi(p, w) = (p - w)l(w, \gamma) + D(p)\mathbf{1}_{w=w_0}$$

With a distribution of productivity, p, higher productivity firms will choose to pay more, as the marginal cost of labor implied by the implicit labor supply function is equated with the marginal revenue product of labor at a higher wage. Intuitively, higher productivity firms want to hire more workers. But since detection of shirking falls with size, this requires them to pay a higher wage to ensure that the no shirking condition holds. Similarly, all of the analysis of firm-side optimization frictions goes through here as well. A low η due to (say) high cost of effort now implies that a large amount of bunching at w_0 can be consistent with a small amount of optimization frictions, δ.

One consequence of this observational equivalence is that we cannot distinguish between efficiency wages and monopsony in our observational analysis. However, in our experimental analysis, we find that the evidence from online labor markets is more consis-

$(1+r)\left[\frac{b}{r+s} - \frac{e}{1-b(1+r} + \frac{\eta E(p)}{(1+\eta)(r-b(1+r)}\right]$

tent with a monopsony interpretation than an effort one, due to the absence of any effect of wages on the number of images tagged correctly. At the same time, it is useful to note that many of the implications from this efficiency wage model are quite similar to a monopsony one: for instance, both imply that minimum wages may increase employment in equilibrium, as Rebitzer and Taylor show. Therefore, while understanding the importance of specific channels is useful, the practical consequences may be less than what may appear at first blush.