Monopsony and Employer Mis-optimization Account for Round Number Bunching in the Wage Distribution

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

December 25, 2017

Abstract

We show that wages in administrative data and in online markets exhibit considerable bunching at round numbers that cannot all 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, and use this to derive bounds on employer market power. We provide experimental validation of these results from online labor markets, 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 both employer mis-optimization and wage setting power are important features of the labor market.

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

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

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

[§]We thank Doruk Cengiz, Jeff Jacobs, and Jon Zytnick for excellent research assistance. We are grateful for helpful comments from Matt Backus, Ellora Derenoncourt, Len Goff, David Green, Ilyana Kuziemko, Jim Rebitzer and from seminar participants at Microsoft Research, Columbia University, and Boston University.

1 Introduction

In the product market, prices are more frequently observed to end in 99 cents than can be explained by chance, and a literature has emerged to document and explain this (e.g Levy et al. 2011). This paper shows that there is similar bunching in the hourly wage distribution, though at "round" numbers. For example, in the Current Population Survey (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 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. This is remarkable given the considerable variation in the level and dispersion of wages across these states.[1] It seems highly unlikely that such bunching at $10.00 is present in the distribution of underlying marginal products of workers.

We use data from both administrative sources and an online labor market to show that there is true bunching of wages at round numbers, and it is not simply an artifact of survey reporting. To explain bunching, we provide an imperfectly competitive model with both workers' left-digit bias, and imperfect firm optimization in the form of employer preferences for round wages. The notion of optimization frictions on the part of employers draws on Chetty 2012. However, unlike the case typically considered in the literature where optimization errors prevent bunching, we consider the case where firms prefer paying round numbers despite potentially lower profits.[2] We show that, in general, excess bunching at round-numbered wages are a function of worker left-digit bias, the percent of

[1] The bottom panel in Figure 1 shows the number of states over time with $10.00 reported as the modal hourly wage. While means, medians, and variances of log wages vary greatly across states, a remarkably large number of number of states show a mode of exactly $10.00, reaching a peak of 48 states prior to the 2008 recession. The middle panel shows that the fraction of wages that end .00 is also a strikingly stable 30-40% of wages over the past 30 years.

[2] As in Chetty (2012), we deliberately abstract from the details of the employer optimization frictions, which may reflect administrative costs, inattention, limits on manager cognition, or norms that constrain wage setting behavior.

profits employers are willing to forgo to pay a round number wage, and the elasticity of labor supply facing the firm.[3] The existence of labor market power makes both worker and firm biases candidate explanations for bunching. Even if workers have left digit bias, in a perfectly competitive labor market with profit maximizing firms, wages will equal marginal products—whose distribution is unlikely to have spikes at round numbers. Moreover, as we show in this paper, lower monopsony power also implies a higher loss in profits from employers mistakenly paying round numbers.

Using both direct experimental as well as indirect evidence, we do not find evidence of worker left-digit bias. This is in contrast to the literature on price-bunching at 99 cents, where the most common hypothesis is left-digit bias— consumers think $9.99 is a much lower price than $10—in a market where producers have some pricing power (Basu 1997,Heidhues and Kőszegi 2008). Subsequently, we show how the degree of bunching can then be used to bound the extent of imperfect competition. Simply put, if—as we find—workers are not misperceiving wages, then a more competitive labor market restricts the space employers have to make mistakes in setting wages. We also consider other explanations, including round-numbered reference points in efficiency wage models and round-numbered focal points in bargaining models, and show that they are inconsistent with patterns in our data—particularly the experimental evidence from the online labor market.

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 two largest U.S. states (Minnesota and Washington) that collect information on hours. We compare the size of the bunches in the administrative data to those in the CPS, also use a unique CPS supplement

[3]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.

which matches respondents' wage information with those from the employers to correct for reporting error in the CPS. 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).

The two explanations—worker versus firm biases—have very different predictions about the the origin of the missing mass corresponding to bunching at the round number. Worker left-digit bias implies an asymmetry in the distribution of missing mass as employers who would otherwise pay a wage slightly below a salient round number have a stonger incentive to bunch than those above. In contrast, employer optimization frictions imply that jobs from both above and below the round number will offer the round number wage, implying a symmetry in the missing mass. Our estimates using administrative data do not indicate an asymmetry in the missing mass distribution, suggesting that left-digit bias is less important than employer mis-optimization as an explanation for bunching at round numbers. Next, we use the estimated extent of bunching along with our model to quantify the extent of optimization frictions and the labor supply elasticity facing firms using "economic standard errors" as in Chetty (2012). Any given quantity of bunching can be explained by a combination of how much profit falls as wages deviate from the firm's optimum, which is given by the extent of labor market competition, together with how much profit employers are will to give up to pay a round number. We estimate the former by bounding the latter. We conclude that if employers are assumed to not give up more than, say, 1% in profits by picking a round number wage, the implied competition in the labor market is quite low, with firm-specific labor supply elasticities of around 1; even allowing a 10% loss in profits, the implied labor supply elasticities are around 3.5. We show these results are robust to allowing very general forms of heterogeneity in both labor supply elasticities facing firms as well as heterogeneity in the extent of firm misoptimization.

As an added validation, 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 offline labor markets, the task reward distribution on MTurk exhibits considerable bunching. However, our experimentally estimated labor supply function shows no evidence of discontinuity as would be predicted by worker left-digit bias. Together with bunching evidence from offline and online observational data, the experimental evidence suggests that employer-side optimization frictions are the most plausible explanation for bunching. At the same time, our experimental (and observational) evidence imply that the labor supply facing online employers is highly inelastic, with elasticities around 0.1, consistent with other research on online labor markets. Together, these findings suggest very small optimization frictions for those who are bunching in the online labor market.

To summarize, this paper makes three main contributions—documenting the existence of bunching in wage distributions that cannot be explained as measurement error, providing evidence that employer optimization frictions rather than employee left-digit biases are a more likely cause, and providing estimates of the size of those frictions and employer market power both of which we find to be economically significant. We show there is a fundamental trade-off between the extent of employer market power and optimization frictions in rationalizing the extent of bunching. The more competitive the labor market, the higher the degree of misoptimization required to rationalize a given level of bunching. The intuition is that the penalty for a given deviation from the optimal wage is larger the more competitive is the labor market.

In observational data, we do not have enough information to separately identify the market power of employers and the size of the optimization frictions—though we show that at least one of them must be large. But in the MTurk data we have a separate experimental estimate of the market power of employers, and we use this together with the missing mass estimate to compute the size of optimization frictions. In the appendix, we replicate the experimental specification on non-experimental observational data, and find similarly low labor supply elasticities and little employee left-digit bias.

The plan of the paper is as follows. In section 2, we briefly review the literature on left-digit bias, bunching, and wage-setting power in the labor market. In section 3, 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 4, we develop a model of bunching that nests worker left-digit bias and firm optimization frictions as special cases. Section 5 recovers the degree of misoptimization and monopsony from the bunching estimates under a variety of assumptions about the degree of heterogeneity in both, and recovers labor supply elasticities consistent with alternative degrees of optimization frictions. Section 6 reports findings from the online experiment, combining them with bunching estimates from the observed online labor market to estimate the extent of optimization friction for employers in the online platform. Section 7 concludes.

2 Literature

A large literature has discussed cognitive biases in processing price information, but little of this has discussed applications to wage determination[4]. 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 higher

[4]Numerous other deviations from the standard model (e.g. concerns about fairness and time-inconsistency) have been documented in a wide variety of labor markets, see e.g. Babcock et al. (2012) for an overview. This suggests that it is not simply the case that workers are sophisticated when it comes to such a high-stakes price as their wage. Particularly relevant to our setting are Chen and Horton (2016) and Della Vigna and Pope (2016) who show a number of behavioral phenomena are present in Mechanical Turk workers, although neither examines left digit bias.

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 et al. (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.[5]

A large literature in behavioral industrial organization has explored how firms choose prices facing behavioral consumers (e.g.Gabaix and Laibson (2006). See Heidhues and Kőszegi (2018) for a survey), to explain these and other pricing anomalies. Theoretical models to explain bunches in prices (e.g. Basu 1997, 2006; Heidhues and Kőszegi 2008) 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) uses a Salop differentiated products model) and this assumption plays an important role in these models as it allows prices to deviate from marginal costs (which do not plausibly have bunches). Our paper is also related to a small but growing literature on behavioral firms (rather than consumers or workers), 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)).

In the models we develop of wage-bunching, it is also important to assume that firms have some labor market power. A recent literature has argued that, far from requiring explicit collusion (as in professional sports) or restrictive non-compete contracts (Starr,

[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 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.

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). One piece of evidence for this comes from significant rent-sharing elasticities and the importance of firm fixed-effects in explaining the distribution of wages (Card et al. 2016), and another piece of evidence comes from minimum wage effects on turnover and tenure (Dube, Lester and Reich (2010)). Direct evidence showing occupation-commuting zone concentration measures negatively affect wages is provided by Azar et al. (2017) A further piece of evidence comes from estimating the impacts of worker deaths on payroll, revenues, and worker substitution (Isen (2013) , Jäger (2016)). Finally, direct estimates of monopsony power from shocks to firm value-added that increase worker wages and employment, as with the patent grants used by Kline et al. (2017), also provide evidence of employer market power. We show that existence of wage-bunching at arbitrary numbers, together with auxiliary evidence we provide, can be used as further evidence in favor of monopsony.

While not the primary focus of this paper, we are also related to a recent literature on platform labor markets. Katz and Krueger (2016) document a large rise in "alternative" work arrangements in the U.S. between 2010 and 2015, including work on platforms such as Amazon Mechanical Turk. Our experimental evidence shows that left-digit biases by workers seem not to explain the pervasive bunching seen on this online platform. The same experimental evidence does show considerable employer market power, however, a fact we corroborate using a wide variety of estimates in our companion paper (Dube et al. (2017)). Calibrating our model with the experimental evidence, we further find that employers on Mechanical Turk seem to exhibit only a small degree of optimization friction, less than 1% of profits worth.

3 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.

3.1 Administrative hourly wage data from select 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 Island) have UI systems that collect detailed information on hours, allowing us to estimate hourly wages, and we have obtained data from the largest two (MN and WA). We have micro-aggregated hourly wage data from Unemployment Insurance payroll records for Minnesota and Washington between 2003q1 and 2007q4. 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 total number of 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 (we report the distributions separately in the Appendix). 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. The wages are clearly bunched at round numbers, with the modal wage at the \$10.00 bin representing more than 0.015 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 and WA minimum wages in this period, suggesting that the hourly wage measure is accurate.

3.2 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 3, 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 4, 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 produces estimates suggesting a similar and substantial amount of bunching in the wage distribution.

3.3 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 done 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 API developed by Ipeirotis to download the near universe of HITs from MTurk from May 2014 to February 2016, resulting in a sample of over 5 million tasks. 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). Descriptive statistics are in the Appendix and are described more fully there.

Figure 5 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 has almost no regulations, suggesting the analogous bunching in real world is not driven by unobserved institutional constraints, including long-term implicit or explicit contracts.

3.4 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 distribution. 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

[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.

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 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:

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

In this expression j sums over 10 cent wage bins (we use 1 cent bins in the MTurk data), and the $\sum_{i=0}^{K} \alpha_i w^i$ 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).

Since Δw is unknown, we use an iterative procedure similar to Kleven and Waseem (2013). Starting with $\Delta w = 10$, we estimate equation (1) 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 (1). We do this until we find a Δw such that the excess and missing masses are equalized. Since Δw is itself estimated, we estimates 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 (1) and add these back to the fitted $\hat{p_w}$ to form a new distribution $\tilde{p}_w$, and estimate regression (1) 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.[7]

[7] Following the literature, our procedure assumes that the missing mass is originating entirely from the

In Figure 6 we show the estimates for the administrative data from MN and WA, using polynomial order $K = 6$. For visual ease, we plot the kernel-smoothed $\hat{\beta}_j$ for the missing mass. 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 $\Delta w = \$0.80$, or $\omega = 0.08$. 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 1, 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; t-statistic for the null hypothesis that they are equal is 0.030. 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 0.023. 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 and WA only. We find a substantially larger estimate for the excess mass, around 0.032. In column 3, we report estimates using the re-weighted CPS counts for MN 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 0.016; 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-WA administrative data

surrounding basin. However, the extent to which estimates is due to the differences in samples (MN 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 2 is in Figure 7.

Since the counterfactual involves fitting a smooth distribution using a polynomial in the estimation range, in Table 2 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 order, K.

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.

4 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, 2006) and the optimization friction 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 et al. (2012) by allowing utility to depend on the true wage w as well as the leading digit.[8] 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 firm that pays wage w is given by:

$$l(w,p) = \frac{\left[we^{\gamma \mathbb{1}_{w \geq w_0}}\right]^{\eta}}{C} k\left(p\right) \tag{2}$$

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: the curve is plotted in figure 8 for specific parameter values. γ is the percentage increase in labor supply that comes from the left-digit bias of workers so the size of γ is a natural measure of the extent of left-digit bias. Our model of labor supply to individual firms can be micro-founded using a multinomial logit model—see Card et al. (2016) 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

[8] However we do not parameterize the extent of "left-digitness" as Lacetera et al. (2012) do. We are implicitly assuming "full inattention" to non-leading digits.

[9] Matejka and McKay (2015) provides 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.

as $\eta \to \infty$. Denote by $l^*(w,p) = \frac{w^\eta}{C} k\left(p\right)$ the "nominal" labor supply curve facing the firm, without any worker left-digit bias.

The other possible explanation for bunching that we consider is employer misoptimization. We now extend the model to allow employers to "benefit" by paying a round number, despite lowered profits.[10] While consistent with employers preferring to pay round numbers, it could reflect internal fairness 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 the exact profit-maximizing wage. These costs may be substantial, as evidenced by the pervasive use of round-numbers in publicly stated wage-policies of large firms.[11]

The presence of δ results in a profit function that looks like:

$$\pi(w,p) = (p-w)l(w,p)e^{\delta \mathbf{1}_{w=w_0}} \tag{3}$$

where δ is the percentage "gain" in profits from paying the round number.[12] 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 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 (2) and (3) profits from paying a wage w to a workers with marginal product p

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

[11]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).

[12]Matějka (2015) shows that rationally inattentive monopoly sellers will choose a discrete number of prices.

can be written as:

$$\pi(w,p) = (p-w)\frac{w^{\eta}}{C}e^{\eta\gamma\mathbb{1}_{w\geq w_0}}e^{\delta\mathbf{1}_{w=w_0}}h\left(p\right) = (p-w)l^{*}(w,p)e^{\eta\gamma\mathbb{1}_{w\geq w_0}}e^{\delta\mathbf{1}_{w=w_0}} \tag{4}$$

Define $\rho\left(w,p\right) = (p-w)l^{*}(w,p)$. Here $\rho\left(w,p\right)$ is, in the language of Chetty (2012), the "nominal model" that parameterizes profits in the absence of left-digit bias or optimization errors. 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 and firm 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:

$$w^{*}(p) = \frac{\eta p}{1+\eta} \tag{5}$$

i.e. 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 will pay the round number wage as opposed to the latent wage if:

$$\pi(w_0,p) > \pi(w^{*}(p),p) \tag{6}$$

which can be written as:

$$e^{\eta\gamma\mathbb{1}_{w*(p)<w_0}}e^{\delta} > \frac{\rho\left(w^{*}\left(p\right),p\right)}{\rho\left(w_0,p\right)} \tag{7}$$

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

$$\delta + \eta\gamma \mathbb{1}_{w*(p)<w_0} > \ln \rho \left(w * (p), p\right) - \ln \rho \left(w_0, p\right) \tag{8}$$

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 (8) can be approximated using the following second-order Taylor series expansion of $\rho\left(w_0, p\right)$ about $w^*\left(p\right)$[13]:

$$\ln \rho \left(w_0, p\right) \simeq \ln \rho \left(w^*, p\right) + \frac{\partial \ln \rho \left(w^*, p\right)}{\partial w}\left[w_0 - w^*\right] + \frac{1}{2}\frac{\partial^2 \ln \rho \left(w^*, p\right)}{\partial w^2}\left[w_0 - w^*\right]^2 \tag{9}$$

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:

$$\frac{\partial^2 \ln \rho \left(w, p\right)}{\partial w^2} = -\frac{1}{\left(p - w\right)^2} - \frac{\eta}{w^2} \tag{10}$$

Using (5) this can be written as:

$$\frac{\partial^2 \ln \rho \left(w^*, p\right)}{\partial w^2} = -\frac{\eta\left(1+\eta\right)}{w^{*2}} \tag{11}$$

where it is convenient to invert (5) and express in terms of the latent wage because wages are observed but marginal products are not. Substituting (11) into (9) and then into (8) leads to the following expression for whether a firm pays the round number:

[13] One can use the actual profit functions not the approximation, but the difference is small for the parameters we use, and the approximation has a clearer intuition.

$$\left[\frac{w_0 - w^*}{w^*}\right]^2 \equiv \omega^2 \leq \frac{\delta + \eta\gamma\mathbb{1}_{w*<w_0}}{\eta\left(1+\eta\right)} \tag{12}$$

The left-hand side (12) 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:

$$z_0 = \frac{\delta + \eta\gamma}{\eta\left(1+\eta\right)}, z_1 = \frac{\delta}{\eta\left(1+\eta\right)} \tag{13}$$

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 $G_0^z(z)$ and a CDF for z_1 of $G_1^z(z)$. From (13) it must be the case that $G_0^z(z) \leq G_1^z(z)$ with equality if there is no left-digit bias. The way in which we use this is the following—suppose the fraction of firms with a latent wage, w^*who bunch is denoted by $\phi(\omega) = \phi\left(\frac{w_0 - w^*}{w^*}\right)$, where ω is the proportionate gap between the optimal wage under the nominal model (w^*) and the round number w_0. Then (12) implies that we will have for $\omega < 0,$:

$$\phi(\omega) = 1 - G_0^z\left[\omega^2\right] \tag{14}$$

and for $w > w_0,$:

$$\phi(\omega) = 1 - G_1^z\left[\omega^2\right] \tag{15}$$

The left-hand side of (14) and (15) have been estimated in the earlier section on the origin of the missing mass. So, (14) and (15) imply that data on the source of the missing mass in the wage distribution can be used to identify, non-parametrically, the distributions of z_0 and z_1, G_0 and G_1. This does not allow us to identify the underlying distribution of (δ, γ, η), the underlying economics parameters of interest.

5 Recovering left-digit bias, monopsony, and 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 $G_0 = G_1$ 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 1 and so we proceed holding $\gamma = 0$.

Note that 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 $\phi(w)$, 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), and the proportional width of the basin on either side of the bunch is $\omega = \frac{\Delta w}{w_0}$, η and δ must satisfy:

$$\frac{\delta}{\eta\,(1+\eta)} = \omega^2 \tag{16}$$

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). [14]

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

[14] 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).

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 $\delta = 0$.

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:

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

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 3 respectively. The implied labor supply elasticity η varies between 0.846 and 3.484 when we vary δ^* between 0.01 and 0.1. Even assuming a substantial amount of misoptimization (around 10% of profits) suggests a labor supply elasticity facing a firm of less than 5; while the 90 percent confidence bounds rule out elasticities greater than 7.4. If we assume, instead, a 1% loss in profits due to optimization friction, the 90 percent confidence bounds rule out $\eta > 2.1$. 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 90 percent confidence interval around it. We can see visually that as we consider higher values of δ^*, the range of admissible $\eta's$ increase and become 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: Kline et al. (2017) estimate a labor-supply elasticity facing the firm of 2.7, using patent decisions as an instrument for firm productivity, 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 4. 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 4 explore changing the degree of the polynomial used to fit the main estimates in table 3, Column 5 uses a quadratic and Column 6 uses a quartic, and our results stay very similar to our main estimates in Table 3.

5.1 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 investigates whether that is the case.

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

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

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

$$\hat{G}(\delta) = 1 - \phi\left(\sqrt{\frac{\delta}{\eta(1+\eta)}}\right) \tag{17}$$

Note that this implies that $\delta \in [0, \delta_{max}]$ where $\delta_{max} = \omega^2\eta(1+\eta)$ where ω^2 is the width 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 asη. Figure 10 shows the distribution along with values of η from equation (17) in the MN-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.

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 the missing mass at ω should be equal to:

$$\phi(\omega) = [1 - \underline{G}]\, H\left(\frac{1}{2}\left(\sqrt{1 + \frac{4\delta^*}{\omega^2}} - 1\right)\right)$$

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:

$$\hat{H}(\eta) = \frac{\hat{\phi}\left(\sqrt{\left(\frac{4\delta^*}{(2\eta+1)^2-1}\right)}\right)}{1 - \underline{G}} \tag{18}$$

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

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

Note that under these assumptions, η is bounded from below at $\eta_{min} = \frac{1}{2}\sqrt{1 + \frac{4\delta^*}{\omega^2}} - 1$. 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 η conditional on $\delta > 0$ (i.e. the bunchers), note that we cannot explain the non-bunching mass by assuming the non-bunchers have $\delta > 0$ but $\eta = \infty$, as 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 11. This figure shows the distribution of η implied by different values of δ from the MN-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.[15]

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 14 (again maintaining that $\gamma = 0$) we get

$$2\ln(\omega) = -\ln(\eta(1+\eta) + \ln(\delta) = -\ln(\eta(1+\eta)) + E[\ln(\delta)\,|\delta > 0] + \ln(\delta_{res}) \tag{19}$$

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 $1 - \phi\left(\exp\{2\ln(\omega)\}\right)$ 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 $-\ln(\eta(1+\eta)) + E[\ln(\delta)\,|\delta > 0]$. Details on using Fourier transforms to recover the distribution $H(\eta)$ are in the Appendix. We use the Stefanski and Carroll (1990) deconvolution kernel estimator. We choose the bandwidth using a bootstrap procedure proposed by Delaigle and Gijbels (2004), taking the bandwidth

[15] This exercise is in the spirit of Saez (2010) who estimated 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).

that minimizes the mean-squared error over 1,000 bootstrap samples.

In Figure 12, 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 $E(\eta)$.

We quantitatively show robustness of our main estimates to alternate specifications in Table **5.** 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 1.143 instead of 0.846 in the baseline estimates from Table 3. 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 1.175 in this case is quite close to the estimate of 1.143 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 4 or greater. Under 5% loss in profits, we get elasticities in columns 1 and 2 that are just over 3, but still quite close to our baseline case of 2.339. Therefore, allowing for heterogeneity in either δ or η only modestly increases the estimated mean η as compared to our baseline estimates.

In columns 3 and 4 we report our estimates using the deconvolution estimator, allowing for an arbitrary distribution for η, along with a lognormal conditional distribution for δ. 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 1.323, which is close to the estimates in columns 1 and 2 (1.143 and 1.175) 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 $E(\eta)$ rises somewhat to 1.590. While the point estimate for this case is larger than the baseline estimate of 0.846 (column 1, Table 3), the 90% confidence interval contains the baseline estimate. Moreover, even in this case, the 95% confidence interval rules out $\eta's$ larger than 4.6, suggesting substantial monopsony power. With $E(\delta|\delta > 0) = 0.05$, we get $E[\eta] = 3.431$ and 4.029 under $\sigma_\delta = 0.1$ and $\sigma_\delta = 1$,respectively. 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 monopsony power remains robust to a wide range of assumptions made about the distribution of δ and η.

5.2 Heterogeneous effects by groups

In Table 6 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 (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 1.014 if employers who are bunching are assumed to be losing 1% of profits from doing so, increasing to 2.7 when firms are allowed to lose up to

5\% in profits.

6 Experimental evidence on nominal wage labor supply elasticity and left-digit bias

Observational data has the advantage that it relates to the labor market as a whole but the disadvantage that it does not offer direct estimates of the economic parameters of interest. This section reports an analysis of an online labor market which offers the advantage of being able to estimate parameters of interest directly, though the disadvantage 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 inverse labor supply elasticity facing the firm, is quite high, 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 et al. 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 et al. 2015 for complementary evidence of requester market power on MTurk) or indeed any online labor market. Indeed, in their original paper on labor economics on Amazon Turk, Horton et al. (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. While the previous section provided indirect evidence on left-digit bias as an explanation for observed bunching, we can take advantage of the Amazon MTurk labor market to run experiments.[16] We designed an experiment to test our model.[17] 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 et al. (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)). 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. Figure 13 shows the screens as seen by participants with (1) the consent form, (2) the initial screening questions and demographic information sheet, (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 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

[16] In a companion paper, Dube et al. (2017) compile labor supply elasticities implicit in the results from 9 previous crowdsourcing compensation experiments on MTurk and find they are uniformly small, generally below 1, and show a similarly low non-experimental labor supply elasticity (0.15) estimated using a double ML procedure on the scraped MTurk data.

[17] Pre-registered as AEA RCT ID AEARCTR-0001349

(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 by randomizing a wage offer for a HIT 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.[18]

6.1 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. We show the exact specifications from the pre-analysis plan in the Appendix.

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

$$Pr(Accept_i) = \beta_0 + \eta_1 log(w_i) + \beta_1 T_i + \beta_2 X_i + \epsilon_i \tag{20}$$

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

[18] 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 Appendix specifications omit this variable).

Appendix). 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 specification testing left-digit bias fits logit regressions allows for a jump in the labor supply at \$0.10, but constrains the slope to the the same on both sides:

$$Pr(Accept_i) = \beta_{0A} + \eta_{1A} log(w_i) + \gamma_{1A} \mathbb{1} \{w_i \geq 0.1\}_i + \beta_{1A} T_i + \beta_2 X_i + \epsilon_i \tag{21}$$

Here left-digit bias is rejected if $\gamma_{A2} = 0$. 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:

$$\begin{aligned} Pr(Accept_i) = \beta_{0B} + \eta_{1B} log(w_i) + \gamma_{2B} \times (log(w_i) - log(0.09)) \times \mathbb{1} \{w_i \geq 0.09\}_i \\ + \gamma_{3B} \times (log(w_i) - log(0.10)) \times \mathbb{1} \{w_i \geq 0.1\}_i + \beta_{2B} T_i + \beta_2 X_i + \epsilon_i \end{aligned} \tag{22}$$

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:

$$Pr(Accept_i) = \sum_{k \in S} \delta_k \mathbb{1} \{w_i = k\}_i + \gamma \beta_{3B} T + \beta_2 X_i + \epsilon_i \tag{23}$$

And then calculate the following statistics:

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

$$\beta_{local} = (\delta_{0.1} - \delta_{0.09}) - \frac{\left(\sum_{k=.08,w\neq .1,}^{.12} \delta_k - \delta_{k-0.01}\right)}{4}$$

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

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 $\frac{1}{10}(\beta_{0.15} - \beta_{0.05})$ 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 $Accept_i$.

6.2 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 14. The binned scatterplot in figure 14 shows the basic pattern of a shallow slope (in levels) with no discontinuity at 10. Table 7 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 specification 21 and tests for a jump at \$0.10 assuming common slopes above and below \$0.10. Column 3 corresponds to equation 22 and allows for slopes to vary on both sides of \$0.10. Finally, column 4, following the flexible specification in equation in 23, 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 close to zero in magnitude, and not statistically significant. We can rule out even small differences in elasticities 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 specification corresponds closely to the theoretical model, where we can recover γ by exponentiating the coefficient on the dummy for 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 zero.

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 table 8, 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, however, 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 cost 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 supply is fairly inelastically supplied to online employers, and an estimated elasticity η generally between 0.1 and 0.2.

In the Appendix, we present complementary evidence from the universe of MTurk jobs (N greater than 4,500,000). By estimating how long a job stays posted before being filled, as a function of the reward posted (and controlling for time posted, requester and task description fixed effects) we can recover another estimate of the labor supply curve facing an employer. The implied labor-supply elasticity under a constant job-filling hazard assumption is close to our experimental estimates (roughly .5) as well as those experimentally estimated in Horton et al. (2011), lending external validity to our experimental design. We also show that tasks with rewards that end in a round number are no more likely to be filled faster than those jobs with rewards that end in any other number, consistent with our experimental findings. Together, the observational and experimental evidence suggest that, at least on Amazon Turk, there is plenty of monopsony, and little left-digit bias.

In addition, we show in the Appendix that the round number bunching on Amazon Turk is not a function of experience: requesters that have posted many tasks or a cumulative large amount of reward money do not differ in their propensity to post round numbers, suggesting that the round numbers observed in the MTurk distribution are not driven by naive or inexperienced requesters.

6.3 Estimates of online optimization frictions

To quantify the extent of implied optimization frictions for MTurk requesters, we first estimate the extent of bunching using scraped reward data from MTurk, using the same methodology as Section 3 with a threshold w_0 = \$1.00. The results are reported in Table 9. Here we use 1 cent bins, estimating the regression between \$0.55 and \$1.55. Again, we find a very clear bunching; the width of the interval for the missing mass is wider here than in the offline labor market data, with $\omega = 0.17$ and a standard error of 0.06. For the online MTurk data, β_0 is again invariant to K at 0.027, while ω varies between 0.17 and 0.24 depending on K. Figure 15 shows the excess and missing mass along with the latent reward distribution in the MTurk data.

Since our estimates for γ were highly concentrated around 1, we impose $\gamma = 1$ which implies symmetric bunching, consistent also with our evidence of symmetry of missing mass above and below \$1.00 in Table 9. This implies we can use estimates for the extent of bunching ω (0.17) and the labor supply elasticity η (0.082) that allows us to recover an estimate for the optimization friction, δ, using equation 16.

This derivation is represented graphically in Figure 16. The solid and dashed lines in red show the $\eta - \delta$ loci consistent with the point estimate of ω and the associated 90 percent confidence interval. For a given value of bunching, ω, the locus is defined by equation 12 with $\gamma = 1$, which implies that a higher labor supply elasticity requires more optimization frictions to rationalize the bunching. Higher values of ω tilt the locus upward: for a given labor supply elasticity, a higher bunching implies greater optimization frictions. The black vertical lines represent the estimated labor supply elasticity and the associated 90 percent confidence intervals. The distribution of δ is derived from sampling on each of these estimates of ω and η. Inverting the point estimate of $\eta = 0.082$ produces an estimate of $\delta^* = 0.003$, well below the 1% threshold we imposed in the offline labor market analysis above.

These estimates are also reported in Table 9. Since both there is sampling error of estimating ω and η, we use bootstrapping (with 500 replicates) to derive the 90 percent confidence interval δ^*, which is estimated as (0.000, 0.007). Even though there is extensive bunching at $1.00 rewards, the small labor supply elasticity implies a small optimization error.

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 from observational data. 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.

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 misoptimization. We show that when there is sizable market power, it requires only 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

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.

Azar, José, Ioana Marinescu, and Marshall I Steinbaum. 2017. "Labor Market Concentration,"Technical report, National Bureau of Economic Research.

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.

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.

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.

Card, David, Ana Rute Cardoso, Jörg Heining, and Patrick Kline. 2016. "Firms and labor market inequality: Evidence and some theory," NBER Working Paper No. 22850.

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.

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, Jeff Jacobs, Suresh Naidu, and Siddarth Suri. 2017. "Monopsony Power In Crowdsourcing Markets," Unpublished manuscript.

Dube, Arindrajit, T William Lester, and Michael Reich. 2010. "Minimum wage effects across state borders: Estimates using contiguous counties," *The review of economics and statistics*, 92(4): 945–964.

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.

Isen, Adam. 2013. "Dying to Know: Are Workers Paid Their Marginal Product?", Unpublished manuscript.

Jäger, Simon. 2016. "How substitutable are workers? evidence from worker deaths."

Katz, Lawrence F and Alan B Krueger. 2016. "The rise and nature of alternative work arrangements in the United States, 1995-2015,"Technical report, National Bureau of Economic Research.

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.

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.”

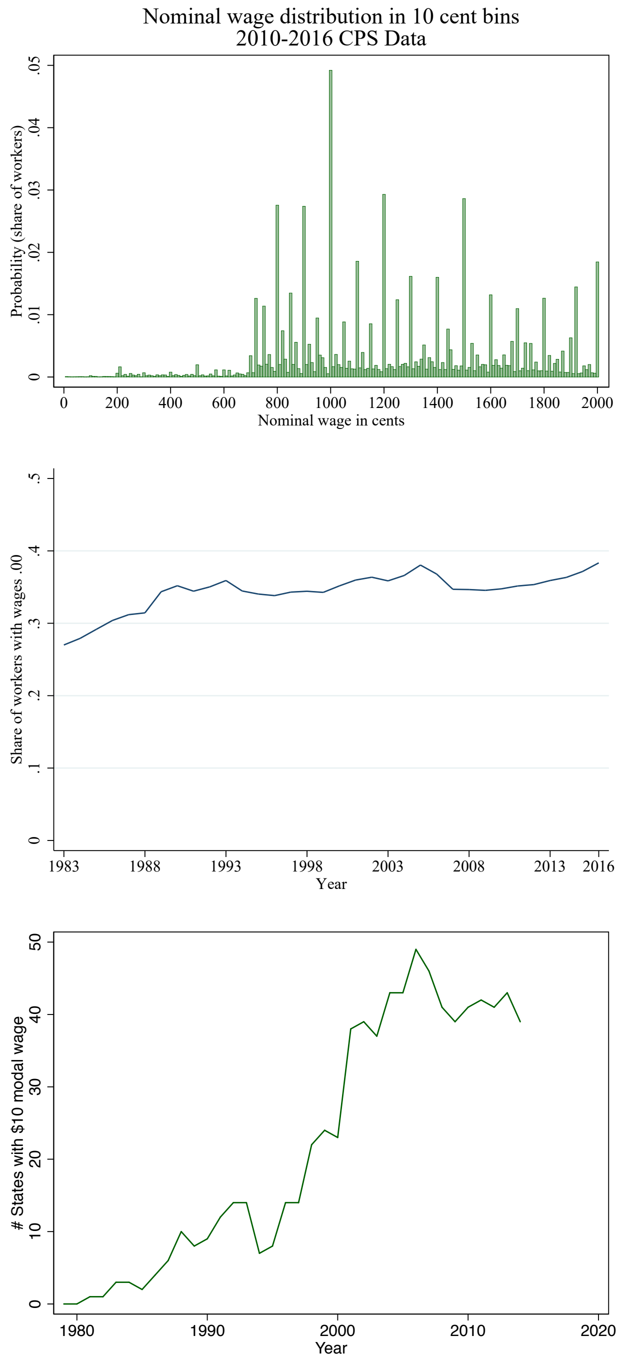
Starr, Evan P, Norman Bishara, and JJ Prescott. 2016. “Noncompetes in the US labor force.”

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.

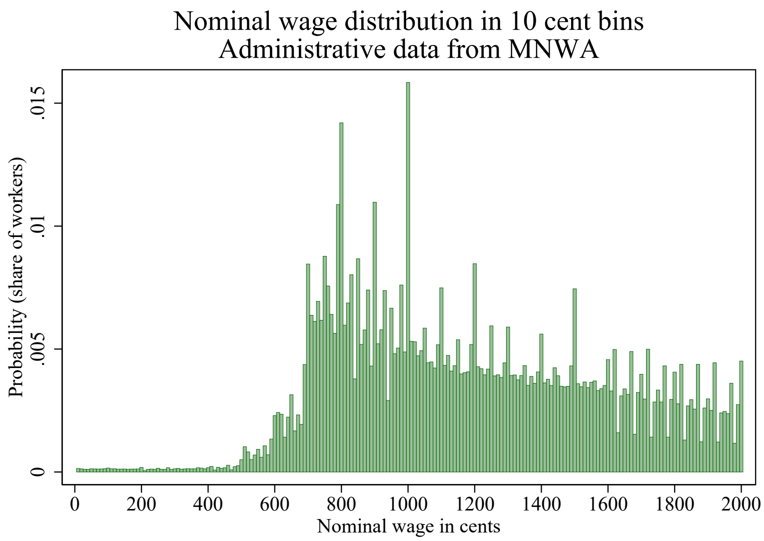
Wang, Xiao-Feng and Bin Wang. 2011. “Deconvolution Estimation in Measurement Error Models: The R Package Decon,” *Journal of Statistical Software*, 39(10), p. i10.

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



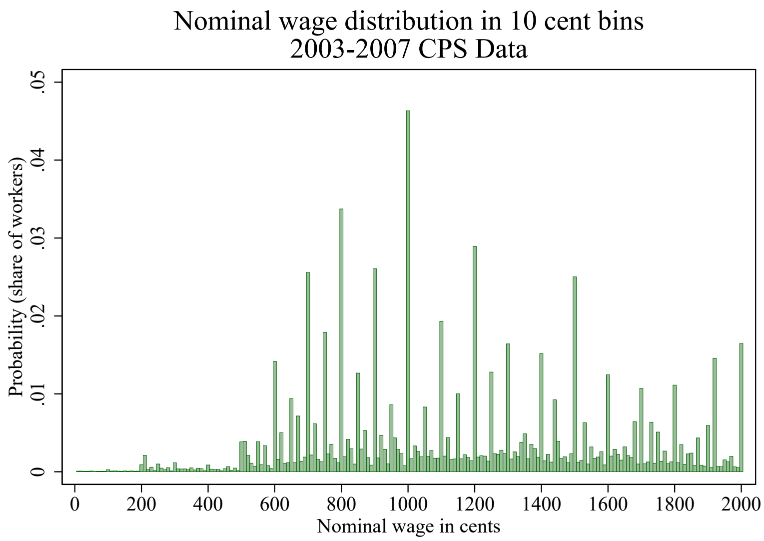
Notes. The top figure shows the CPS hourly nominal wage distribution, pooled between 2010 and 2016, in 10 cent bins. The middle figure in the middle shows the fraction of hourly wages in the CPS that end in .00 from 2003 through 2016. The bottom figure shows the fraction of states with 10.00 modal wages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

Figure 2: **Histogram of Hourly Wages In Administrative Payroll Data from MN and WA, 2003-2007**



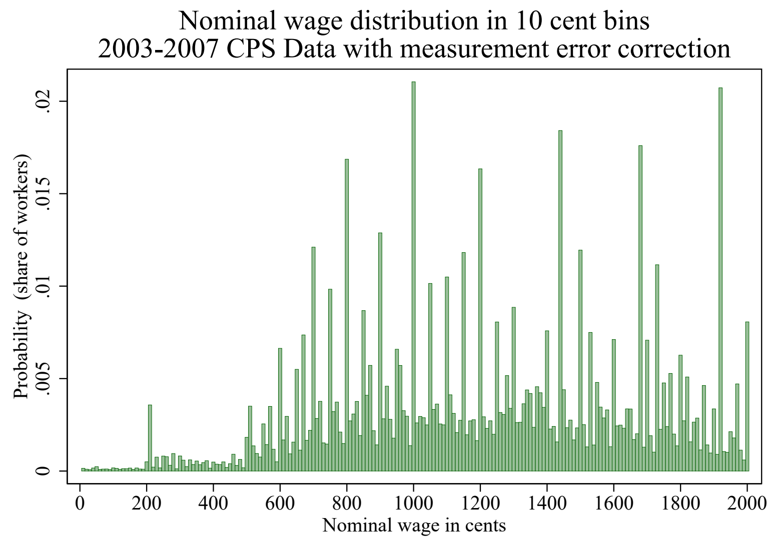
Notes. The figure shows a histogram of hourly wages from administrative Unemployment Insurance payroll records reported to states of Minnesota and Washington. The UI payroll records cover over 95% of all wage and salary civilian employment in the states. Hourly wages are constructed by dividing quarterly earnings by the total number of hours worked in the quarter. The counts here exclude NAICS 6241 and 814, home-health and household sectors, which were identified by the state data administrators are having substantial reporting errors. 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.

Figure 3: **Histogram of Hourly Wages in National CPS data, 2003-2007**



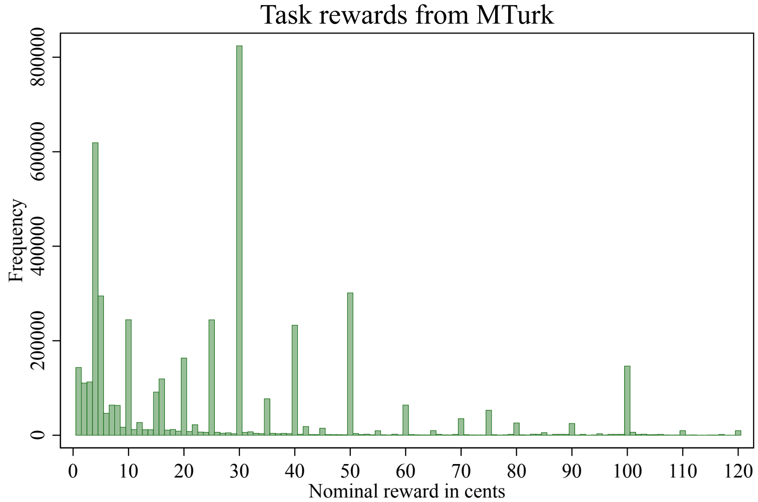
Notes. The figure shows a histogram of hourly wages by $0.10 (nominal) wage bins, averaged over 2003q1 and 2007q4, using CPS MORG files. Hourly wages are constructed by average weekly earnings by usual hours worked. The sample is restricted to those without imputed earnings. The counts here exclude NAICS 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.

Figure 4: **Wage Bunching in CPS data, 2003-2007, Corrected for Reporting Error Using 1977 CPS supplement**



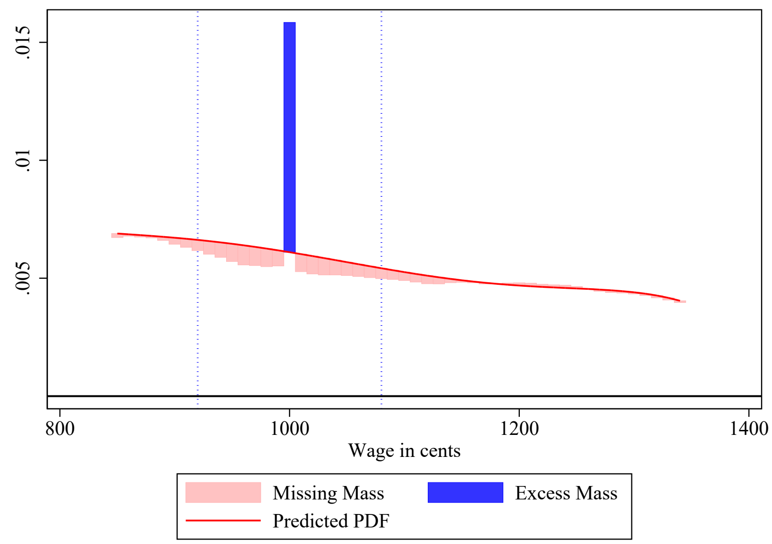
Notes. The figure shows a histogram of hourly wages by $0.10 (nominal) wage bins, averaged over 2003q1 and 2007q4, using CPS MORG files, where individual observations were re-weighted to correct for overreporting of wages ending in particular two-digit cents using the 1977 CPS supplement. Hourly wages are constructed by 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.

Figure 5: **Bunching in Task Rewards in Online Labor Markets - MTurk**



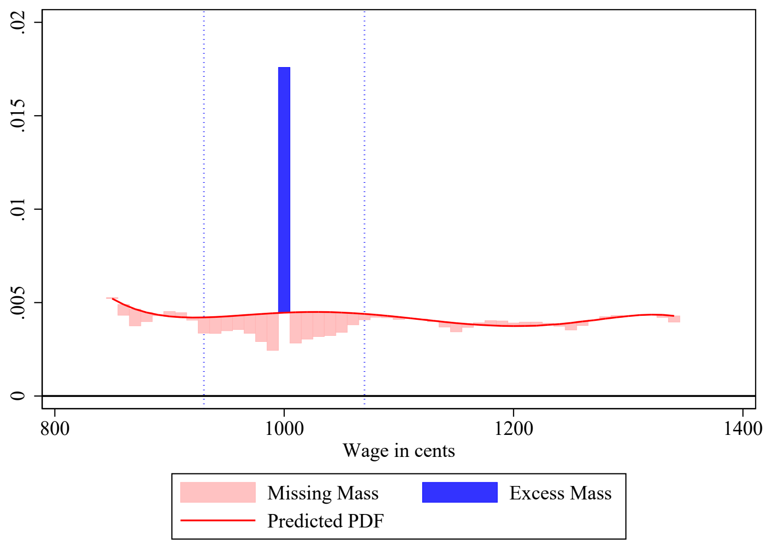
Notes. The figure shows a histogram of posted rewards by $0.01 (nominal) bins scraped from MTurk. The sample represent all posted rewards on MTurk between May 01, 2014 and September 3, 2016.

Figure 6: **Excess Bunching and Missing Mass Around $10.00 Using Administrative Data on Hourly Wages (MN, WA)**



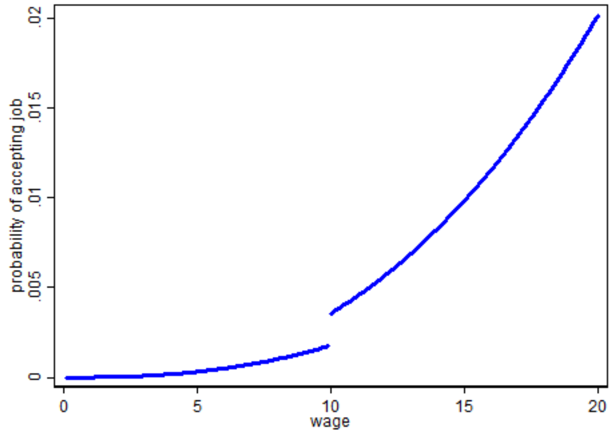
Notes. The reported estimates of excess bunching at $10.00, and missing mass in the interval around $10.00 as compared to the smoothed predicted probability density function, using administrative hourly wage counts from MN and WA, aggregated by $0.10 bins, over the 2003q1-2007q4 period. The darker shaded 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 7: **Excess Bunching and Missing Mass Around $10.00 Using Measurement Error Corrected CPS Data**



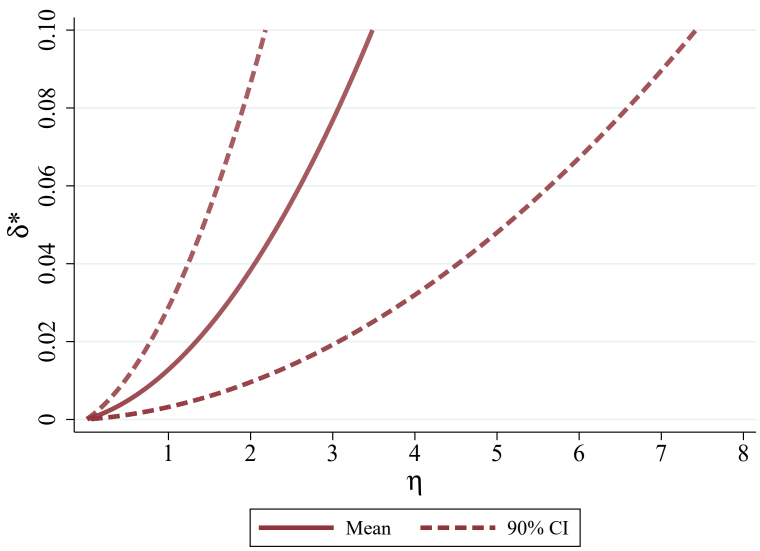
Notes. The reported estimates of excess bunching at $10.00, and missing mass in the interval around $10.00 as compared to the smoothed predicted probability density function, using CPS data corrected for measurement error using the 1977 administrative supplement. The darker shaded blue bar at $10.00 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 8: **Labor Supply Function with Left-digit Bias**



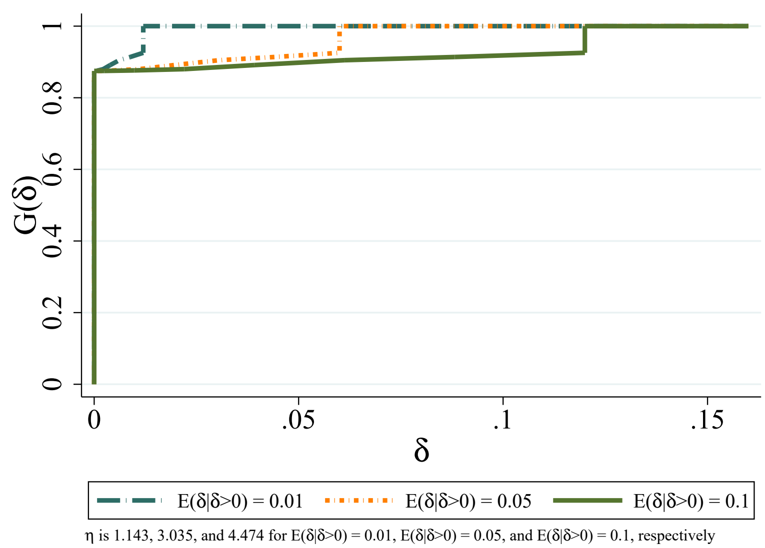
Notes. The figure shows a hypothetical labor supply function with left-digit bias, indicated by a discontinuous jump at $10. Here the elasticity $\eta = 1.5$, the constant parameter is $C = 50$ and the left-digit bias parameter $\gamma = 2$.

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



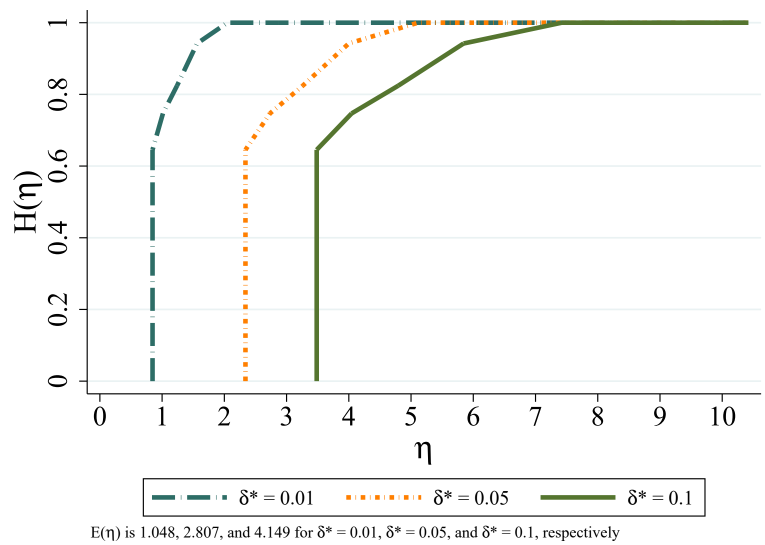
Notes. The solid, red, upward sloping line shows the locus of the labor supply elasticity η and optimization frictions $\delta^* = E[\delta|\delta > 0]$ consistent with the extent of bunching ω estimated using the administrative hourly wage data from MN and WA between 2003q1-2007q4, as described in equation 16 in the paper. The dashed lines are the 90 percent confidence intervals estimated using 500 bootstrap replicates.

Figure 10: **Implied Distribution of δ Under Constant η**



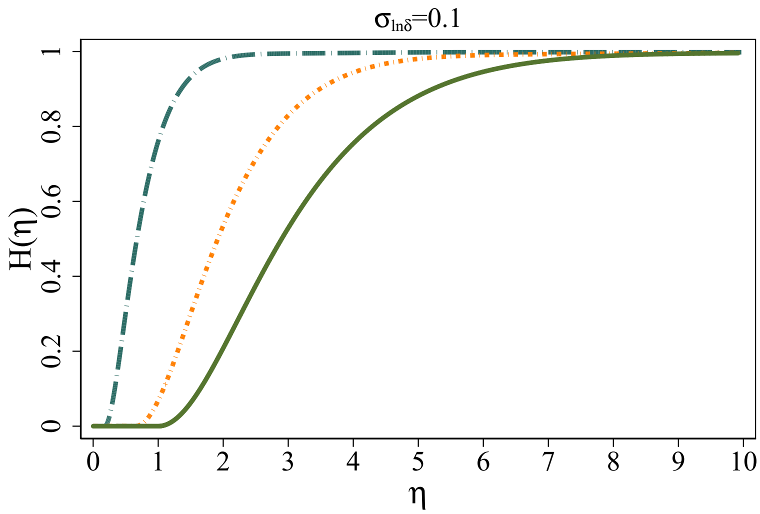
Notes. The figure plots the cumulative distributions $G(\delta)$ based on equation 17, for alternative values of $E(\delta|\delta > 0)$. The elasticity η is assumed to be a constant. The estimates use administrative hourly wage data from MN and WA.

Figure 11: **Implied Distribution of η with a 2-point Distribution of δ**

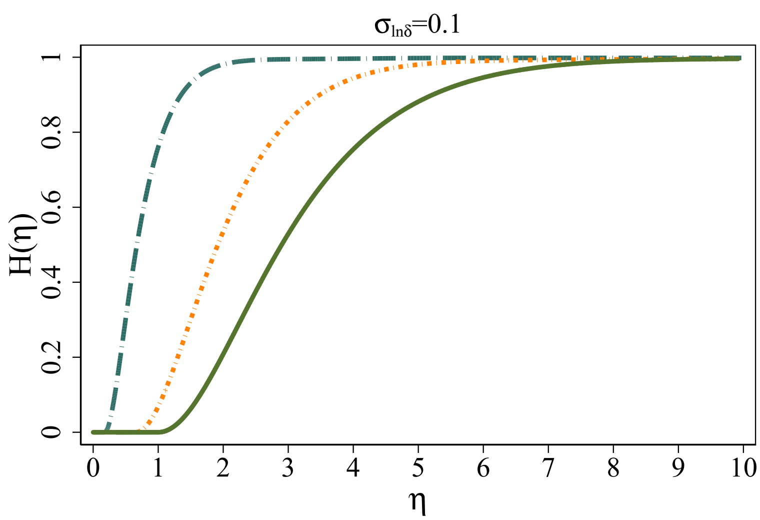


Notes. The figure plots the cumulative distributions $H(\eta)$ based on equation 18, for alternative values of $\delta^* = E(\delta|\delta > 0)$. δ is assumed to follow a 2-point distribution with $\delta = 0$ with probability $\underline{G}$ and $\delta = \delta^*$ with probability $1 - \underline{G}$. The estimates use administrative hourly wage data from MN and WA.

Figure 12: **Implied Distribution of η using a Deconvolution Estimator where δ has a Conditional Lognormal Distribution**



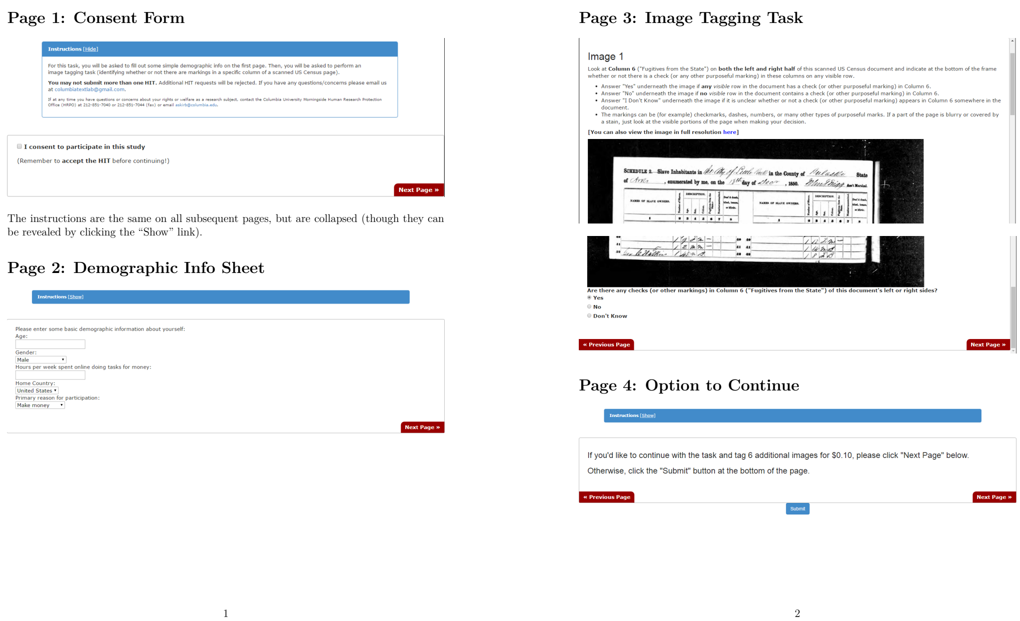
E(η) is 3.414, 8.038, and 11.644 for E(δ|δ>0) = 0.01, E(δ|δ>0) = 0.05, and E(δ|δ>0) = 0.1, respectively



E(η) is 3.414, 8.038, and 11.644 for E(δ|δ>0) = 0.01, E(δ|δ>0) = 0.05, and E(δ|δ>0) = 0.1, respectively

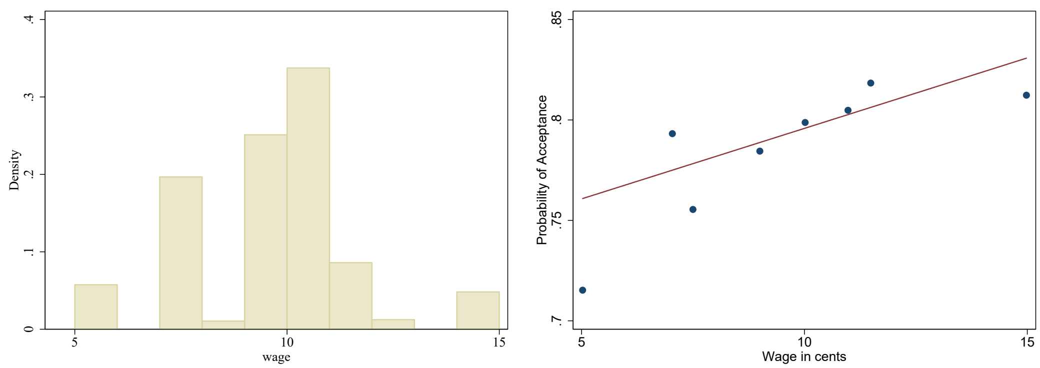
Notes. The figure plots the cumulative distributions $H(\eta)$ using a deconvolution estimator based on equation 19, for alternative values of $E(\delta|\delta > 0)$. The procedure allows for an arbitrary smooth distribution of η, while assuming δ is lognormally distributed (conditional on being non-zero) 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 and WA.

Figure 13: **Online Labor Supply Experiment on MTurk**



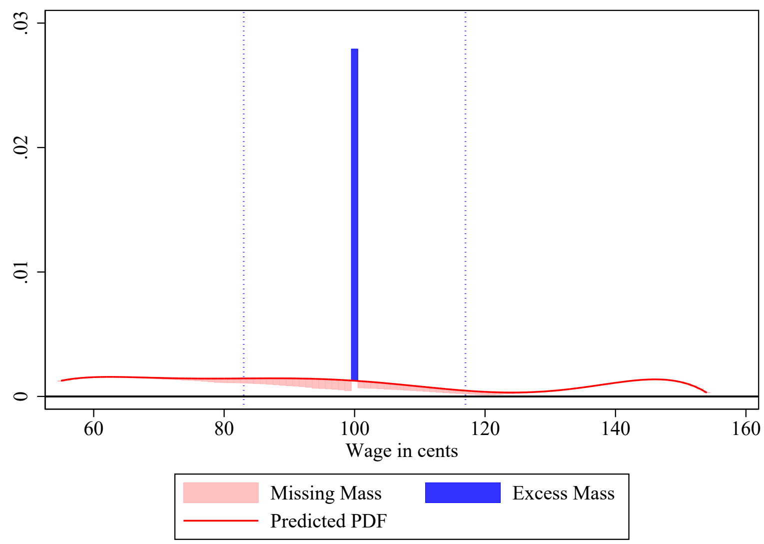
Notes. The figure shows the screen shots for the consent form and tasks associated with the online labor supply experiment on MTurk.

Figure 14: **Distribution of Randomized Rewards in the MTurk Experiment, and Resulting Probability of Task Acceptance**



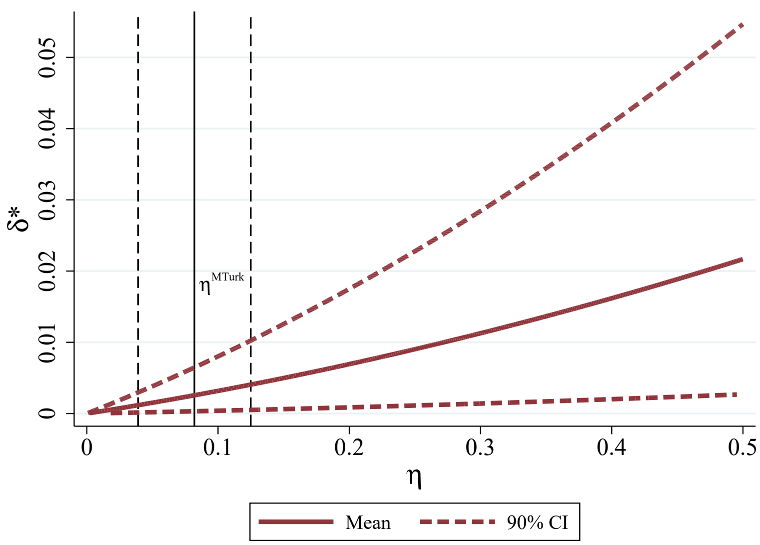
Notes. The left panel shows the density of randomized rewards in the online experiment on MTurk. The right panel shows the acceptance probabilities associated with each value of the reward.

Figure 15: **Excess Bunching and Missing Mass Around $10.00 Using Administrative Data on Rewards from Amazon Mechanical Turk**



Notes. The reported estimates of excess bunching at $1.00, and missing mass in the interval around $100 as compared to the smoothed predicted probability density function, using the universe of rewards from Amazon Mechanical Turk. The darker shaded blue bar at $100 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 $0.01 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 16: **Relationship Between Labor Supply Elasticity (η) and Optimization Frictions (δ) and Size of Bunching (ω): MTurk Data**



Notes. The solid, red, upward sloping line shows the locus of the labor supply elasticity η and optimization frictions δ consistent with the extent of bunching ω estimated using the MTurk data, as described in equation 16 in the paper. The dashed lines are the 90 percent confidence interval estimated using 500 bootstrap replicates. The vertical line shows the experimentally estimated labor supply elasticity η and the dotted vertical lines are the 90 percent confidence intervals for η.

Table 1: **Estimates for Excess Bunching, Missing Mass, and Interval around Threshold**

	(1)	(2)	(3)	(4)
Value of w_0	\$10.00	\$10.00	\$10.00	\$10.00
Excess mass at w_0	0.010	0.032	0.013	0.041
	(0.002)	(0.007)	(0.003)	(0.007)
Total missing mass	-0.013	-0.044	-0.018	-0.033
	(0.005)	(0.014)	(0.006)	(0.017)
Missing mass below	-0.006	-0.025	-0.009	-0.019
	(0.005)	(0.015)	(0.007)	(0.021)
Missing mass above	-0.007	-0.019	-0.009	-0.014
	(0.004)	(0.015)	(0.006)	(0.017)
Test of equality of missing mass below and above w_0:				
t-statistic	0.030	-0.156	-0.042	-0.159
Bunching = $\frac{Actual\ mass}{Latent\ density}$	2.596	6.229	3.942	8.394
	(0.293)	(4.386)	(1.332)	(4.689)
w_L	\$9.20	\$9.30	\$9.30	\$9.30
w_H	\$10.80	\$10.70	\$10.70	\$10.70
$\omega = \frac{(w_H - w_0)}{w_0}$	0.080	0.070	0.070	0.070
	(0.023)	(0.027)	(0.030)	(0.029)
Data:	Admin MN & WA	CPS-Raw MN & WA	CPS-MEC MN & WA	CPS-Raw

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

	Dum. for \$0.5	Dum. for \$0.25 & \$0.5	Poly. of degree 2	Poly. of degree 4	Fourier, degree 3	Fourier, degree 6
	(1)	(2)	(3)	(4)	(5)	(6)
Value of w_0	\$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
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Total missing mass	-0.012	-0.012	-0.010	-0.011	-0.008	-0.017
	(0.005)	(0.005)	(0.003)	(0.005)	(0.004)	(0.007)
Missing mass below	-0.008	-0.008	-0.005	-0.006	-0.006	-0.009
	(0.004)	(0.004)	(0.003)	(0.005)	(0.004)	(0.005)
Missing mass above	-0.004	-0.004	-0.005	-0.006	-0.002	-0.009
	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Test of equality of missing mass below and above w_0: t-statistic	-0.657	-0.729	0.150	-0.022	-0.624	0.057
Bunching = $\frac{Actual\ mass}{Latent\ density}$	2.656	2.621	2.693	2.649	2.694	2.254
	(0.312)	(0.322)	(0.258)	(0.272)	(0.251)	(0.326)
w_L	\$9.40	\$9.40	\$9.20	\$9.20	\$9.30	\$9.40
w_H	\$10.60	\$10.60	\$10.80	\$10.80	\$10.70	\$10.60
$\omega = \frac{(w_H - w_0)}{w_0}$	0.060	0.060	0.080	0.080	0.070	0.060
	(0.021)	(0.021)	(0.027)	(0.024)	(0.037)	(0.023)
Data:	Admin MN & WA	Admin MN & WA	Admin MN & WA	Admin MN & WA	Admin MN & WA	Admin MN & WA

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. 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. Bootstrap standard errors based on 500 draws are in parentheses.

Table 3: **Bounds for Labor Supply Elasticity in Offline Labor Market**

	(1)	(2)	(3)	(4)
A. $\delta^* = 0.01$				
$\bar{\delta}$	0.001	0.003	0.003	0.002
η	0.846	1.014	1.014	1.014
90% CI	[0.472, 2.050]	[0.538, 4.525]	[0.538, 9.512]	[0.503, 4.525]
95% CI	[0.417, 2.871]	[0.538, 4.525]	[0.472, 9.512]	[0.417, 4.525]
B. $\delta^* = 0.05$				
$\bar{\delta}$	0.005	0.015	0.013	0.011
η	2.339	2.733	2.733	2.733
90% CI	[1.429, 5.112]	[1.593, 10.692]	[1.593, 21.866]	[1.508, 10.692]
95% CI	[1.291, 6.970]	[1.593, 10.692]	[1.429, 21.866]	[1.291, 10.692]
C. $\delta^* = 0.1$				
$\bar{\delta}$	0.011	0.030	0.025	0.023
η	3.484	4.045	4.045	4.045
90% CI	[2.182, 7.421]	[2.418, 15.319]	[2.418, 31.127]	[2.295, 15.319]
95% CI	[1.983, 10.053]	[2.418, 15.319]	[2.182, 31.127]	[1.983, 15.319]
$G(0) = \underline{G}$	0.894	0.703	0.750	0.772
Data:	Admin MN & WA	CPS-Raw MN & WA	CPS-MEC MN & WA	CPS-Raw

Notes. The table reports point estimate and associated 90 and 95 percent confidence intervals for labor supply elasticties, η, associated with different values of optimization friction δ for the offline labor market. The datasets are administrative hourly wage data ,CPS-MEC, and CPS from MN and WA as well as national CPS data. In rows A, B and C, we use hypothesized values of δ of 0.01, 0.05 and 0.1 respectively. The labor supply elasticity, η, is estimated using the estimated extent of bunching, ω, and the hypothesized δ, using equation 16 in the paper. The 90 and 95 percent confidence intervals in square brackets are estimated using 500 boostrap draws.

Table 4: **Bounds for Labor Supply Elasticity in Offline Labor Market — Robustness to Specifications of Latent Wage**

	Dum. for \$0.5 (1)	Dum. for \$0.25 & \$0.5 (2)	Poly. of degree 2 (3)	Poly. of degree 4 (4)	Fourier, degree 3 (5)	Fourier, degree 6 (6)
A. $E(\delta \mid \delta > 0)$= 0.01						
$\overline{\delta}$	0.001	0.001	0.001	0.001	0.001	0.002
η	1.240	1.240	0.846	0.846	1.014	1.240
90% CI	[0.538, 2.050]	[0.472, 2.050]	[0.417, 2.050]	[0.472, 2.050]	[0.300, 2.871]	[0.618, 2.871]
95% CI	[0.472, 2.050]	[0.472, 2.050]	[0.372, 2.871]	[0.372, 2.050]	[0.247, 2.871]	[0.472, 4.525]
B. $E(\delta \mid \delta > 0)$= 0.05						
$\overline{\delta}$	0.006	0.007	0.004	0.005	0.005	0.008
η	3.260	3.260	2.339	2.339	2.733	3.260
90% CI	[1.593, 5.112]	[1.429, 5.112]	[1.291, 5.112]	[1.429, 5.112]	[0.984, 6.970]	[1.791, 6.970]
95% CI	[1.429, 5.112]	[1.429, 5.112]	[1.174, 6.970]	[1.174, 5.112]	[0.839, 6.970]	[1.429, 10.692]
$G(0)=\underline{G}$	0.871	0.865	0.917	0.907	0.908	0.830
Data:	Admin MN & WA	Admin MN & WA	Admin MN & WA	Admin MN & WA	Admin MN & WA	Admin MN & WA

Notes. The table reports point estimate and associated 90 and 95 percent confidence intervals for labor supply elasticities, η, associated with hypothesized δ=0.01 and δ=0.05 for the offline labor market. All columns employ the administrative hourly wage data 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. In row A, we hypothesize δ = 0.01; whereas it is δ = 0.05 in row B. The labor supply elasticity, η, is estimated using the estimated extent of bunching, ω, and the hypothesized δ, using equation 16 in the paper. The 90 and 95 percent confidence intervals in square brackets are estimated using 500 boostrap draws.

Table 5: **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
Excess mass at w_0	0.010	0.010	0.010	0.010
	(0.002)	(0.002)	(0.002)	(0.002)
Total missing mass	-0.013	-0.013	-0.013	-0.013
	(0.005)	(0.005)	(0.005)	(0.005)
Bunching = $\frac{\textit{Actual mass}}{\textit{Latent density}}$	2.596	2.596	2.596	2.596
	(0.293)	(0.293)	(0.293)	(0.293)
A. $E(\delta\|\delta > 0)$= 0.01				
$\bar{\delta}$	0.001	0.001	0.001	0.001
η	1.143	1.175	1.323	1.590
90% CI	[0.618, 3.080]	[0.604, 3.403]	[0.561, 3.026]	[0.697, 3.868]
95% CI	[0.543, 3.619]	[0.526, 3.907]	[0.496, 3.494]	[0.575, 4.565]
B. $E(\delta\|\delta > 0)$= 0.05				
$\bar{\delta}$	0.006	0.006	0.006	0.006
η	3.035	3.097	3.431	4.029
90% CI	[1.791, 7.443]	[1.755, 8.126]	[1.648, 7.270]	[1.966, 9.186]
95% CI	[1.606, 8.655]	[1.557, 9.286]	[1.486, 8.355]	[1.672, 10.733]
$G(0)$= $\underline{G}$	0.875	0.875	0.875	0.875
Data:	Admin MN & WA	Admin MN & WA	Admin MN & WA	Admin MN & WA

Notes. The table reports point estimate and associated 90 and 95 percent confidence intervals for labor supply elasticities, η, associated with hypothesized δ=0.01 and δ=0.05 for the offline labor market. All columns employ the administrative hourly wage data. Heterogeneous δ, and η are allowed in columns 1 and 2, using equations 17 and 18, respectively. Columns 3 and 4 allow heterogeneous δ and η, and assuming a conditional lognormal distribution of δ, using a deconvolution estimator based on equation 19. 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.

Table 6: **Bounds for Labor Supply Elasticity in Offline 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	0.015	0.029
	(0.003)	(0.004)	(0.006)	(0.003)	(0.003)	(0.006)
Total missing mass	-0.011	-0.012	-0.042	-0.012	-0.016	-0.024
	(0.009)	(0.007)	(0.013)	(0.006)	(0.007)	(0.013)
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. $E(\delta\|\delta > 0)$= 0.01						
$\overline{\delta}$	0.002	0.001	0.003	0.001	0.002	0.003
η	1.014	0.846	0.846	0.846	0.846	1.014
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. $E(\delta\|\delta > 0)$= 0.05						
$\overline{\delta}$	0.009	0.005	0.014	0.007	0.008	0.013
η	2.733	2.339	2.339	2.339	2.339	2.733
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 estimate and associated 90 and 95 percent confidence intervals for labor supply elasticties, η, associated with hypothesized δ=0.01 and δ=0.05 for the offline labor market. All columns employ the national 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 δ = 0.01; whereas it is δ = 0.05 in row B. The labor supply elasticity, η, is estimated using the estimated extent of bunching, ω, and the hypothesized δ, using equation 16 in the paper. The 90 and 95 percent confidence intervals in square brackets are estimated using 500 boostrap draws.

Table 7: **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 8: **Task quality by offered task reward on MTurk**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Wage	-0.006	-0.002	0.011		0.001	0.011	0.031	
	(0.012)	(0.017)	(0.017)		(0.022)	(0.033)	(0.034)	
Jump at 10		-0.002				-0.006		
		(0.007)				(0.013)		
Spline			-0.019				-0.052	
			(0.067)				(0.127)	
Local				0.003				0.012
				(0.011)				(0.022)
Global				-0.003				-0.002
				(0.006)				(0.012)
η	-0.006	-0.002	0.011		0.001	0.011	0.032	
	(0.012)	(0.017)	(0.017)		(0.023)	(0.034)	(0.035)	
Sample	Pooled	Pooled	Pooled	Pooled	Sophist.	Sophist.	Sophist.	Sophist.
Sample Size	4073	4073	4073	4031	1407	1407	1407	1396

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 5-8 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 9: **Estimates for Round Number Bunching, Labor Supply Elasticity and Optimization Frictions: MTurk Data**

Value of w_0	\$1.00
Excess mass at w_0	0.027
	(0.003)
Total missing mass	-0.023
	(0.010)
Missing mass below	-0.014
	(0.013)
Missing mass above	-0.009
	(0.013)
Test of equality of missing mass below and above w_0:	
t-statistic	-0.212
Bunching = $\frac{Actual\ mass}{Latent\ density}$	22.104
	(16.040)
w_L	\$0.83
w_H	\$1.17
$\omega = \frac{(w_H - w_0)}{w_0}$	0.170
	(0.064)
η	0.082
	(0.026)
δ^*	0.003
90% CI for δ^*	[0.000, 0.007]
95% CI for δ^*	[0.000, 0.008]
$G(0)= \underline{G}$	0.748

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 \$(w_{L},w_{H})\$ from which the missing mass is drawn. It also reports the bunching, and ω, both estimated using observational MTurk data, along with the experimentally estimated labor supply elasticity, η. Finally, the extent of optimization frictions is estimated using η and ω using equation 16 in the paper. The 90 and 95 percent confidence intervals in square brackets are estimated using 500 bootstrap replicates. Bootstrap standard errors based on 500 draws are in parentheses.

Online Appendix A Observational Results From Amazon Turk

Online Appendix A.1 Upwards Sloping Job-Specific Labor Supply Curves in an Online Labor Market.

We define a task as a unique combination of description, reward, time allotted, expiration date, and requester. In our 856 day population of 4,504,696 posted tasks, 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, and make 2 estimates of the time of the task. The first, which can only be calculated for a subset of the data, involves parsing the text in the description of the task, which will sometimes contain time information like "1 hour 20 minutes 30 seconds". This is defined for just over 10% of our data. The other, which is much less precise but exists for the whole dataset is the time allotted by the requester, which will almost always be an overestimate of the actual time, but hopefully tracks the actual time. Note that time allotted is also how much time a Turker has to do the task, and if the task is too long relative to the time allotted, it may expire before the Turker can do the task.

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. Thus we measure the time until the job disappears as duration of the posting. While tasks may disappear due to requesters canceling them rather than being filled, this is rare. We thus treat the duration of the task posting as a measure of time until 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. We estimate regressions of the form:

$$log(duration_{hrdt}) = \sum_{k=0}^{10} \delta_k \mathbb{1}\left\{mod(reward_h, 10) = k\right\} + \beta log(reward_h)$$
$$+\gamma log(timealloted_h) + \delta_d + \delta_r + \delta_t + \epsilon_{trdh}$$

Where h indexes hit tasks, d indexes description, r indexes requesters, and t indexes the first 6 minute interval that the job was posted in.

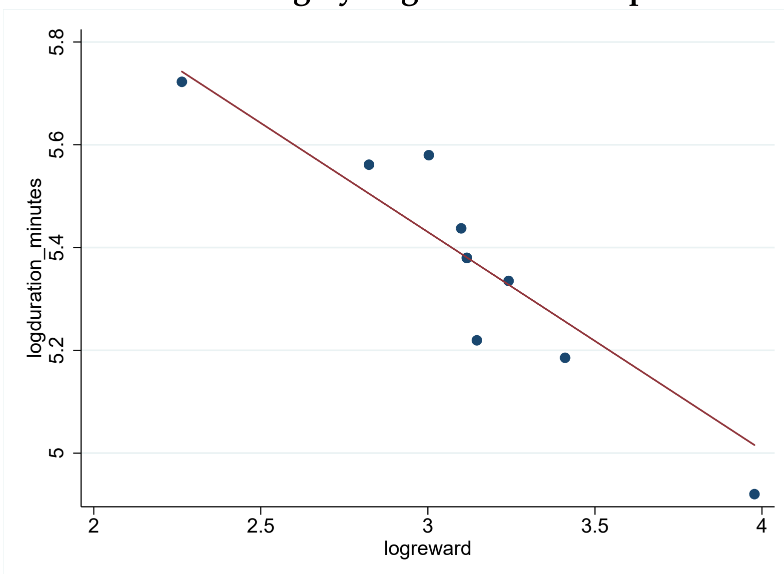
Online Appendix A.2 Do Experienced Requesters Use Fewer Round Numbers?

In this Appendix section, we show that there does not seem to be a decay in the probability of posting a round number with requester experience. Our primary measure of requester experience is number of days since first posting a task request on Amazon Turk.We estimate specifications of the form:

$$RewardDivisibleby10_{trdh} = \beta Experience_r + \delta_r + \delta_d + \delta_t + \epsilon_{trdh} \tag{24}$$

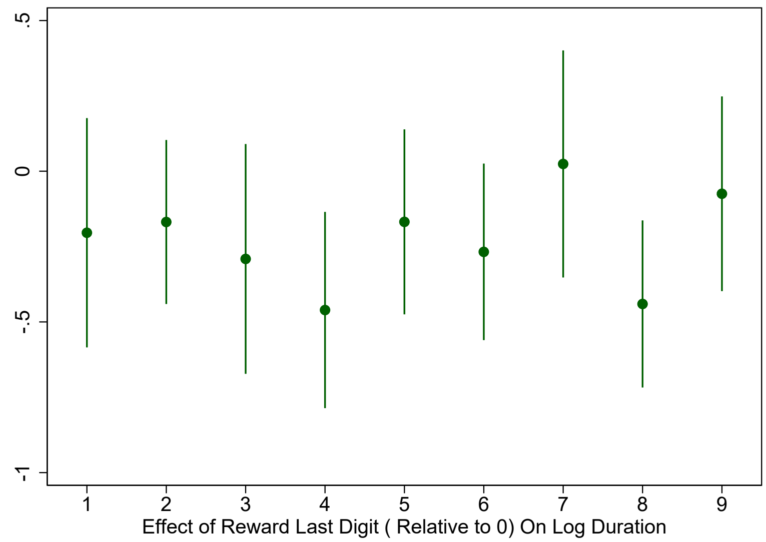
If sophistication explained the lack of bunching, we would expect more experienced posters to use fewer round numbers. Table A2 suggests that there is little evidence of this: for any measure of divisibility (by 10,5, or 100) experience has no effect on the divisibility of rewards posted. This suggests that the bunching on Amazon Turk is not an artifact of naive or inexperienced requesters.

Figure A.1: **Duration of Task Posting by Log Reward - Scraped Observational Sample**



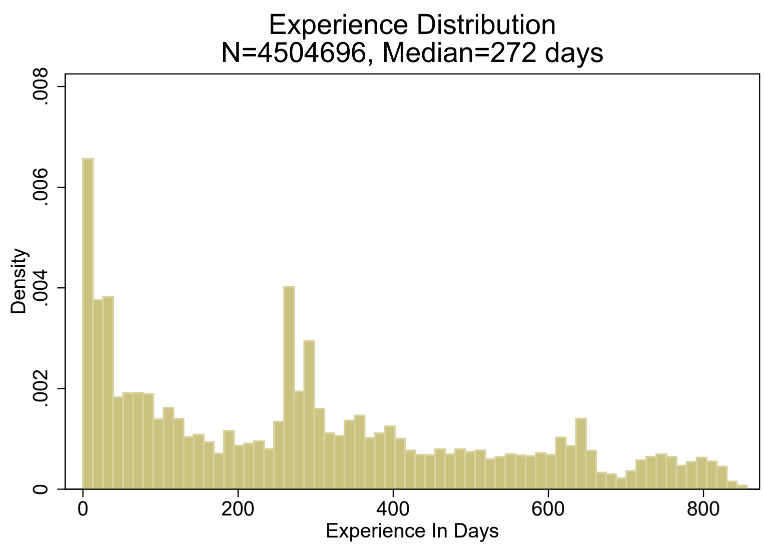
Notes: The figure shows the a binned scatterplot of log duration (in minutes) of how long a task was posted on MTurk on log reward (in cents), controlling for requester fixed effects and log time allotted. The scraped sample constitutes of all HITs posted on MTurk between May 1, 2014 and September 3, 2016.

Figure A.2: **Effects of Reward Last Digit on Duration of Task Posting**



Notes: The figure shows effects of the last digit of the reward on log duration in minutes (with 0 as the omitted category) , controlling for log wage in cents, requester fixed effects and log time allotted - for the scraped observational sample.

Figure A.3: **Density of Experience in Scraped Sample**



Notes: The figure shows the distribution of experience in days in the observational sample.

Table A.1: **Duration of Task Posting by Wage**

	(1)	(2)	(3)	(4)	(5)	(6)
Log(Reward)	-0.268***	-0.508**	-0.507**	-0.512**	-1.066	-1.223
	(0.0502)	(0.185)	(0.186)	(0.185)	(0.657)	(0.742)
Log(Time Alloted)	-0.0304	-0.823***	-0.850***	-0.848***		
	(0.111)	(0.135)	(0.128)	(0.128)		
Ends in 1	-0.204	-0.166	-0.171	-0.153	-0.621*	-0.119
	(0.194)	(0.240)	(0.241)	(0.248)	(0.276)	(0.192)
Ends in 2	-0.168	-0.0325	-0.0323	-0.0326	-0.0913	-0.0499
	(0.139)	(0.162)	(0.163)	(0.164)	(0.133)	(0.157)
Ends in 3	-0.291	-0.0782	-0.0840	-0.0787	-0.190	-0.165
	(0.194)	(0.151)	(0.152)	(0.151)	(0.655)	(0.664)
Ends in 4	-0.461**	-0.434**	-0.426**	-0.420*	-0.0866	-0.108
	(0.166)	(0.158)	(0.161)	(0.164)	(0.441)	(0.419)
Ends in 5	-0.168	0.129	0.132	0.132	-0.277	-0.262
	(0.157)	(0.152)	(0.152)	(0.152)	(0.476)	(0.476)
Ends in 6	-0.268	-0.0533	-0.0565	-0.0563	-0.274	-0.286
	(0.149)	(0.154)	(0.156)	(0.154)	(0.464)	(0.453)
Ends in 7	0.0241	-0.110	-0.110	-0.107	-0.956	-0.950
	(0.192)	(0.227)	(0.227)	(0.224)	(0.570)	(0.604)
Ends in 8	-0.440**	-0.105	-0.0958	-0.0926	0.0620	0.0657
	(0.141)	(0.133)	(0.135)	(0.135)	(0.258)	(0.273)
Ends in 9	-0.0748	0.163	0.165	0.125	0.255	-0.162
	(0.165)	(0.180)	(0.180)	(0.158)	(0.570)	(0.387)
Log(Task Time)					0.0738	-0.267
					(1.764)	(1.584)
Controls	No	No	No	Yes	No	Yes
N	2795741	2772258	2772138	2771570	374541	374541
Clusters	10418	8645	8535	8477	2202	2202
Requester FE	Yes	No	Yes	No	Yes	No
Description FE	No	Yes	Yes	No	No	No
Req. X Desc. FE	No	No	No	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Within R2	0.00276	0.00762	0.00769	0.00920	0.0134	0.0337
R2	0.849	0.902	0.902	0.902	0.969	0.970

Notes. Significance levels are * 0.10, ** 0.05, *** 0.01.

Table A.2: **Divisibility of Reward by Experience of Requester:Scraped Sample**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
experience	0.000306	0.00446	0.102	0.000348	0.0182	0.00506	0.000185	0.00887	-0.000163
	(0.000195)	(0.0197)	(0.0829)	(0.000208)	(0.0194)	(0.00298)	(0.000111)	(0.0119)	(0.000272)
N	4504696	4470771	4442166	4504696	4470771	4442166	4504696	4470771	4442166
Clusters	11680	11069	9147	11680	11069	9147	11680	11069	9147
Requester FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Description FE	No	No	Yes	No	No	Yes	No	No	Yes
Time FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Within R2		2.86e-08	0.00000904		0.000000712	2.10e-08		0.000000441	2.36e-10
R2	0.0206	0.932	0.984	0.0309	0.948	0.980	0.0208	0.952	0.996

Notes. Columns 1-3 have Divisibility by 10, Columns 4-6 have Divisibility by 5, and Columns 7-9 have Divisibility by 100. Even Columns control for description and time fixed effects. Standard Errors clustered by requester.
*Significance levels are * 0.10, ** 0.05, *** 0.01.*

Online Appendix B Additional Experimental Specifications

Online Appendix B.1 Pre-analysis plan specifications

In Tables B.1 and B.2 we show specifications from our pre-analysis plan that parallel those in 7 and 8, 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. (2016), and improves precision on the elasticity estimate.

Table B.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 probabilties 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 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 B.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 probabilties 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 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.
* $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$

Online Appendix C 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 limited consequence of optimization errors 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. Shirking is detected by employers with probability D. Following Rebitzer and Taylor (1995), we allow the detection of shirking to falling in the amount of employment $l(w)$.[19] Workers quit with an exogenous rate q. An unemployed worker receives benefit b and finds an offer at rate s. 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 and Shapiro and Stiglitz, we allow the wages offered by firms to vary; indeed our model will predict higher productivity firms will pay higher wages—leading to equilibrium wage dispersion.

[19] 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. 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 can write the value of not shirking can be written as:

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

The value of shirking can be written as:

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

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 $EV^N$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 decling 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 marginal cost of labor to the marginal revenue product of labor (p_i), i.e., $w_i = \frac{\eta}{1+\eta} p_i$.[20]

Finally, we can augment this labor supply function to exhibit left-digit bias. Consider the case where for wage $w \geq w_0$, the perceived wage 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:

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

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

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

Note that under the condition that $d(x)$ has a constant elasticity, the implicit labor supply elasticity continues to constant both below and above w_0. However, there is a discontinuous jump up in $l(w)$ 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 go through here as well. A low η due to (say) high cost of effort now implies 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 on-line labor markets is more consistent with a monopsony interpretation than an effort one. 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.

Online Appendix D Deconvolution estimator

In this appendix, we describe the deconvolution estimator we use to estimate the distributions of the the elasticity η and δ. Recall that if we condition on $\delta > 0$,we can take logs of equation 14 to obtain:

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

We make the assumption that δ_{res} is lognormally distributed, so that $\ln(\delta_{res}) \sim N(0, \sigma_\delta^2)$, 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 $1 - \hat{\phi}(\exp\{2\ln(\omega)\})$ to numerically obtain a density for $2\ln(\omega)$, where $\hat{\phi}$ is empirically estimated from the shape of the missing mass. 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 $-\ln(\eta(1+\eta))$.

To see this, 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:

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

This estimator also requires a choice of bandwidth which is a function of sample size. Delaigle and Gijbels (2004) also 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.