1 Misestimation

We elicit recalls from workers about their main job outcomes. We define those to be weekly pay, weekly hours, gross hourly pay, net hourly pay and the sharethat expenses represent out of total pay. We ask as well for forecasts of job outcomes. The forecast for gross weekly pay is incentivized for accuracy through a Quadratic Scoring Rule, with a bonus that goes up to $5. Forecasts for weekly hours and hourly pay are also elicited but are not incentivized.

To calculate misestimation of outcomes (either for recall or forecast beliefs), we subtract actual job outcomes from beliefs. As such, when this measure is positive, it implies overestimation, and when it is negative, it implies underestimation. The exception is expenses misestimation. Due to its nature as a cost, this measure is defined as actual expenses minus belief of expenses. In this case, a positive value entails underestimation of expenses.

2 Information Treatment

We now discuss our randomized information treatment. In the Baseline survey, after submitting their recall and forecast beliefs, subjects are randomized into the treatment group with 50% probability. Next, in a single page we tell them: (i) how we calculated their gross hourly pay and its value; (ii) show an example of their expected expenses share given their car; (iii) calculate their actual net hourly pay based on this information; (iv) compare the actual net hourly pay with their recall, informing them if they are either under or overestimating it. On the following page, we tell them that overestimation is common among gig workers in our sample and provide them with a brief explanation of the concept of overconfidence.

In the Midline survey, the treatment group is shown the exact same information treatment pages they were presented in the Baseline survey. This is meant to work as a a reinforcement of the information initially presented to them. Furthermore, they are also given feedback on their forecast, and told whether they over or underestimated it. If a worker was randomized into the control group, we provide them with the same information, but only at the end of the Endline survey.

Our estimation strategy aims to causally identify the effect of information provision and knowledge about job outcomes on beliefs and decisions. In practice, we use three main specifications. The first one is

$$z_{it} = \beta_0 + \beta_1 Treat_i + X_{i0}\Gamma + \varepsilon_{it} \tag{1}$$

where z_{it} are either practices related to information acquisition (such as checking weekly pay) or log misestimation of job outcomes for individual i at period t. Expenses share (which ranges from 0 to 1) is

not in log form. $Treat_i$ is the treatment binary variable, equal to 1 if an individual is in the treatment group. X_{i0} is the covariates matrix, composed of binary variables for gender, race, education, age, among others. Finally, ε_{it} is the regression error and t is 0 to 16 weeks after the Baseline survey. The average treatment effect is identified by β_1.

Most effects from our information treatment should be a function of what signal workers receive: are they told they overestimate or underestimate their net hourly pay? That is, we expect opposite responses from a worker that receives a positive signal (e.g., they think their net hourly pay is \$10/hour, when it is actually \$12/hour) and a worker that receives a negative signal (e.g., they think their net hourly pay is \$14/hour, when it is actually \$12/hour).

The above specification compares the treatment group as a whole with the control group, with no regard for differential effects due to the initial level of misestimation of net hourly pay. To take that into account, we use one of the two following specifications:

$$y_{it} = \beta_0 + \beta_1 Over_i + \beta_2 Treat_i + \beta_3 Treat_i \cdot Over_i + X_{i0}\Gamma + \varepsilon_{it} \tag{2}$$

$$y_{it} = \beta_0 + \beta_1 Mis_i + \beta_2 Treat_i + \beta_3 Treat_i \cdot Mis_i + X_{i0}\Gamma + \varepsilon_{it} \tag{3}$$

where y_{it}, for individual i at period t, can be the log recall beliefs of job outcomes, the log misestimation of job outcomes (defined as the difference between the log of the recall belief and the log of the actual outcome), the log actual job outcomes, a binary variable for job search or information on other jobs, or an extensive margin of labor supply measure. The expenses share (which ranges from 0 to 1) is again not in log form. Mis_i is the initial log misestimation in net hourly pay (defined as the initial log recall belief minus the initial log actual outcome) and $Over_i$ equals 1 if an individual is initially overestimating net hourly pay. X_{i0}, in addition to our set of controls defined above, may also include the pre-treatment outcome variable. Period t is 0 to 16 weeks after the Baseline survey. All regressions include standard errors robust to heteroskedasticity.

In equation (2), β_2 identifies the average treatment effect for those initially underestimating net hourly pay, while β_3 is the differential effect of the treatment on those initially overestimating. In equation (3), β_2 identifies the intercept of the treatment effect, while β_3 identifies its slope: it identifies how the treatment effect varies depending on the value of the log initial misestimation of net hourly pay.

As mentioned above, we do a log transformation of most of our outcome variables. This is done in order to minimize the influence of outliers, as these outcomes vary over 3 orders of magnitude in our sample. On equation (3), we make the simplifying assumption that the treatment effect is symmetric between initially overestimating and underestimating workers, which we relax in robustness checks.

The effect of the information treatment on the intensive margin of labor supply is ambiguous: on the one hand, workers learning they are overestimating their pay should reduce hours worked due to the substitution effect. On the other hand, budget-constrained workers in the same situation can actually increase their hours worked to be able to fulfill their household budget. Thus, we do an heterogeneity analysis along this dimension for the treatment effects on labor supply decisions.