Testing Expectation-Based Reference Dependent Utility Theory: A Reanalysis of A Framed Field Experiment with Low-Income Brazilian Piece-Rate Workers *

Ketaki Sarin†

November 30, 2024

Abstract

An abundance of literature has sought to identify the determinants of labour supply, gradually moving away from the neoclassical model towards behavioural theories that incorporate loss aversion and income-targetting by workers. This paper leverages data from a framed field experiment conducted in Northeast Brazil (Stockley (2018)), that sought to analyse the influence of expectations of income on subjects' output and effort decision. I re-examine the results using novel statistical innovations in machine learning to improve the precision of the estimates, and test for the existence of reference-dependent and neoclassical effects through a stylized model of the optimal stopping decision of workers. Much like what can be drawn from existing literature on this topic, the labour supply decision seems to internalize components of both reference-dependent and neoclassical models. In general, there is a pronounced income effect of increasing the expected earnings of participants. Moreover, I find evidence that income targets are operationalized from expectations based on past experience.

*This paper was written in fulfillment of the requirements of ECO499 at the University of Toronto. I thank Lisa Stockley (PhD) for sharing this data from her doctoral dissertation. I also extend my heartfelt gratitude to Prof. Gustavo J. Bobonis, University of Toronto, for his supervision and mentoring throughout the process of writing this paper. I am solely responsible for any errors.

†Economics and Mathematics Specialist, University of Toronto, email: ketaki.sarin@mail.utoronto.ca

1 Introduction

The standard intertemporal model of labour supply predicts that workers respond to a transitory increase in wages by increasing their labour supply. This theory is justified by an increase in the opportunity cost of leisure with a higher wage, which - all else held constant - results in a higher level of labour supplied in equilibrium. However, there is considerable empirical evidence showing negative or insignificant wage elasticities, giving rise to a range of alternative formulations of the theory of labour supply.

A growing body of literature in behavioural economics suggests that labour supply decisions are strongly influenced by reference points and narrow-bracketing (Camerer et al. (1997) Fehr and Goette (2007) Abeler et al. (2011) Crawford and Meng (2011)). It advances that in addition to viewing income in levels, individuals have preferences over 'gains' and 'losses' relative to an income target (Kahneman and Tversky (1979)). In this paper, I seek to conduct a reanalysis of the framed field experiment by Stockley (2018), with a focus on a theory of expectation-based reference points as proposed by Kőszegi and Rabin [henceforth KR]. According to this theory, reference points are determined by individuals' recently held probabilistic beliefs (i.e expectations), and utility is derived from gains or losses relative to these rational expectations. For loss averse individuals, the drop in utility from a loss is felt more severely than a rise in utility from a gain, resulting in a discontinuous drop in the marginal benefit of working once income surpasses the reference point (Kahneman and Tversky (1979)).

Abeler et al. (2011) tested the predictions of the KR model in a laboratory experiment, finding overwhelming evidence in support of the existence of reference points derived from rational expectations. In their experiment, subjects were made to complete a tedious and repetitive task for a piece-rate wage, allowing them to flexibly choose when to stop working. Payment was made on the basis of a lottery - with probability 0.5 they were paid their accumulated earnings upto the point they chose to work, and with probability 0.5 they were paid a known fixed payment. The experiment sought to exogenously manipulate expectations by offering one of two possible fixed payments (High or Low). Under the assumption that the fixed payment acts as a rational income target for the subjects, the combination of loss aversion and reference targetting meant that individuals would choose to reduce the disparity between the two possible states of payment. Consequently, individuals who were offered the higher fixed payment are predicted to work more (till accumulated earnings = fixed payment) and vice-versa. Strong evidence in favour of this prediction was realized under Abeler et. al's innovative experiment design.

This paper uses data collected by Lisa Stockley, as a part of her doctoral dissertation at the University of Toronto. Stockley (2018) conducted a real world replication of Abeler et. al's experiment with a sample of low-income individuals involved in piece-rate work in Northeast Brazil. The experiment was conducted in two shifts. The first shift simulated the design of Abeler et. al, where workers' expectations were manipulated by a lottery-based payment contract in an open-ended work shift. After subjects flexibly chose the amount of output to produce, their payment was determined by a coin flip - either to receive their accumulated piece-rate earnings or the fixed payment. Under a neoclassical framework, the fixed payment changes neither the marginal cost or marginal benefit to effort, and hence should not influence the labour supply decision. However, if individuals' preferences align with the KR model, then effort should be increasing in their probabilistic beliefs of income, which is proxied by the fixed payment. The second shift of the experiment was a novel extension by Stockley, conducted three weeks after the first. A sub sample of subjects (initially offered the higher fixed payment) agreed to return for the same open-ended work shift with a lottery payment, but were unexpectedly faced with a wage shocks (i.e. either an increase or decrease from the piece-rate wage in Shift 1). This channel potentially actualized an evolution of the expectations of income for the subjects, based on their outcome in the previous shift, which did not necessarily coincide with the value of the fixed payment.

In general, I find that the fixed payment in the first shift does not significantly affect the probability of stopping for the subjects, for the full sample of participants with variations in both the wage and fixed payment. In fact, in the estimation that is the closest analogue to Abeler et al. (2011), I find a significantly negative elasticity of output with respect to the fixed payment offered, a result that favours the dominance of an income effect rather than KR preferences. Based on these results alone, I fail to conclude that rational expectations of income induce reference-dependence in the optimal labour supply decision.

The second shift, however, which analyses 'adaptive' expectations based on a subject's own experience in Shift 1, shows stronger evidence in favour of reference-dependence. Specifically, I find that matching the income target proxied by the earnings that were realized (after the lottery) in Shift 1 significantly affected the optimal stopping decision. This effect was more pronounced for subjects who faced an unanticipated wage cut going from Shift 1 to Shift 2.

These results add to the growing body of field evidence on labour supply decisions, leveraging a framed experimental set-up. It highlights the importance of behavioural models in capturing the essence of labour supply decisions in naturally-occurring economic and work environments, where the neoclassical models falls short. Moreover, by virtue of the population of this experiment, it is one of the only few studies that targets an important segment of underrepresented workers - those engaged in piece-rate daily production in less developed regions.

The rest of this paper is organised as follows: Section 2 provides an overview of existing literature on the topic of labour supply, with an emphasis on behavioural findings. Section 3 outlines the design of the experiment that was conducted by Stockley (2018). In section 5.1, I provide a brief overview of the population the experiment was conducted on, followed by a detailed description of my empirical methodology in Section

5.2. Results for both shifts are presented in Section 6. Finally, in Section 7, I present a discussion on the interpretations of the observed estimates and their consistency with the models discussed. Section 8 concludes.

2 Background

For a long time, labour supply was derived from the neoclassical model of the inter-temporal consumption-leisure decision, This model predicts a positive inter-temporal substitution effect on labour supply upon transitory wage increases, giving rise to a positive wage elasticity of labour (under the assumption of negligible income effects). Findings explained by this model tend to estimate the inter-temporal elasticity of substitution to be between 0 and 0.4 (Ghez and Becker (1975), Altonji (1986)). One of the main criticisms of such models is that in most work environments, workers are not free to choose their working hours, and hence the estimates above are attenuated from their true value (Chetty et al. (2011)).

More recent studies have shed light on the power of behavioural models in explaining observed labour supply decisions, particularly within the framework of reference-dependence. These models find their roots in Kahneman and Tversky's "prospect theory" (1979), a model proposed as an alternative to standard expected utility theory to analyse decisions made under uncertainty. The main implication of this structure on preferences is that individuals derive utility from losses and gains of income relative to a reference point, rather than just from levels. Moreover, if subjects exhibit "loss aversion", the marginal disutility from losses is steeper than the marginal utility from gains, making subjects more sensitive to losses (Kahneman and Tversky (1979)).

To address these hypotheses, research emerged from novel samples of workers who were free to choose their hours of labour supply. Oettinger (1999) studied food and beverage vendors in a baseball stadium, where as desired, the workers had a flexible labour supply and more importantly, could predict periods of high and low earnings. Accounting for the endogeneity in vendor wages, which arises from shifts in the demand, they find a substantially positive wage elasticity of supply. Fehr and Goette (2007) conducted a field experiment testing the effect of transitory wage shocks on the labour supply choices of a sample of bike messengers in Zurich, who were doubly flexible in choosing the hours worked per week, as well as the effort per hour. They found that while worker's would choose a greater number of shifts under a temporarily high wage environment, they would optimally choose a lower level of effort per hour. Their preferred explanation of this behaviour uses a model of daily income-targetting, wherein workers offered a higher wage will hit their income target earlier and hence choose a lower equilibrium level of effort. In an attempt to simultaneously link three popular labour supply models - a simple neoclassical model along with two behavioural variants - Andersen et al. (2014) conduct a framed field experiment with vendors in an open air market, introducing manipulations both in anticipated earnings/wages and unexpected windfall gains. Aligned with the neoclassical model and reference-dependent theories, vendors increase their labour supply in the face of anticipated wage increases; but contrary to behavioural models (including KR), they do not respond to unanticipated windfall gains early in the shift.

Camerer et al. (1997) introduced a novel dataset of New York cab drivers, recording the number of shifts worked per day, the number of trips per shift and the per trip time and earnings. The high frequency of the data, along with the flexibility of subjects to choose their working hours, created a unique opportunity to test responses of labour supply to wages and expected earnings. They find a significantly negative wage elasticity of labour supply in their sample, which they explain in the spirit of "prospect theory", assuming that drivers have daily income targets. Drivers work until their target is met, resulting in lower effort on days with higher wages (Camerer et al. (1997)).

Farber (2005) criticised the results of Camerer et al. (1997) for their model being misspecified and biased. In his 2008 extension, Farber generated a structural model with income targets as latent, random variables with driver-specific means, and found a jump in the stopping probabilities at the income target. He still, however, fails to identify sufficient reference-dependence in preferences owing to the smooth relationship between the aggregate realized income and continuation and stopping probabilities. Moreover, the large variation in the average day-to-day targets within drivers implies that there is little consistency in drivers' reference points in the week, rendering the model unrealistic and bearing low predictive power (Farber (2008)).

Kőszegi and Rabin (2006) developed a robust theory of reference-dependent preferences to rationalize findings of income-targetting, by specifying a target-generating process. According to this theory, reference points are endogenously determined by workers' rational expectations of income. Utility is partly derived from the gains and losses relative to these expectation-based targets. By proposing a theory wherein reference points are generated by probabilistic beliefs, this model therefore provides a systematic way of thinking about responses in labour supply to expected and unexpected changes in income.

Abeler et al. (2011) devise a clever experimental design to test the KR theory of expectation-based reference-dependence in the laboratory. They use a lottery-based payment contract which offers, with 50 percent probability, earnings equal to a predetermined fixed payment, and exogenously manipulate the rational expectations of income by treating subjects with either a high or low fixed payment. They find a significantly positive elasticity of labour supply with respect to the fixed payment offered, providing strong evidence in favour of the KR model (Abeler et al. (2011)). Whether these results extend beyond the laboratory, however, is not well-justified.

Crawford and Meng (2011) integrate Farber's income-targetting structural model with the KR model by operationalizing the targets as rational expectations of the drivers (thereby avoiding Farber's critcism of unstable income targets). These point expectations are calculated using carefully chosen driver/day-of-the-week sample averages that limit endogeneity problems between the targets and the outcomes they are used to explain.[1] In their analysis, they emphasize on the heterogeneity of outcomes across drivers that have early earnings that are more or less than expected, claiming that the second target a driver reaches on a particular day has a more significant effect in their stopping decision than the first. This kind of target-reversal phenomenon across the two groups, they claim, is inconsistent with the neoclassical model, but is gracefully characterized by a reference-dependent model (Crawford and Meng (2011)).

Although the different studies do not produce results that are consistent with one another, or with any single model of labour supply, they do collectively present substantial evidence challenging the neoclassical model. Often finding a negative wage elasticity, and effort choices that vary in the expectations of future earnings, the orientation of these results with reference-dependence warrants further examination of its influence on labour supply, particularly in real world settings. Stockley (2018) borrows the experimental design of Abeler et al. (2011) to conduct one of the first real world replications of their experiment, by drawing on a sample of piece-rate daily workers in the garment manufacturing industry in Northeastern Brazil. I re-examine the labour supply behaviour of the subjects in this field experiment, and further present a reduced form estimation of the stopping probabilities of subjects in response to matching their expectations of income and minutes worked[2]. I therefore add to the discussion of the relatively limited body of direct field evidence on reference-dependence, relying on a framed field experiment rather than observational data.

3 Experimental Design

The experiment was conducted in July and August of 2014 by Lisa Stockley and her team in the interiors of Northeast Brazil. The entire experiment constituted two shifts, the second one occurring around 3 week after the first and involving only a sub sample of the Shift 1 subjects. The experimenters visited the homes of the participants and assigned a simple production task. Participants were given a roll of circular stickers, and a deck of 4.25" x 5.5." cards with ten randomly placed black dots. They were then asked to completely cover the black dots with the stickers. Each card they completed counted as one 'task' or one 'unit of output'. This production task was carefully selected to meet several goals:

- Effort was easily observed, generating precise measures of labour supply in output and duration.

[1]This proxy of expectations using the drivers' past earnings is a design that I carry forward in my analysis of Shift 2 outcomes, as is described in Section 5.2.

[2]similar to the methodology of Farber (2008) and Crawford and Meng (2011)

- The characteristics of the production process mimicked day-to-day work involved in the local output market
- The subjects had no practice or expectations around the task other than the information revealed during the experiment
- Since the task was unfamiliar, menial, and unrelated to their actual work, subjects would not mistake it to be adding value for the experimenter or their own skills

In the first segment of both shifts, participants were asked to complete as many cards as they could within a 4-minute window, for which they were immediately rewarded the promised 0.05BRL[3] payment per unit of output. The purpose of this activity was to (1) allow subjects to familiarize themselves with the effort associated with the production task (2) gauge their relative productivities in the unfamiliar task and (3) establish credibility of the experimenters with the payment contracts.

The goal of the two shifts was to manipulate workers' expectations on labour income, and observe the effect of these expectations on labour supply. These manipulations were controlled exogenously by creating a lottery-based payment system. Workers were offered either a high or low wage rate (w_h and w_l, respectively) and a high or low fixed payment (f_h or f_l, respectively). The wage rate represented the piece-rate earnings (per unit of output completed). Unlike the first activity of the shift that instructed them to stop at the 4-minute mark, workers were free to choose when to stop producing output in the main shifts. After the experiment was over, participant $i's$ payment was decided based on the following lottery, determined by a coin-flip:

- with probability p=0.5, they received their accumulated earnings $w_i * e_i$, where w_i is the wage rate offered and e_i is the completed level of output
- with probability p=0.5, they received the fixed payment f_i

Although workers were fully informed of the payment lottery before making their effort choice, the actual payment was only realized *after* they made their optimal labour supply decisions. This uncertainty created probabilistic beliefs about their income. If subjects had KR preferences, then the fixed payment that could potentially be realized would influence these expectations, and hence their effort choice.

In the first shift, workers were randomly assigned a wage rate and fixed payment, from the 2x2 matrix of possible payment contracts, captured in Table 4(a). For this first shift, since subjects were new to such a

[3] Brazilian Real

work/payment structure, expectations of income are assumed to be determined by the salience of the fixed payment.

Approximately three weeks after the first shift, the team of experimenters revisited subjects who were assigned the high fixed payment in Shift 1 and offered them the opportunity to participate in another shift of work. After completing the 4-minute introductory work shift and being paid their accumulated earnings, they were again offered a lottery-based payment contract. However, for this shift, variation was only introduced along the wage dimension. Specifically, this meant that all subjects were paid the high fixed payment (f_h), but wages varied as follows:

- Among those who were offered w_l in Shift 1, half the subjects were now offered the higher wage w_h, i.e. a positive wage shock. The remaining half continued with their Shift 1 contract (f_h, w_l).
- Among those who were offered w_h in Shift 1, half the subjects were now offered the lower wage w_l, i.e. a negative wage shock. The remaining half continued with their Shift 1 contract (f_h, w_h).

The 2x2 matrix of possible Shift 2 payment contracts is outlined in Table 4(b). If past payment contracts were to influence the probabilistic beliefs around income, then we would expect those with a wage shock to choose a different level of effort than those without, for the same payment contract in Shift 2. More generally, if subjects have KR preferences, then their effort choice in Shift 2 would react not only to the present payment contract, but also to their past choices. The inclusion of a second shift of work, in addition to the replication of Abeler et. al's experiment, was instrumental in channeling another source of expectations. Now, other than the fixed payment, the income earned in Shift 1 and past effort choice could also play the role of targets which influence Shift 2 behaviour.

The effects of changes in the beliefs around income, either exogenous (from the fixed payment) or endogenous (from past effort), on the subjects' labour supply decisions are analysed in Section 6.

4 Theoretical Framework

In this section, I put forward the hypotheses that map the predictions of the theoretical model to this experiment. The entire theoretical framework is built on the assumption of utility maximizing behaviour under the model of Kőszegi and Rabin (2006), as introduced in Section 2. Reference-dependence implies that individuals value gains and losses relative to a rationally-generated target, in addition to levels of income. At the same time, loss aversion construes losses to be more heavily weighted in utility than gains. This reflects as a high marginal utility of income just before the target, followed by a discontinuous drop as soon as the target is met (Kahneman and Tversky (1979)). In the context of Shift 1 of this experiment, subjects do not know which payment state will be realized, but want to minimize losses in the realization of either state. One of the possible payments is fixed (and known) before the start of production, and hence acts as a benchmark for the other payment, which is decided by individuals' optimal effort choice. Therefore, losses are avoided only if subjects work to the point that equates their accumulated earnings to the fixed payment offered (or whatever expectation of income is assumed). For Shift 2, while payment is again determined by a lottery, subjects may have adapted their targets of income (and time) to the outcomes of Shift 1. Insofar as their choice of labour determines part of the realized Shift 2 income (with the remainder dependent on the exogenous fixed payment), we may assume that their labour supply reflects the desire to meet the new income (and time) target. Therefore, under the assumption that subjects in the experiment rationally target expectation-based reference points, the model predicts that individuals optimally seek to equate their accumulated earnings to the income target.

Under the design of Shift 1, where the expectations that generate the target are determined by the fixed payment, with regards to the level of output I predict that:

Hypothesis 1: Individuals who are offered the higher fixed payment work more than those offered the lower fixed payment, while controlling for differences in the piece-rate wage.

Relatedly, the discontinuity in the the marginal utility of working just before and after the target motivates the following conjecture about the probability of stopping:

Hypothesis 2: A subject's marginal probability of stopping jumps when they reach their income (or minutes) target.

The model of the optimal stopping decision is applied to both Shift 1 and Shift 2, but the target(s) of the subjects is assumed to vary based on the beliefs at the start of the respective shifts. Lastly, Crawford and Meng (2011) observe heterogeneity in the influence of income and minutes targets across subsamples, depending on whether their earnings early in the work shift are more or less than expected. This distinction determines the chronological order of reaching the two targets in the process of production. Put informally, they also propose that these targets are operationalized by an individual's past average of income earned (and time worked) in previous comparable work shifts. To describe one such scenario, consider a subject who formulated an income and time target based entirely on their Shift 1 outcomes, before the start of Shift 2. If their wage rate were to stay the same, they would take the same amount of time to accumulate the same total level of income, i.e. the income and time targets will be met simultaneously. However, if on commencing production we introduce a higher piece-rate wage, then they need to work *less* to achieve the same income as at the lower wage rate. Alternatively, if they work as long as they did with the lower wage rate, they would end up accumulating greater earnings than their income 'target'. To explain behaviour consistent with their findings, Crawford and Meng propose a modification to the optimal stopping decision, according to which subjects are significantly influenced by the second target they reach during the shift, and not by the first.

If such a utility-generating process were to be assumed, then subjects' stopping decision would align with the following hypothesis in the context of the second shift of this experiment:

Hypothesis 3: The marginal probability of stopping of subjects who are offered a positive wage shock is significantly affected by their minutes target, but varies smoothly over their income target. The marginal probability of stopping of subjects who are offered a negative wage shock is significantly affected by their income target, but varies smoothly over their time target.

5 Data and Methodology

5.1 Data

This data is obtained from a framed field experiment conducted in the municipalities of Bezerros, Caruaru, and São Caitano, of the state Pernambuco in Northeast Brazil. Details on the demographics and working conditions of this subpopulation can be found in Stockley's dissertation (Stockley (2018)). An important feature of this population was that they were engaged in piece-rate unsupervised garment labour for a living. Their working hours were flexible, allowing them to substitute away from labour depending on the opportunity cost of their time and choose when to quit. Moreover, their work was largely unskilled, manual and repetitive in nature. The nature of the production process involved in the experiment was therefore familiar to the subjects, although involving a new task to be performed.

For this experiment, 366 participants from 43 neighbourhood clusters were recruited from a baseline survey for the first shift. The distribution of participants across the four possible payment contracts is displayed at the bottom of the left panel of Table 2. A sub sample of 239 of these 366 participants returned for the second shift. The distribution of these returning participants across the four possible payment contracts is also displayed (at the bottom of the right panel) in Table 2. The rest of Table 2 contains self-reported characteristics of this sample at the time of the first experiment. Overall, sample characteristics appear well balanced across all the 8 treatment cells. The average participant in the experiment was a female aged around 40 years.

In the practice activity of each shift, subjects were given 4 minutes to complete as many cards as they wished, to assess their relative productivities. For the first practice shift, the median participant produced 1.5 cards per minute, with the fastest completing 3.2 cards per minute and the slowest completing 0.5 cards per minute. For the second shift, participants were statistically faster, suggesting some degree of learning-by-doing from Shift 1 to Shift 2. The median participant of the second practice shift produced 1.75 cards per minute, with the fastest completing 3.75 cards per minute and the slowest completing 0.25 cards per minute.

During the main open-ended shift of work, subject were free to choose their effort, in both time worked and cards completed. In Shift 1, the mean output produced by participants was 19.97 cards, working 12.5 minutes on average. In Shift 2, the average labour supply was lower, with participants completing 16.64 cards on average and working for a mean of 9.4 minutes. Since we are interested in the difference in labour supply across payment contracts, Tables 1(a) and 1(b) present the unconditional mean (and standard deviation) of cards completed, while Tables 1(c) and 1(d) present the unconditional mean (and standard deviation) of duration worked, for the different treatment assignments in the first and second shift, respectively.

5.2 Methodology

In this section, I discuss the empirical models used to identify the effect of expectation-based reference points on the optimal labour supply decision of participants.

5.2.1 Shift 1 Analysis

In Shift 1, subjects' expectations were manipulated by offering a HIGH or LOW fixed payment in the payment lottery contract.

Before running the model for the optimal stopping decision, I conduct a simple regression analysis to estimate the effect of offering a higher fixed payment and higher wage on effort. The regression equation is:

$$ln(e_i) = \alpha + \beta_f * f_i^{\text{HIGH}} + \beta_w * w_i^{\text{HIGH}} + \beta_{fw} * f_i^{\text{HIGH}} * w_i^{\text{HIGH}} + \gamma * X_i + \epsilon_i \tag{1}$$

Here $ln(e_i)$ is the log of the output (no. of cards) produced produced by individual i. f_i^{HIGH} is a binary variable indicating whether the subject was offered the higher fixed payment. w_i^{HIGH} is a binary variable indicating whether the subject was offered the higher wage. X_i is a list of covariates of individual i. The neoclassical model predicts that f_i^{HIGH} should have no effect on labour supply (since it enters the optimal labour supply decision as a constant in expected income). Moreover, negative wage elasticity of labour supply would require labour supply to be increasing in w_i^{HIGH} (by the substitution effect, the opportunity cost of leisure increases as the wage offered increases). Results of this estimation are presented in Table 4.

Next, I run a regression specification that most closely resembles Abeler et al.'s identification strategy. The following equation aims to estimate *only* the effect of varying the fixed payment on output produced, by restricting the sample to those offered the high wage.

$$ln(e_i) = \alpha + \beta_f * f_i^{\text{HIGH}} + \gamma * X_i + \epsilon_i \Big|_{w_i^{\text{HIGH}}=1} \tag{2}$$

If subjects were to exhibit preferences in accordance with the KR model of reference-dependent utility, reference points should be derived from the expectations of income. In this shift, these expectations are manipulated by changing the fixed payment offered in the lottery. Therefore, we would expect those with the higher fixed payment ($f_i^{\text{HIGH}} = 1$) to produce more output than those with the lower fixed payment ($f_i^{\text{HIGH}} = 0$)), given the same wage rate. Results of this specification are presented in Table 3.

Double Machine Learning Estimates: To improve precision of the estimates on the treatment variables of the regressions presented above, I employ the Double Machine Learning (DML) algorithm as an extension of my analysis. DML is useful in a randomized control experiment such as this, when there exist high-dimensional covariates that are correlated with the outcome (effort in producing output) but uncorrelated with the treatment. The algorithm creates an optimal vector of covariates that reduces the standard error of the OLS estimates without introducing bias, thereby increasing precision. Tables 4 and 3 present the results for OLS and DML estimates of the coefficients in equations (1) and (2), respectively.

Finally, I present a stylized model to estimate the optimal stopping decision of the subjects as a binary variable:

$$\begin{aligned} \phi_{ih}^{\text{STOP}} = \alpha + \beta_f * f_i^{\text{HIGH}} + \beta_y * \text{cum\_income}_{ih} + \beta_t * \text{cum\_mins\_worked}_{ih} + \\ \delta * (\text{cum\_income}_{ih} \geq \text{income\_target}_i) + \gamma * X_i + \epsilon_i \end{aligned} \tag{3}$$

Here ϕ_{ih}^{STOP} is a binary variable that is equal to 1 at the last unit produced by individual i, i.e. if they stop producing after h units of output. f_i^{HIGH} is a binary variable indicating whether the subject was offered the higher fixed payment, as before. The variable cum_income_{ih} is a continuous variable that captures earnings accumulated by i after producing the h^{th} unit of output. Similarly, $\text{cum\_mins\_worked}_{ih}$ is a continuous variable that captures the total minutes worked by i after producing the h^{th} unit of output. Since the unit of observation in the data was at an individual rather than individual-output level, I interpolated the minutes worked per unit of output by assuming subjects had a constant productivity (equal to their average productivity) during the shift. Therefore, the variable $\text{cum\_mins\_worked}_{ih}$ is constructed as

$$\text{cum\_mins\_worked}_{ih} = \text{avg\_productivity}_i * h$$

$$\left(\text{avg\_productivity}_i = \text{total\_mins}_i / \text{total\_output}_i\right)$$

where $\text{avg\_productivity}_i$ measures the minutes taken per card by i in Shift 1.

Lastly, the variable $(\text{cum\_income}_{ih} \geq \text{income\_target}_i)$ is a dummy variable indicating whether individual i has hit their income target after producing h unit of output. In this case, the income target is just the fixed payment f_i offered to individual i.

Following the same idea as in the simple regression estimation, the KR model predicts the probability of stopping should jump (increase) once the income target is met. Moreover, it follows from the diminishing marginal utility of leisure (increasing marginal cost of effort) that the probability of stopping be positively related to both cumulative income and minutes worked.

The left panel of Table 6 first presents the marginal effects on the probability of stopping from a probit regression, at an income evaluation point of 3 BRL (the value of the lower fixed payment). Those offered the lower fixed payment should have a higher probability of stopping at this level of income, compared to those with a higher income target. The right panel of the same table presents the marginal effects at the median level of accumulated earnings (2 BRL) and minutes worked (19.5 mins) in Shift 1.[4] Table 5 presents linear probability estimates of the same equation.

5.2.2 Shift 2 Analysis

The analysis of Shift 2 rests on the assumption that the return of a fraction of subjects in another shift of the same task generated new probabilistic beliefs about the income and effort targets. I borrow from the econometric methodology of Farber (2008) and Crawford and Meng (2011) in this section, treating drivers' targets as rational expectations operationalized by their previous shift performance. Specifically, the recent income earned and duration worked in Shift 1 formed adapted targets, that served as reference points for the Shift 2 labour supply decision.

To determine the effect of these new expectation proxies on the optimal stopping decision, I estimate the following model:

[4] These estimates are robust to using different evaluation points, including the 25^{th} and 75^{th} percentiles of the total accumulated earnings and total minutes worked across all individuals in Shift 1.

$$\phi_{ih}^{\text{STOP}} = \alpha + \beta_y \text{cum\_income}_{ih} + \beta_t \text{cum\_mins\_worked}_{ih} + \delta_y^{\text{ACCUM}}(\text{cum\_income}_{ih} \geq \text{income}^{\text{ACCUM}}\text{target}_i) + \delta_y^{\text{PAID}}(\text{cum\_income}_{ih} \geq \text{income}^{\text{PAID}}\text{target}_i) + \delta_t(\text{cum\_mins\_worked}_{ih} \geq \text{mins\_target}_i) + \gamma X_i + \epsilon_i \tag{4}$$

Since all subjects were offered the same higher fixed payment in this shift, the variable f_i^{HIGH} is no longer relevant to the analysis. Variables cum_income$_{ih}$ and cum_mins_worked$_{ih}$ have the same interpretation as before, now looking at performance in Shift 2 instead. The variable (cum_income$_{ih} \geq$ income$^{\text{ACCUM}}$target$_i$) is a dummy to indicate whether the earnings accumulated by output h exceeds the piece-rate earnings accumulated in Shift 1 work. The variable (cum_income$_{ih} \geq$ income$^{\text{PAID}}$target$_i$), on the other hand, is a dummy to indicate whether the earnings accumulated by output h exceeds the payment actually received in Shift 1. Note that since Shift 1 payment was decided by a lottery, this payment was the same as the accumulated earnings for some subjects but equal to the fixed payment for others. Lastly, the variable (cum_mins_worked$_{ih} \geq$ mins_target$_i$) is a dummy that indicates whether individual i has met/exceeded the total time worked in Shift 1, after producing output h.

Another feature of the experimental design of Shift 2 was that returning subjects may have been exposed to a wage shock from Shift 1. Continuing to follow the methodology of Crawford and Meng (2011), I therefore incorporate a split-sample analysis for Shift 2, separately estimating the stopping probabilities generated in equation (4) for individuals with a positive wage shock and negative wage shock. These estimations are represented by the equations below.

$$\phi_{ih}^{\text{STOP}} = \alpha + \beta_y \text{cum\_income}_{ih} + \beta_t \text{cum\_mins\_worked}_{ih} + \delta_y^{\text{ACCUM}}(\text{cum\_income}_{ih} \geq \text{income}^{\text{ACCUM}}\text{target}_i) + \delta_y^{\text{PAID}}(\text{cum\_income}_{ih} \geq \text{income}^{\text{PAID}}\text{target}_i) + \delta_t(\text{cum\_mins\_worked}_{ih} \geq \text{mins\_target}_i) + \gamma X_i + \epsilon_i \Big|_{wage_i^{SHIFT\_2} > wage_i^{SHIFT\_1}} \tag{5}$$

$$\phi_{ih}^{\text{STOP}} = \alpha + \beta_y \text{cum\_income}_{ih} + \beta_t \text{cum\_mins\_worked}_{ih} + \delta_y^{\text{ACCUM}}(\text{cum\_income}_{ih} \geq \text{income}^{\text{ACCUM}}\text{target}_i) + \delta_y^{\text{PAID}}(\text{cum\_income}_{ih} \geq \text{income}^{\text{PAID}}\text{target}_i) + \delta_t(\text{cum\_mins\_worked}_{ih} \geq \text{mins\_target}_i) + \gamma X_i + \epsilon_i \Big|_{wage_i^{SHIFT\_2} < wage_i^{SHIFT\_1}} \tag{6}$$

Table 9 displays the probit estimation results of equation (4) for the full sample of returning subjects, while Table 10 estimates the same for the split-sample. For completeness, estimates for those given no wage shock are presented alongside the subsamples with positive and negative wage shocks. Marginal effects on the probability of stopping are reported at (an evaluation point of) the median level of income accumulated and minutes worked across all subjects in Shift 2[5]. Table 7 and 8 display the linear stopping probability estimates for the same specifications.

6 Results

6.1 Shift 1

Table 3 presents estimates for the attempted replication of Abeler et al.'s experiment in this setting of Brazilian piece-rate workers. Restricting the sample to those subjects that received the higher wage of 0.02 BRL/card in Shift 1 permits variation only along the fixed payment dimension. I find the elasticity of labour supply with respect to the fixed payment to be negative and large. The DML results are even statistically significant the 5 percent level, estimating subjects who receive the higher fixed payment to be producing roughly 50 percent *less* output than those with the lower treatment, after controlling for fixed effects and other productivity measures. Decisively opposing Abeler et al.'s evidence, the implications of these surprising results are analysed in Section 7.

Since there was a 2x2 assignment of treatment based on wages and fixed payment, my next analysis (based on equation (1)) includes the effect of this variation in wages, and is presented in Table 4. The negative direction on the coefficient for the higher wage treatment suggests that being offered a higher wage induced workers to work less, on average. What is commonly interpreted as an income effect, this negative effect of higher wages was stronger for the group of subjects in the higher fixed payment treatment, as measured by the interaction term. On the whole, however, all coefficients are imprecise, preventing us from concluding that either the higher wage or the higher fixed payment have an overall significant effect on the level of output subjects optimally choose to produce.

The final outlook on Shift 1 outcomes from a model of an optimal stopping decision as a function of the income target as well as levels of income and time worked. Table 6 presents the marginal effects on the probability of stopping, at two distinct evaluation points. Since the estimates are nearly identical across different choices of evaluation points, I discuss only the results from the panel on the left. Columns (1) and (2) present the coefficients on cumulative income and cumulative minutes worked alone, before and after introducing age, sex, and productivity controls and interviewer, time of day and weekday fixed effects, respectively. Column (3) includes a dummy for the fixed payment offered, which indicates that those who

[5]These estimates are robust to using different evaluation points, including the 25^{th} and 75^{th} percentiles of the total accumulated earnings and total minutes worked across all individuals in Shift 2.

received the higher fixed payment in the lottery are less likely to quit working at the lower fixed payment mark of 3 BRL. Though this result is imprecisely estimated, the weakly positive direction does support the prediction of the KR model.

Column (4) presents the estimation of equation (3). The only conclusive result from this specification is that subjects who have worked for longer have a higher marginal probability of stopping production. The negative and statistically significant coefficient on cumulative minutes worked throughout all specifications, says that once they accumulate 3 BRL of income, subjects who work an extra minute have a 0.12 percentage point higher probability of quitting, on average. This result can be explained by convexity of the cost of effort, i.e. marginal effort costs that are increasing in the number of cards produced. Cumulative income seems to have no significant effect on the probability of stopping. More surprisingly, we would expect (based on the KR model) that those offered the higher fixed payment would be less likely to quit at an income of 3 BRL (which was the lower fixed payment amount); but the near zero coefficient on the HI treatment suggests that being offered the higher fixed payment also did not systematically influence the probability of stopping. Lastly, the positive direction of the coefficient on the income target dummy weakly supports the predictions of KR, but that lack of significance on this result means we can not conclude that meeting the income target had any overall effect. An important distinction here is that the interpretation of results on marginal stopping probabilities are not directly comparable the results of Table 4 which suggests a decrease in the *average level of output* for subjects in the HI treatment. The linear probability estimates in Table 5 maintain the same results as the probit estimates, with slightly lower significance.

6.2 Shift 2

Results from the probit estimation of the optimal stopping decision in Shift 2 are detailed in Table 9. All estimates represent the marginal probability effects at an evaluation point of the median of total income accumulated (2 BRL) and duration (7.4 minutes) among Shift 2 participants.

Table 9 obtains results for the full returning sample, while Table 10 obtains the same estimates for the split-sample. The coefficients on cumulative income and cumulative minutes worked follows an identical interpretation to the Shift 1 results, discussed in Section 6.1. Columns (3) and (4) of Table 9 estimate equation (4), but for the two defined income targets individually. Based on the estimations in column (3), subjects who match their income *accumulated* in the previous shift are 1.5 percentage points more likely to stop working, but this result lacks statistical significance. Column (4) estimates that subjects who match their income *received* in the previous shift have higher probability of quitting production, by around 4.2 percentage points on average, while controlling for whether they received the higher of the two possible payments. This is a result of high statistical and economic significance, given that the mean probability of quitting at the evaluation point was around 5.8 percent. Both estimations also suggest that those who hit their minutes target, determined by the total duration worked in Shift 1, are statistically more likely to quit by around 3.8 percentage points than those who have not yet met the target.

Columns (1) and (2) of Table 10 are identical estimations as columns (3) and (4) of Table 9, but restricted to the subsample of individuals who received the lower wage in Shift 1 but the higher wage in Shift 2 (positive wage shock), all else the same. For these individuals, the effect of both targets, income accumulated and income paid in Shift 1, have a significant positive effect on the probability of stopping work at the median evaluation point, of 3.75 and 4.17 percentage points respectively. Columns (3) and (4) are instead restricting the estimates to the subsample of individuals who received the higher wage in Shift 1 but the lower wage in Shift 2 (negative wage shock). The estimates on both the income targets are the largest for this sub sample of subjects. Matching the income accumulated in Shift 1 increases the probability of stopping by approximately 10.7 percentage points, which is significant at the 10 percent level. Yet more striking, is the estimate on matching the income paid in Shift 1, which leads to an increase in the marginal probability of stopping by 18.87 percentage points with significance at the 1 percent level. Additionally, for both groups with wage shocks, the coefficient on the minutes target is weakly positive and imprecisely estimated. Finally, columns (5) and (6) of table 10 present the same estimates for the sub sample of participants that were given no wage shock. These participants essentially repeated another shift identical to Shift 1. The estimate on the target of income accumulated is much smaller and insignificant for this group. The minutes target, however, is significant and suggests that subjects who have worked the same amount of time as Shift 1 are on average 4.5 percentage points more likely to quit.

The linear stopping probability estimates, which estimate the average marginal probability of stopping over the entire shift (rather than at an evaluation point), provide results in the same direction but with weaker precision. These results are presented in Table 7 and Table 8 for the full sample and split sample, respectively.

7 Discussion

Hypothesis 1: Fixed payments as Rational Expectations of Income:

If subjects aligned with KR preferences, they would attempt to reduce the disparity between the two states of payment in the lottery-based payment contract (using the fixed payment as a reference point). As a result, they would rationally choose a labour supply level that equates their accumulated earnings with the fixed payment offered. Corroborating this theory, in their laboratory experiment Abeler et. al found that subjects offered a higher fixed payment produce statistically more output than those offered the lower fixed payment.

In the attempted replication presented in Table 3, no concurrent results are no observed. Instead, the negative and significant effect of the treatment provides strong evidence in favour of an income effect. If subjects treat the fixed payment offered as potential income, those assigned to the higher payment treatment would have a higher expected income. The neoclassical utility model would predict that those individuals with a stronger income effect view themselves as richer and hence work less. This is what we observe with significance in our sample.

Can we conclude that subjects do not have reference-dependent preferences? Not necessarily. The evidence in favour of the income effect does not allow us to eliminate the presence of some degree of reference-dependence in the labour supply decision. Rather, I interpret the results as supporting any one of the following possibilities:

(1) Subjects do not exhibit reference-dependent preferences.

(2) *If* subjects do have reference-dependent preferences, the fixed payment offered in this experimental design did not systematically influence their target.

(3) Subjects exhibit reference-dependence, the fixed payment influenced the income target, but the neoclassical income effect dominates this reference-dependence, and hence determines the overall direction of labour supply.

Canonical labour supply models, however, also propose that transitory wage shocks or changes in expected income that are insignificant compared to overall lifetime earnings should have a negligible income effect on labour supply. In this pool of subjects, the strong income effect that dominates potential substitution and reference-dependence effects challenges this assumption, and aligns better with the behavioural theory of "narrow bracketing" (Rabin and Weizsäcker (2009)). Based on this theory, subjects do not aggregate different sources of income into one account of 'lifetime earnings'; rather, they view the income earned from the experiment as isolated from other sources of income. Therefore, any changes to the expected income in the experiment are no longer negligible in the labour supply decision within the experiment, and will have a significant influence on their optimal effort choice.

The existence of a dominant income effect is further evidenced by the negative elasticity of output with respect to the wage rate, when extending the analysis to the entire Shift 1 sample (not just those receiving the higher wage). All in all, the results of Shift 1 provide strong evidence of an income effect in favour of narrow bracketing, rejecting the hypothesis that the fixed payment manipulates reference-points to significantly affect the labour supply decision.

Hypothesis 2: Past Experience as Expectations of Income/Minutes Targets

For this analysis, the income earned/realized and time worked during Shift 1 proxy the expectations of subjects in their reference-dependent preferences during Shift 2. To put these results into perspective, a neoclassical model would predict output produced to vary smoothly with the income accumulated, with no regard to whether it is higher or lower than the expected level. Farber's (2008) pure income-targetting model would predict a jump in the probability of stopping once the rational income target is met, but a linear relation with time worked. My results for the full sample seem to rationalize a model where both the income and minute targets are significant in the stopping decision. As far as the income target is concerned, reaching the level of income that was actually *received* in Shift 1 led to a higher probability of stopping. Engaging in a repeated experiment roughly three weeks after the first, earnings that subjects ultimately took home appear to have greater salience in forming rational expectations of income in the second experiment, as opposed to the level of income accumulated through production. An interesting implication of this result is that although the fixed payment did not induce strong reference-dependence in the first shift, subjects may have come to view it as a target the second time the experiment was conducted (insofar as they were paid the fixed payment in Shift 1). If this is true, it admits the possibility of a learning effect in the setting of rational expectation-based targets. While the worker sample in Brazil may not have assumed immediate consideration of this fixed payment in their expectations, the distinct sample of Abeler et al., used to participation in laboratory experiments with exogenous manipulations, may have demonstrated a faster uptake of such a target.

The level of income accumulated upto a certain point, similar to Crawford and Meng's results, does not have a significant influence on the stopping decision. The significant, positive marginal effect of cumulative time worked on the stopping probability is rationally explained by convexity associated with the cost of effort. The statistical significance of minutes worked, both in its linear form and as a target, suggests that time has both a strong neoclassical effect as well as a reference-dependent effect on labour supply decisions.

Hypothesis 3: The Two-Target Hypothesis of Crawford and Meng

Turning to the split-sample analysis, Crawford and Meng (2011) predict that the second target an individual reaches in the production process plays a significant role in their stopping decision, not their first. Subjects who enjoyed a positive wage shock are more likely to reach the level of income earned in Shift 1 before working for as long as they did. On the other hand, subjects who suffered a negative wage shock would likely exceed the time worked in Shift 1 before they reach the same income level. Therefore, they predict a target-reversal phenomenon where the significance switches from the time target to the income target, in going from a positive to negative wage shock.

My results testing these hypotheses are in general inconsistent with predictions, and are not robust to whether the estimation assumes a linear or probit model. Based on the linear probability estimates, reference-dependence on the income paid during Shift 1 is evidenced only by the subsample experiencing a negative wage shock. The effect of reaching the time target is significant for the negative and no shock treatments, but the proximity of the point estimates across all groups suggests this may just be an issue of precision. Based on the probit estimates, the stopping probability of both the positive and negative treatment groups increases on hitting their income targets (both accumulated and realized Shift 1 income). Only the group experiencing no wage shock responds to hitting the minutes target, which again may be a result of lack of precision.

Although the prediction of significance of the second target is consistent for the negative wage shock subsample in this experiment, there is no definitive pattern of target reversal between the positive and negative wage shock treatment groups. In general, therefore, I do not interpret these results as conclusive of Crawford and Meng's hypotheses.

There is, however, a unique result that emerges from the split-sample analysis. The income targetting effect, as discussed under *Hypothesis 2*, seems to be markedly stronger for the subsample of workers who experienced a negative wage shock. While a robust psychological analysis of such behaviour is beyond the scope of this paper, it roughly ties into behavioural theories of mental accounting, observing that individuals tend to aggregate losses into the same mental account (Thaler (1985)). Similar to how individuals who lose bets have a tendency to continue gambling until they reverse their fortune[6], it seems that workers who perceive themselves as starting out behind the curve (faced with a negative wage shock), levy greater importance on meeting their income target.

8 Conclusion

This paper analysed data obtained from a framed field experiment in a novel setting of workers engaged in production that pays at a piece-rate wage. It aimed to estimate the effect of income (and time) targetting on the labour supply decision, where the targets are born out of the rational probabilistic beliefs of individuals. The first experiment exogenously manipulated these expectations but setting a fixed payment that could have been earned with a 50 percent chance, but the results find that responses of workers in this sample to variation in the exogenous payment contradicts KR predictions. They instead seem to be dominantly influenced by an income effect, working less when offered the higher fixed payment in their lottery-based

[6]this was an example presented in Thaler (1985).

contract. This result endorses workers who follow a “narrow bracketing” mental accounting framework when making these labour supply decisions. It may, however, be worthwhile to note that the optimal effort choices observed pertain to a repetitive, menial, and blatantly ‘unproductive’ task, which may not extend to a worker making labour supply decisions in less arbitrary environments, such as their primary occupation.

Illuminating the overall results of both shifts, we can not reject the presence of reference-dependence in preferences; however, the expectation of income is more significantly influenced by one’s past income than the fixed payment offered. Workers’ optimal stopping probability appears to jump once they match the income received in their previous shift. An interesting upshot of the split-sample analysis of the second shift reveals that when workers are faced with an unexpected wage cut, the income target assumes an even stronger influence on their labour supply decision. Results are also obtained on the variation of labour supply along the dimension of time; I find that time worked plays an important part in the labour supply story by entering the optimal stopping decision both in levels (of total minutes worked at a given level of output) and as a reference-point (working as long as one did in the previous shift).

In conclusion, the results are not wholly consistent with any single labour supply model, with components of neoclassical utility (income effect), as well as reference-dependence where the reference points arise from adaptive expectations based on past experience. As clear as laboratory results such as Abeler et al’s may seem, real world labour supply decisions assume a more complex nature with a variety of exogenous and endogenous influences that can not always be captured by a model. This paper contributes to our understanding of behavioural labour supply models by drawing results from a low-income sample in a developing nation, and challenges the external validity of the seemingly robust results obtained from laboratories and other WEIRD[7] settings.

9 Tables

Table 1(a): Output Produced in Shift 1, by Payment Contract

	HIGH Fixed Payment	**LOW Fixed Payment**
HIGH Wage	18.8 (18.75)	19.2 (15.4)
LOW Wage	21.9 (22.4)	19.1 (19.7)

Note: Table reports the unconditional mean (and standard deviation) of the cards completed in each treatment cell of the first shift.

[7] White, Educated, Industrialized, Rich, and Democratic.

Table 1(b): Output Produced in Shift 2, by Payment Contract

	HIGH Shift 2 Wage	LOW Shift 2 Wage
HIGH Shift 1 Wage	18.0 (15.5)	15.6 (13.7)
LOW Shift 1 Wage	16.7 (12.1)	16.3 (13.2)

Note: Table reports the unconditional mean (and standard deviation) of the cards completed in each treatment cell of the second shift.

Table 1(c): Minutes Worked in Shift 1, by Payment Contract

	HIGH Fixed Payment	LOW Fixed Payment
HIGH Wage	11.8 (10.0)	12.1 (8.6)
LOW Wage	14.0 (12.5)	11.2 (9.2)

Note: Table reports the unconditional mean (and standard deviation) of the minutes worked in each treatment cell of the first shift.

Table 1(d): Minutes Worked in Shift 2, by Payment Contract

	HIGH Shift 2 Wage	LOW Shift 2 Wage
HIGH Shift 1 Wage	10.0 (7.8)	8.6 (6.5)
LOW Shift 1 Wage	9.4 (5.8)	9.6 (7.1)

Note: Table reports the unconditional mean (and standard deviation) of the minutes worked in each treatment cell of the second shift.

Table 2: Self-Reported Sample Characteristics, by Treatment

	First Shift Sample				Second Shift Sample			
	(1) Low Wage-High F	(2) Low Wage-Low F	(3) High Wage-High F	(4) High Wage-Low F	(5) Low Wage-No Shock	(6) Low Wage-Positive Shock	(7) High Wage-No Shock	(8) High Wage-Negative Shock
Age	43.11	39.65	42.64	42.54	41.61	44.65	42.36	43.77
	[18.52]	[15.77]	[15.41]	[15.63]	[20.31]	[17.04]	[15.64]	[15.18]
Male	0.38	0.47	0.44	0.39	0.34	0.42	0.48	0.42
	[0.49]	[0.50]	[0.50]	[0.49]	[0.48]	[0.50]	[0.50]	[0.50]
Employed	0.34	0.50	0.41	0.38	0.30	0.37	0.40	0.38
	[0.48]	[0.50]	[0.49]	[0.49]	[0.46]	[0.49]	[0.49]	[0.49]
- works from home	0.38	0.26	0.47	0.29	0.35	0.41	0.48	0.40
	[0.49]	[0.44]	[0.50]	[0.46]	[0.49]	[0.50]	[0.51]	[0.50]
- works for a piece-rate	0.41	0.52	0.38	0.35	0.38	0.42	0.32	0.39
	[0.50]	[0.51]	[0.49]	[0.49]	[0.50]	[0.51]	[0.48]	[0.50]
Observations	125	62	120	59	61	60	58	60

Note: The first four columns (left) review the first shift, while the last four (right) review the second shift sample. Columns are based on the assignment of different treatments in each shift. Reported values are the mean (sd) of self-reported characterstics at the time of the first shift.

Table 3: Replicating Abeler et. al: OLS and DDML estimates

	Dependent Variable: Log(Cards Produced)				
	OLS estimates			DDML Estimates	
	(1)	(2)	(3)	(4)	(5)
HI treatment	-0.205	-0.480	-0.487*	-0.489**	-0.539**
	(0.356)	(0.303)	(0.250)	(0.193)	(0.215)
Productivity		1.150***	1.372***		
		(0.359)	(0.329)		
Interviwer FE's	No	No	Yes	No	Yes
Time of Day FE's	No	No	Yes	No	Yes
Weekday FE's	No	No	Yes	No	Yes
Observations	103	103	103	103	103
Mean of dep. var.	2.538	2.538	2.538	2.580	2.580

Note: Standard errors are clustered at the neighbourhood level. Estimates are for subjects offered the higher wage (0.2 BRL/card) only. Coefficient estimates statistically significant at * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels, respectively. DDML potential controls are a list of covariates collected from the participants. For the DML estimates, colummns (4) and (5) impose Productivity as a necessary control, while column (5) also imposes the fixed effects as necessary controls, allowing flexibility to choose among all other covariates.

Table 4: Effect of Treatment on Cards Produced: OLS and DDML estimates

	Dependent Variable: Log(Cards Produced)				
	OLS estimates			DDML Estimates	
	(1)	(2)	(3)	(4)	(5)
HI treatment	-0.089	-0.053	-0.292	-0.212	-0.143
	(0.311)	(0.301)	(0.398)	(0.225)	(0.244)
High Wage	0.099	0.342	-0.034	0.110	-0.051
	(0.345)	(0.305)	(0.295)	(0.182)	(0.205)
HI treatment X High Wage	-0.116	-0.493	-0.102	-0.269	-0.220
	(0.454)	(0.407)	(0.409)	(0.294)	(0.309)
Productivity		1.424***	1.684***		
		(0.395)	(0.405)		
Interviwer FE's	No	No	Yes	No	Yes
Time of Day FE's	No	No	Yes	No	Yes
Weekday FE's	No	No	Yes	No	Yes
Observations	227	227	227	227	227
Mean of dep. var.	2.479	2.479	2.479	2.544	2.544

Note: Standard errors are clustered at the neighbourhood level. Coefficient estimates statistically significant at * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels, respectively. Variable HI treatment" is a dummy that indicates whether the subject was offered the higher fixed payement (6 BRL). Variable High Wage is a dummy that indicates whether the subject was offered the higher wage (0.2 BRL/card). DDML potential controls are a list of covariates collected from the participants. For the DML estimates, colummns (4) and (5) impose Productivity as a necessary control, while column (5) also imposes the fixed effects as necessary controls, allowing flexibility to choose among all other covariates."

Table 5: Probability of Stopping: Shift 1 (Linear Probability Estimates)

	Dependent variable: Binary Stopping Decision Full Sample			
	(1)	(2)	(3)	(4)
High Fixed Payment Treatment		-0.0100	-0.0033	-0.0005
		(0.0091)	(0.0109)	(0.0115)
Cumulative Income	-0.0047***	-0.0048***	-0.0006	-0.0018
	(0.0014)	(0.0014)	(0.0019)	(0.0020)
Cumulative Mins. Worked	0.0018***	0.0019***	0.0013*	0.0013*
	(0.0007)	(0.0007)	(0.0007)	(0.0008)
Cum. Income > Income Target				0.0130
				(0.0126)
Productivity			-0.0611***	-0.0613***
			(0.0182)	(0.0182)
Controls (λ)	No	No	Yes	Yes
Interviwer FE's	No	No	Yes	Yes
Time of Day FE's	No	No	Yes	Yes
Weekday FE's	No	No	Yes	Yes
Observations	4,703	4,703	4,703	4,703
Mean of dep. var.	0.0474	0.0474	0.0474	0.0474

Note: Coefficient estimates and standard errors from OLS linear probability estimates are presented; standard errors are clustered at the neighbourhood level. Coefficient estimates statistically significant at * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels, respectively. The 'Income Target' is set as the fixed payment offered, according to the treatment group. Controls are the age and sex of the participant.

Table 6: Marginal Effects: Shift 1 (Probit Estimates)

	Dependent variable: Binary Stopping Decision							
	Income Evaluation Point: LO fixed payment [3]				Evaluation Points: Median Income [2] and Time [9.5]			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High Fixed Payment Treatment			-0.0028	0.0001			-0.0027	0.0001
			(0.0097)	(0.0103)			(0.0092)	(0.0099)
Cumulative Income	-0.0044***	0.0002	0.0002	-0.0010	-0.0044***	0.0002	0.0002	-0.0010
	(0.0014)	(0.0015)	(0.0015)	(0.0015)	(0.0014)	(0.0014)	(0.0014)	(0.0014)
Cumulative Mins. Worked	0.0016***	0.0012**	0.0012**	0.0012**	0.0016***	0.0011**	0.0011**	0.0011**
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Cum. Income > Income Target				0.0131				0.0126
				(0.0120)				(0.0117)
Controls (λ)	No	Yes	Yes	No	Yes	Yes	Yes	
Interviwer FE's	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Time of Day FE's	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Weekday FE's	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	4,703	4,703	4,703	4,703	4,703	4,703	4,703	4,703
Mean of dep. var.	0.0474	0.0474	0.0474	0.0474	0.0474	0.0474	0.0474	0.0474

Note: Coefficient estimates and standard errors from probit estimates are presented; standard errors are clustered at the neighbourhood level. Coefficient estimates statistically significant at * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels, respectively. The 'Income Target' is set as the fixed payment offered, according to the treatment group. Controls are the age, sex and productivity of the participant.

Table 7: Probability of Stopping: Shift 2 (Full Sample, Linear Probability Estimates)

	Dependent Variable: Binary Stopping Decision Full Sample			
	(1)	(2)	(3)	(4)
Cumulative Income	-0.0092**	-0.0014	-0.0028	-0.0057
	(0.0043)	(0.0049)	(0.0057)	(0.0053)
Cumulative Mins. Worked	0.0057***	0.0044***	0.0033**	0.0038***
	(0.0015)	(0.0014)	(0.0014)	(0.0013)
Cum. Income > Income Target (accum.)			0.0169	
			(0.0143)	
Cum. Income > Income Target (paid)				0.0469**
				(0.0189)
Cum. Mins. > Mins. Target			0.0501**	0.0503***
			(0.0185)	(0.0174)
Productivity		-0.1343***	-0.1404***	-0.1405***
		(0.0377)	(0.0371)	(0.0367)
Controls (λ)	No	Yes	Yes	Yes
Interviwer FE's	No	Yes	Yes	Yes
Time of Day FE's	No	Yes	Yes	Yes
Weekday FE's	No	Yes	Yes	Yes
Observations	2,715	2,715	2,715	2,715
Mean of dep. var.	0.0582	0.0582	0.0582	0.0582

Note: Coefficient estimates and standard errors from OLS linear probability estimates are presented; standard errors are clustered at the neighbourhood level. Coefficient estimates statistically significant at * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels, respectively. The Income Target (accum.) and Income Target (paid) are set to be the individual-specific total income accumulated and paid during Shift 1, respectively. The Mins. Target is set to be the individual-specific time worked during Shift 1. Columns (4) controls for whether subject recevied the higher payment in the Shift 1 lottery. Controls are the age and sex of the participant.

Table 8: Probability of Stopping: Shift 2 (Split Sample, Linear Probability Estimates)

	Dependent variable: Binary Stopping Decision					
	Positive Wage Shock		Negative Wage Shock		No Wage Shock	
	(1)	(2)	(3)	(4)	(5)	(6)
Cumulative Income	0.0034	0.0051	0.1110	0.0835	-0.0079	-0.0104
	(0.0149)	(0.0101)	(0.0706)	(0.0760)	(0.0076)	(0.0066)
Cumulative Mins. Worked	0.0042	0.0041	-0.0126	-0.0082	0.0051**	0.0058***
	(0.0070)	(0.0052)	(0.0127)	(0.0134)	(0.0022)	(0.0020)
Cum. Income > Income Target (accum.)	0.0338		0.0427		0.0139	
	(0.0266)		(0.0836)		(0.0202)	
Cum. Income > Income Target (paid)		0.0269		0.2045***		0.0436
		(0.0284)		(0.0521)		(0.0331)
Cum. Mins. > Mins. Target	0.0540	0.0615	0.0559	0.0547*	0.0596*	0.0554*
	(0.0389)	(0.0381)	(0.0539)	(0.0273)	(0.0303)	(0.0292)
Productivity	-0.1630*	-0.1503	0.1121	0.1164	-0.1507**	-0.1399*
	(0.0835)	(0.0815)	(0.1545)	(0.1434)	(0.0684)	(0.0700)
Controls (λ)	Yes	Yes	Yes	Yes	Yes	Yes
Interviwer FE's	Yes	Yes	Yes	Yes	Yes	Yes
Time of Day FE's	Yes	Yes	Yes	Yes	Yes	Yes
Weekday FE's	Yes	Yes	Yes	Yes	Yes	Yes
Observations	627	627	524	524	1,564	1,564
Mean of dep. var.	0.0558	0.0558	0.0687	0.0687	0.0556	0.0556

Note: Coefficient estimates and standard errors from OLS linear probability estimates are presented; standard errors are clustered at the neighbourhood level. Coefficient estimates statistically significant at * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels, respectively. The Income Target (accum.) and Income Target (paid) are set to be the individual-specific total income accumulated and paid during Shift 1, respectively. The Mins. Target is set to be the individual-specific time worked during Shift 1. Columns (2), (4) and (6) control for whether subject recevied the higher payment in the Shift 1 lottery. Controls are the age and sex of the participant.

Table 9: Marginal Effects: Shift 2 (Full Sample, Probit Estimates)

	Evaluation Points: Median Accumulated Income [2] and Time [7.8] Full Sample			
	(1)	(2)	(3)	(4)
Cumulative Income	-0.0075**	-0.0024	-0.0018	-0.0044
	(0.0038)	(0.0035)	(0.0046)	(0.0044)
Cumulative Mins. Worked	0.0048***	0.0036***	0.0031***	0.0036***
	(0.0013)	(0.0011)	(0.0012)	(0.0011)
Cum. Income > Income Target (accum.)			0.0151	
			(0.0120)	
Cum. Income > Income Target (paid)				0.0424***
				(0.0153)
Cum. Mins. > Mins. Target			0.0379***	0.0384***
			(0.0130)	(0.0120)
Controls (λ)	No	No	Yes	Yes
Interviwer FE's	No	Yes	Yes	Yes
Time of Day FE's	No	Yes	Yes	Yes
Weekday FE's	No	Yes	Yes	Yes
Observations	2,715	2,715	2,715	2,715
Mean of dep. var.	0.0582	0.0582	0.0582	0.0582

Note: Coefficient estimates and standard errors from probit estimates are presented; standard errors are clustered at the neighbourhood level. Coefficient estimates statistically significant at * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels, respectively. The Income Target (accum.) and Income Target (paid) are set as the individual-specific total income accumulated and paid during Shift 1, respectively. The Mins. Target is set as the individual-specific time worked during Shift 1. Columns (4) controls for whether subject received the higher payment in the Shift 1 lottery. Controls are the age, sex and productivity of the participant.

Table 10: Marginal Effects: Shift 2 (Split Sample, Probit Estimates)

	Evaluation Points: Median Accumulated Income [2] and Time [7.8]					
	Positive Wage Shock		Negative Wage Shock		No Wage Shock	
	(1)	(2)	(3)	(4)	(5)	(6)
Cumulative Income	0.0061	0.0058	0.1741	0.0905	-0.0060	-0.0085
	(0.0088)	(0.0062)	(0.1712)	(0.1422)	(0.0070)	(0.0061)
Cumulative Mins. Worked	0.0031	0.0035	-0.0159	-0.0048	0.0048**	0.0055***
	(0.0047)	(0.0035)	(0.0268)	(0.0212)	(0.0021)	(0.0017)
Cum. Income > Income Target (accum.)	0.0375**		0.1068*		0.0076	
	(0.0174)		(0.0597)		(0.0177)	
Cum. Income > Income Target (paid)		0.0417**		0.1887***		0.0424
		(0.0173)		(0.0606)		(0.0272)
Cum. Mins. > Mins. Target	0.0258	0.0347	0.0611	0.0629	0.0481**	0.0425**
	(0.0214)	(0.0217)	(0.0541)	(0.0401)	(0.0210)	(0.0187)
Controls (λ)	Yes	Yes	Yes	Yes	Yes	Yes
Interviwer FE's	Yes	Yes	Yes	Yes	Yes	Yes
Time of Day FE's	Yes	Yes	Yes	Yes	Yes	Yes
Weekday FE's	Yes	Yes	Yes	Yes	Yes	Yes
Observations	627	627	524	524	1,564	1,564
Mean of dep. var.	0.0558	0.0558	0.0687	0.0687	0.0556	0.0556

Note: Coefficient estimates and standard errors from probit estimates are presented; standard errors are clustered at the neighbourhood level. Coefficient estimates statistically significant at * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ levels, respectively. The Income Target (accum.) and Income Target (paid) are set to be the individual-specific total income accumulated and paid during Shift 1, respectively. The Mins. Target is set to be the individual-specific time worked during Shift 1. Columns (2), (4) and (6) control for whether subject recevied the higher payment in the Shift 1 lottery. Controls are the age, sex and productivity of the participant.

References

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *The American Economic Review*, 101(2):470–492.

Altonji, J. G. (1986). Intertemporal substitution in labor supply: Evidence from micro data. *Journal of Political Economy*, 94(3):S176–S215.

Andersen, S., Brandon, A., Gneezy, U., and List, J. A. (2014). Toward an understanding of reference-dependent labor supply: Theory and evidence from a field experiment. Working Paper 20695, National Bureau of Economic Research.

Camerer, C., Babcock, L., Loewenstein, G., and Thaler, R. (1997). Labor Supply of New York City Cabdrivers: One Day at a Time*. *The Quarterly Journal of Economics*, 112(2):407–441.

Chetty, R., Guren, A., Manoli, D., and Weber, A. (2011). Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *American Economic Review*, 101(3):471–75.

Crawford, V. P. and Meng, J. (2011). New york city cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. *American Economic Review*, 101(5):1912–32.

Farber, H. (2005). Is tomorrow another day? the labor supply of new york city cabdrivers. *Journal of Political Economy*, 113(1):46–82.

Farber, H. S. (2008). Reference-dependent preferences and labor supply: The case of new york city taxi drivers. *American Economic Review*, 98(3):1069–82.

Fehr, E. and Goette, L. (2007). Do workers work more if wages are high? evidence from a randomized field experiment. *American Economic Review*, 97(1):298–317.

Ghez, G. and Becker, G. (1975). A theory of the allocation of time and goods over the life cycle. In *The Allocation of Time and Goods over the Life Cycle*, pages 1–45. National Bureau of Economic Research, Inc.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Oettinger, G. S. (1999). An empirical analysis of the daily labor supply of stadium venors. *Journal of Political Economy*, 107(2):360–392.

Rabin, M. and Weizsäcker, G. (2009). Narrow bracketing and dominated choices. *The American Economic Review*, 99(4):1508–1543.

Stockley, L. M. (2018). *How Do Expectations Influence Labour Supply? Evidence from Framed Field Experiments*. PhD thesis, University of Toronto.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, 4(3):199–214.