FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

JOÃO FRANCISCO PUGLIESE

REFERENCE-DEPENDENCE AND LABOR SUPPLY DECISIONS

São Paulo

2020

JOÃO FRANCISCO PUGLIESE

REFERENCE-DEPENDENCE AND LABOR SUPPLY DECISIONS

Dissertação apresentada à Escola de Economia de São Paulo como pré-requisito à obtenção de título de mestre em Economia de Empresas.

Orientador: Bruno Ferman.

.

São Paulo

2020

Pugliese, João Francisco.
Reference-dependence and labor supply decisions / João Francisco Pugliese. - 2020.
27 f.

Orientador: Bruno Ferman.
Dissertação (mestrado CMEE) – Fundação Getulio Vargas, Escola de Economia de São Paulo.

1. Economia - Processo decisório. 2. Mercado de trabalho. 3. Renda. 4. Salários. I. Ferman, Bruno. II. Dissertação (mestrado CMEE) – Escola de Economia de São Paulo. III. Fundação Getulio Vargas. IV. Título.

CDU 331.5

Ficha Catalográfica elaborada por: Isabele Oliveira dos Santos Garcia CRB SP-010191/O
Biblioteca Karl A. Boedecker da Fundação Getulio Vargas - SP

JOÃO FRANCISCO PUGLIESE

REFERENCE-DEPENDENCE AND LABOR SUPPLY DECISIONS

Dissertação apresentada à Escola de Economia de São Paulo como pré-requisito à obtenção de título de mestre em Economia de Empresas.

Data de aprovação: __/__/____

Banca examinadora:

Prof. Dr. Bruno Ferman
FGV-EESP (Orientador)

Prof. Dr. Guilherme Lichand
University of Zurich

Prof. Dr. Nicolás Matias Ajzenman
FGV-EESP

Agradecimentos

Gostaria de agradecer à minha família pelo apoio e suporte. Gostaria de agradecer à Ana por sempre estar ao meu lado durante todo esse período. Gostaria de agradecer aos meus amigos da Salinha, que tornaram o Mestrado uma experiência muito melhor, e todos os colegas com quem convivi nesse período. Agradeço ao corpo docente da Escola de Economia por todo o conhecimento transmitido em aulas, conversas e seminários, e em especial o professor Bruno Ferman por aceitar orientar esse artigo. Agradeço também aos membros da banca, Prof. Guilherme Lichand e Prof. Nicolás Ajzenman, por disponibilizarem seu tempo. Por fim, agradeço ao meu co-autor Pedro Bessone pelos ensinamentos e oportunidades.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Resumo

O modelo de oferta de trabalho de ciclo-da-vida prediz que choques de renda transitórios deveriam ter um efeito negligível na provisão de trabalho. Essa predição foi desafiada por uma série de artigos mostrando que trabalhadores tendem a reduzir sua oferta de trabalho depois de receber um choque positivo e transitório de renda, ao qual associam teorias de *income targeting*. Nesse artigo, nós argumentamos que a maioria da evidência existente à favor da hipótese de *income targeting* não leva em consideração um problema de identificação que chamamos de "seleção dinâmica". Um simples porém geral modelo de oferta de trabalho dinâmica mostra que, na existência de um componente não observado no custo marginal do trabalho, a seleção dinâmica levará à identificação espúria de efeitos-renda negativos, mesmo na ausência destes. Nós mostramos a relevância empírica da seleção dinâmica no contexto de um experimento aleatório com variação exógena de salários intra-dia. Ignorando a seleção dinâmica, nós achamos evidência consistente com *income targeting* em diferentes margens da oferta de trabalho. Nós então desenvolvemos e aplicamos uma técnica para solucionar a seleção dinâmica, e após isso não encontramos evidência de *income targeting*. Nós também mostramos que não há evidência de *income targeting* quando usamos choques de renda inesperados e não relacionados à seleção dinâmica. Nossos resultados indicam que a seleção dinâmica é teoreticamente e empiricamente relevante e pode afetar a interpretação dos resultados na literatura de *income targeting*.

Palavras-chave: oferta de trabalho, seleção dinâmica, preferências com dependência de referência.

Abstract

The life-cycle model of labor supply predicts that transitory income shocks should have a negligible effect over the provision of labor. This prediction is defied by a series of papers showing workers tend to reduce labor supply after receiving a transitory positive income shock, which they associated with theories of income targeting. In this paper, we argue that most of the existing evidence in favor of the income targeting hypothesis does not account for an identification problem we name dynamic selection. A simple, yet general model of dynamic labor supply shows that if there is an unobserved component in the marginal cost of work, dynamic selection will lead to the spurious identification of negative income effects even when none are present. We show the empirical relevance of dynamic selection in the context of a randomized controlled trial with exogenous intra-day variation on wages. Ignoring dynamic selection, we find evidence consistent with income targeting in different margins of labor supply. We then develop and apply a technique to address dynamic selection, after which no evidence of income targeting is found. We also show no evidence of income targeting from exogenous monetary windfalls, which should not be affected by dynamic selection. Our results indicate that dynamic selection is theoretically and empirically relevant and could affect the interpretation of the results in the income targeting literature.

Keywords: labor supply, dynamic selection, reference-dependent preferences.

Contents

1 Introduction

The neoclassical life-cycle model of labor supply predicts that, short of winning in the lottery, transitory changes in daily income should have a negligible impact on labor supply.[1] However, this basic prediction has been challenged by a series of prominent papers that show a negative relation between daily income and labor supply for NYC cab drivers (Camerer et al., 1997; Farber, 2015; Crawford and Meng, 2011; Thakral and Tô, 2019). Such behavior can be explained with a theory of reference-dependent preferences, in which drivers quit working for the day upon reaching a target level of earnings.

The evidence on income targeting from cab drivers provides key insights for the broader literature of reference-dependent preferences. It is evidence that reference-dependent preferences matter in a real-world economic setting and it remains to this day one of the key evidences of reference-dependent preferences outside of the lab (O'Donoghue and Sprenger, 2018). Moreover, Thakral and Tô (2019) presents evidence that cab drivers' reference point formation follows an adaptive process throughout the day.

We propose an alternative explanation for the empirical findings in the cab driver literature that does not rely on reference-dependent preferences. We develop a simple, yet general dynamic model of labor supply that, purposefully, has no income effects. According to this model, even in the absence of income effects, the duration models used to estimate income effects in the cab driver literature should find a negative relationship between earnings earlier in the day and labor supply later in the day.

In our model, what drives this negative relationship is a selection bias we call *dynamic selection*. As workers decide to leave throughout the day, the sample of workers who faced high wages early in the day cease to be comparable with the sample that received a low wages early in the day. The intuition is that, absent income effects, workers who received low wages early in the day have lower incentives to keep working, and will tend to quit for the day, leaving the sample. Those who, in spite of low wages, do not quit early on, will therefore be positively selected. In other words, they find worthwhile to keep working under low incentives because they will have, on average, a lower marginal cost of labor in that day. Then, workers who faced low wages earlier in the day and

[1] The provision of labor supply by forward-looking workers should only be affected by their life-time wealth, which in general will not be meaningfully altered by daily income shocks.

accumulated less income will also be less likely to stop working at any moment later in the day, explaining the negative relationship between income early in the day and labor supply later on.

We test the empirical relevance of our theory using data from an experiment with low-income workers conducted in Chennai, India. In this study, designed to investigate the economic consequences to increase sleep among the urban poor, 452 participants worked for up to 28 days in a full-time data-entry job (Bessone et al., 2019). The workers had considerable discretion over their labor supply: they could decide their time of arrival and departure, when to take breaks, and how much effort to exert in their work, making it similar to other settings with flexible working hours.

Two features of this setting are essential to disentangle income targeting from dynamic selection. First, a large part of participants earnings was determined by a piece-rate level that fluctuated randomly between a low and a high value throughout the day. This allows us to investigate the impact of variations without the typical concern that variation in income might be correlated with aggregate supply shocks directly affecting labor supply.[2] Second, unlike in the cab driver setting, we can observe the wage level at any given point in time. As we explain in detail below, this is crucial for our strategy to control for dynamic selection.

Ignoring the issue of dynamic selection, we replicate what we consider to be the main results of the income targeting literature in our setting. We find that workers randomly assigned to higher piece rates early in the day had a higher probability to quit, provided less effort as measured by the total output produced in the data-entry work, and spent more time in working breaks later in the day. The results are precisely estimated and comparable in magnitude to the income targeting literature. Using a similar specification to previous papers, we find that a 10% increase in cumulative earnings early in the day is associated with an increase of roughly 9.5% in the stopping probability later in the day.[3]

Should these results be interpreted as evidence of income targeting? To answer that question, we must correct for the dynamic selection bias. We propose two solutions,

[2] To our knowledge, this is the first paper in the income targeting literature to explore exogenous intra-day wage variation. An exception is Fehr and Goette (2007), which explores random wage variation *across* days. Since they have no intra-day wage variation, they do not perform the same test of income targeting that this paper and the cab driver literature has performed.

[3] In comparison, in samples of NYC cab drivers, Thakral and Tô (2019) report a 3.3% increase in the stopping probability, while Farber (2015) reports a 9% increase during daytime shifts.

one using insights from the model and another based on the experimental setting. First, we show in our model that workers whose immediately preceding piece rate was low will have on average the same marginal cost of work regardless of their wage history up to that point. In our model, the marginal cost of work is increasing over time while the marginal benefit increases in wages. By selecting those participants who experienced the lowest possible wage in the most recent period, and still decided to keep working, we average-out differences in marginal costs of effort that drives dynamic selection. Second, we take advantage of random variation in "short days", in which participants could only work on a limited number of hours and were incentivized to quit working exactly at 5PM. On those days, the extensive margin of labor supply does not operate, and we can investigate the presence of income effects on overall effort and work breaks without the concern of dynamic selection.

After controlling for dynamic selection, our results change considerably: we find no evidence of negative income effects on labor supply, giving support for the neoclassical model. We find precisely estimate nulls for stopping decision, effort provision, and work breaks when we focus on the sample that had received a low piece rate in the previous period. While this method does not allow us to estimate income effects from variations in the first lag of piece rates, it is remarkable that all the other wage lags go from negative and significant in the standard estimation to very small and insignificant – or even *positive* and significant in some cases – when we control for dynamic selection. Moreover, we find very limited evidence of income targeting in work breaks and effort on short days, when the scope for dynamic selection is much smaller.

We corroborate our null finding using another exogenous source of income variation that is impervious to dynamic selection. We use exogenous variation in the payment for the completion of a cognitive task, which increased cumulative earnings by 12% on average. This variation can be interpreted as a monetary windfall because the task was short, performed at a time of the day when the vast majority of participants have not yet left, and very inelastic to the variation in payment. This monetary windfall should have a negligible effect on labor supply except through income effects. Again, we find no evidence of negative income effects. If anything, we find that participants have a small boost in output just after receiving the monetary windfall, in stark contrast with the predictions from models of income targeting.[4]

[4] One caveat is that participants may have different mental accounts depending on the source of income,

Our paper makes three contributions to the literature. First, we introduce the issue of dynamic selection to the estimation of daily income effects, which, in tandem with our empirical results, casts doubts on the interpretation of part of the income targeting literature (Farber, 2005, 2015; Crawford and Meng, 2011; Thakral and Tô, 2019). Future works focused on estimating daily income effects should take dynamic selection into account. In the case of this paper, the setting provided natural, non-parametric solutions. When such solutions are infeasible, an alternative is to explicitly model the dynamic selection in their empirical models, in the spirit of Card and Hyslop (2005), which face a similar issue when estimating a hazard model for Unemployment Insurance recipients.[5]

A second contribution to the literature is that we are the first paper to explore variation in earned (wages) and non-earned income on labor supply in the same setting. Contrary to the findings of papers using an intra-day variation on wages, the two papers we are aware that use exogenous variation on monetary windfalls to estimate income effects find no evidence of income targeting (Dupas et al., 2019; Andersen et al., 2018). Our results suggest that a possible reason behind these different results is that income effect estimates from monetary windfalls are not affected by dynamic selection issues.

Finally, our paper also complements the income targeting literature in developing, non-Western settings. With the exception of Dupas et al. (2019) and Andersen et al. (2018), all other papers in the literature study income targeting in rich, Western countries.[6] In comparison to these two papers in developing countries, we contribute by having exogenous variation in earned income. This may be important if non-earned income does not enter the same mental account as earned income, one intuition backed by the literature on narrow bracketing and mental accounting (Read et al., 1999).

The rest of this paper is organized as follows: Section 2 presents the setting and describes the data. Section 3 present the empirical framework and formalizes the issue of dynamic selection. Section 4 shows the estimates of income effect and Section 5 summarises our results and takeaways for future work in this literature.

---

and do not consider monetary windfalls in their targets. If this is true, a null effect of monetary windfalls do not rule out the existence of income targeting (Dupas et al., 2019).

[5] See also Ham and Lalonde (1996), who noted a similar selection issue to the one we point here in the literature of welfare dependency.

[6] Some examples include Thakral and Tô (2019), Farber (2005), Crawford and Meng (2011), Chang and Gross (2014), and Oettinger (1999) in the US; Fehr and Goette (2007) in Switzerland; and Jonason and Wållgren (2013) in Sweden.

2 Setting

We test for the existence of daily income effects using data collected to evaluate the consequences of improving the sleep of low income, urban workers in Chennai, India (Bessone et al., 2019). In order to study the impact of sleep on work-related outcomes, they set up a data-entry firm in their lab and recruited 452 participants to work full time for up to 28 workdays, or "shifts". During the study, the participants were exposed to a series of random shocks to their income, allowing us to study the role of income effect on labor supply.

Recruitment and sample characteristics. Recruitment followed two strategies: first, recruiters went to low-income neighborhoods in Chennai and spread information about the study. Advertisements for the study offered a one-month data entry job, and recruiters provided interested individuals with additional information. Second, past participants could refer to potential new participants in the study. In both cases, recruiters approached interested individuals to interview them and determine their eligibility to participate in the study.

Interested individuals participated in a screening process involving a home visit to check whether the individual met the study's eligibility criteria. Relevant to the present study, the eligibility criteria were: (i) being between 25 and 55 years old; (ii) fluency in Tamil (the local language); (iii) the ability to read and write numbers; (iv) having worked fewer than five days per week in the previous month; (v) earning less than Rs. 700 ($10) per day in the previous month, (vi) having the intention of staying in Chennai for the following 5 weeks; and (vii) no children in the household younger than three years.[1] 62% of individuals approached on the street agreed to take the eligibility survey, of which 39% were deemed eligible to participate in the study.

Table 1, Panel A, displays sample characteristics at the participant level. Sixty-six percent of study participants were female, and the typical study participant was roughly 33 years old and had 1.3 children on average. The participants had about ten years of

[1] In addition, participants also needed to live in a dwelling able to accommodate the home improvements that were part of the intervention in Bessone et al. (2019). They also had to own three or fewer of the sleep aids offered in their study.

education on average, and only about 30% of them had experience using computers.[2]

Data-entry work. The data-entry task consisted of digitizing text and numeric data designed to mimic a real-world, data-entry job, using the interface represented in Figure 1. The participants were presented with a "typing page" in the left of the screen and should re-enter the data into text boxes, or "fields", located in the right of their screen.[3]. After submitting the data from the last field in a typing page, the participants received a new typing page to be digitized.

The data that participants digitized were artificially generated by Bessone et al. (2019), which allows them to keep the average difficulty of the task constant across participants. Moreover, since the entry data was artificially generated, we are able to measure perfectly whether the participants are entering the data correctly or not.

Working hours. On average, the participants spent 66% of their time in the office doing the data-entry work (Table 1, Panel B). The rest of their time was spent completing mandatory surveys and other experimental tasks not relevant for this study. In the time allotted for the data entry work, participants could choose how long they spent typing and how long they spent on work breaks.

The participants were also free to choose their arrival and departure times. Their shifts pertained to two types. Most shifts were "long shifts" (or long days) when participants could work from 9:30 AM to 8 PM. In the remaining shifts, which we call "short shifts" (or shot days), the participants could only work from 11 AM to 5 PM. Whether a shift was short or long was randomized at the participant level, so the type of shift should not be correlated with calendar dates or any other participants' characteristics.

The short shifts were designed to shut down the extensive margin of labor supply in a part of the workdays. This is essential to deal with selection issues related to the decision to stop working, which we discuss in detail in Section 3. To keep working hours constant on short days, the participants received a Rs. 50 bonus (roughly 20% of average daily earnings for this type of workday) if they were in the office from 11 AM to 5 PM. Figure

[2] During the first day of the study, participants were introduced to the data entry work and were trained to use computers.

[3] They had to fill the fields sequentially. The software did not allow returning to a field nor skipping any field.

A1a shows that exit time was very concentrated at 5 PM for short shifts. In contrast, on long days, participants presented much more variability in terms of exiting time (Figure A1b), indicating that the extensive margin of labor supply was much less active in short than in long shifts. Because of that, most of our analysis is focused on long shifts, unless in instances in which we explicitly want to shut down the extensive margin of labor supply.

Earnings. Participants' overall earnings consisted of a series of payments ranging from their performance in the data-entry work to payments for completing surveys, experimental tasks, and for being at the office on time. By far, the largest component of participants' payment comes from data-entry, representing Rs. 343 per day or 71% of their daily earnings on average in long shifts (1, Panel B).

The payment for the data-entry work had two components. First, the participants received Rs. 21.60 ($0.30) per hour of active typing. If a participant spent two consecutive minutes without typing, the software would automatically pause, and the participant would not receive the time-based payment until they began typing again. Second, participants received a performance-based payment consisting of a piece-rate per correct character and a penalty per mistake.[4] The piece-rate was randomized between a "High" value (Rs. 2 per 100 correct characters) and a "Low" value (Rs. 0.5 per 100 correct characters), while the penalty rate was always at Rs. 10 per 100 mistakes.

The piece-rates' value was randomized within-shifts, mimicking the instantaneous wage variation in real jobs such as for cab, ride-share, and rickshaw drivers. Each day was divided into time windows called "incentive sessions", which lasted for at least 30 minutes. After 30 minutes had passed in a given session, the next submitted field would trigger an incentive switch, in which point a new piece-rate would be randomly assigned to the participant and a new incentive session would begin.[5] At each incentive session switch, a piece-rate was drawn independently with equal probabilities between High and Low.

The participant could only adopt an income targeting strategy if income, or a component of income, is observable throughout the day. Throughout the whole study, the

4 The number of mistakes was defined by the Levenshtein distance between the entered and the original texts. The Levenshtein distance between two strings is the minimum number of changes (insertions, deletions, or substitutions) required for both strings to be equal.

5 Thus, some incentive sessions can last for a few additional minutes. For example, a long paragraph, which takes minutes to be completed, may begin at the very end of an incentive session, and the change will only be triggered after it is submitted.

new piece-rate values appear on the bottom right corner of the screen (see Figure 1) on every incentive session switch. As an indication that participants notice the piece-rates, the output produced in High incentive sessions was 15% higher on average than in Low incentive sessions (See Section 4 for more details). In the first five days in the study, a screen summarizing participants' earnings appeared after each typing page was submitted. This summary presented the total earnings from transcribing the page, as well as breaking the payment down into gains from correct entries, penalties from mistakes, and typing time. After the sixth day, participants no longer viewed that screen, but they were still able to track their piece-rate history. Since the variation on piece-rates accounts for 63% of the within-participant variation in earnings, participants could infer their earnings with relative precision by tracking their piece-rate history.

Outcome variables. Our first measure of labor supply is a quality-adjusted measure of output defined as the number of correct entries minus eight times the number of mistake per incentive session.[6] This measure captures three margins of labor supply that may be affected by daily income: effort in the data entry task, work breaks, and the decision to stop working.[7] On average, participants produce 18,328 units of output on long shifts, or 3,410 units per hour (Table 1, Panel B).

Our second measure of labor supply is the amount of time spent with voluntary work breaks. On average, participants spend 16 minutes per day or 1.5 minutes per incentive session on voluntary pauses, which represents 3.6% of the time in the average incentive session. (Table 1, Panel C).

Our third measure of labor supply is the participant's stopping decision, defined as the time the participant entered the last data point in any given day. This measure captures the "extensive margin" of labor supply and has the advantage of matching the measure of labor supply used in most of the income targeting literature (Farber, 2005, 2015; Thakral and Tô, 2019).

6 This measure was pre-registered in Bessone et al. (2019). The weight on mistakes was set to equal the ratio between the penalty rate of Rs. 0.1 per mistake and the average piece-rate of Rs. 0.0125 per correct entry.

7 The effort margin is mostly driven by typing speed because the mistake rate is very low (below 0.01 mistakes per entry). In fact, the output measure correlates almost perfectly with the number of correct entries ($\rho \approx .92$), which drives the effort margin.

3 Empirical Framework

In this section, we lay out the empirical framework for estimating daily income effects on labor supply. We show that the empirical models currently used to estimate income effect on labor supply do not take into account a selection issue, which we call "dynamic selection". This selection issue leads to a negative bias on the conventional estimates of income effect on labor supply. Finally, we propose an empirical strategy that controls for dynamic selection.

3.1 Output and work breaks

We investigate the presence of income effects on output and work breaks by estimating the equation

$$y_{ids} = \sum_{j=0}^{4} \beta_j \text{High}_{ids-j} + \gamma X_{ids} + \varepsilon_{ids} \tag{3.1}$$

where y_{ids} is a labor supply measure of participant i at date d and incentive session s. We estimate this model with our three preferred measures: output, work breaks, and a dummy indicating stopping decision. We choose the incentive session s as the unit of observation because it is the level of variation of piece rates. This is important as it shuts down any auto-correlation between wages during the day, an issue we discuss below. The variables High_{idj} are dummies indicating whether session j had a high piece rate. Note that we include the contemporaneous piece rate, with the associated coefficient β_0 capturing the substitution effect of high piece rates on labor supply. We also include the piece rates from the previous four sessions and the associated coefficients β_j capture flexibly the impact of previous piece rates on contemporaneous labor supply. Since wages are random, we do not need to add covariates for identification, but we still control for covariates X_{ids} to improve the precision of our estimates. The controls include individual, date, day in study, and incentive session fixed effects.

According to the neoclassical model of labor supply, the β_j coefficients should be zero for any $j > 0$. This theory posits that workers should take into account their life-time wealth when deciding how much labor supply to provide at any given day (Farber, 2005). Therefore, short of winning in the lottery, transitory income shocks early in the day should have a negligible effect on labor supply provision later on.

On the other hand, in the theory of income targeting, a positive income shock early on should decrease labor supply later in the day. In this theory, the participants set a daily target of how much work they would like to do and either try to work exactly towards that goal (Camerer et al., 1997) or have a higher marginal utility of income while they are below this target(Köszegi and Rabin, 2006). In both cases, the income targeting theory predicts that workers should exert more effort or work more when they are below their target. In particular, in our empirical model, a theory of daily income targeting implies that when the outcome variable is output $\beta_j < 0$ for $j > 0$. Conversely, when the outcome is either work breaks or stopping decision, income targeting implies that $\beta_j > 0$ for $j > 0$.

There are a series of identification issues with models of daily labor supply. First, in the real world, past wages may be correlated with future wages. This may affect the worker's decision of labor supply decision. For example, if wages are positively correlated over time, a worker might decide to stay for longer in the job since the probability of receiving a high wage in the future would be higher. This is not an issue in our setting, since, by design, wages are not correlated across incentive sessions, and we control for the piece rate in the current incentive session.

Second, whenever wages are higher, workers may decide to exert more effort. Then, a worker that received a series of high wages earlier in the day not only accumulated more income but also may have exerted more effort earlier in the day. If exerting effort early in the day reduces the marginal utility of supplying labor later on (because of fatigue, for example), they will supply less labor in the current session. In that case, the income effect estimates would be negatively biased. Although this may be true in our setting, this issue is not crucial to our main results since we do not find evidence of negative income effects. In addition, in Section 4 we explore an alternative source of income variation that does not suffer from this issue.

Third, because at any given incentive session we can only observe workers who have not yet left, this may cause an issue of dynamic selection, which will tend to bias the estimates of income effects negatively. We discuss this issue in-depth in Section 3.3.

3.2 Stopping Decision

For the sake of comparability with the previous literature (Farber, 2005, 2008, 2015; Thakral and Tô, 2019), we estimate an alternative specification to identify income effects on the stopping decision. Following Thakral and Tô (2019), we estimate

$$Q_{idp} = \sum_j \mathbb{1}\{h_{idp} \in H_j\}(\alpha_j h_{idp} + \beta_j I_{idp} + \gamma_j X_{idp}) + \varepsilon_{idp} \quad (3.2)$$

where Q_{idp} is a dummy denoting whether participant i at date d stops working for the day after typing page p.[1] Note that we have changed the unit of observation in comparison to equation 3.1. We did that because most participants make the decision to stop or keep working after submitting a typing page.[2]

The variables I_{idp} and h_{idp} represent, respectively, the cumulative income and hours since the start of the shift up to the beginning of page p. Because these variables are positively correlated, we need to control for h_{idp} to be able to causally interpret the marginal effect of cumulative income on the decision to stop working. Following Farber (2005) and Thakral and Tô (2019), we control for cumulative working flexibly, to avoid bias from functional form misspecification on equation 3.2. To do so, we divide the shift into 30 minutes windows H_j with j indexing these windows[3] and estimate one model per window separately.

The parameters of interest in this regression are the β_j, which captures the marginal effect of cumulative income on the stopping decision at different moments of the shift j. Similarly to the previous section, under the neoclassical model of labor supply, β_j should be equal to zero, while theories of income targeting predict that the β_j's should be positive.

This specification has a key advantage in comparison to specification 3.1. By allowing the income effects to vary flexibly throughout the day, we are able to isolate income effects estimates closer to the end of the shift. Because at the beginning of the day, virtually no participant is in the margin between quit and keep working, one cannot hope to identify income effects at these early sessions. This flexible specification allows us to deal with this issue in a transparent way. Following Thakral and Tô (2019), we also give special focus to the β_j in the time window j that includes the median shift duration of participants, 7.5 hours.

[1] As described in Section 2, the participants must digitize data coming from a fictitious form.
[2] 76% of participants stop to work after completing a page, rather than in the middle of a page.
[3] Windows are centered, meaning that at window h pages ending with ± 15 minutes of h are included.

On the other hand, a challenge to this specification is that the ability to accumulate income might be correlated with other factors behind the stopping decision. For example, more productive participants may also have longer endurance, which would bias the estimates of income effect positively. To address that, we instrument cumulative income with the share of High piece rates up to, but not including, the current incentive session. Thus, all our estimates from Equation 3.2 come from 2SLS estimators that we fit for each time window. We consider this approach a step forward in comparison to the previous literature, which relies on (admittedly rich) covariates to deal with the endogeneity of accumulated income.

3.3 Dynamic selection

In addition to the empirical challenges explained above, we identify a novel and empirically relevant source of bias pervasive in most papers in the income targeting literature. This challenge to identification, which we denote by dynamic selection, should generally generate a negative bias in the estimates of income effect on labor supply, which could lead us to erroneously attribute a negative effect of past wages on labor supply to income targeting.

The crux of the issue is that the sample of workers who *decided* to keep working even after receiving lower wages earlier in the day is positively selected in comparison to the workers who kept working after receiving a series of high wages earlier on. The high wage workers had heightened incentives to keep working than the low wage workers up to the time of the day we are observing them. If the low wage workers kept working in spite of that, it might be because they have a low cost of working on that day. In that case, when faced with a fixed piece rate in a given moment of the day, the worker who received low wages earlier on and kept working will be more likely to keep working than the worker who received high wages earlier on. Crucially, this might be simply because the low wage workers are positively selected, rather than because of a causal effect of past wages on labor supply.

We formalize our point with a simple yet general model of dynamic labor supply. We shut-off income effects and any causal influence between previous wages and contemporaneous labor supply decisions. Even so, an econometrician would still find a (spurious) relationship between past wages and contemporaneous stopping decision. Although we do not model that explicitly, this issue could also affect other margins of labor supply, since the workers who receive high and low wages early in the day are not comparable at later moments of the day.

The agent's shift is divided in time sessions indexed by $t \in \{1, 2, ...\}$ and in each session they decide whether to keep or to stop working, denoted by $d_t = 1$ and $d_t = 0$, respectively. If they choose $d_t = 0$, then $d_{t'} = 0$ for any $t' > t$.

Before the agent chooses d_t, she observes three relevant state variables at the beginning of each session. First, the time of the session, denoted by t, which captures how long they have been working for. Second, in each section, the worker is randomized to receive either a low or high wage. We denote the wage at t by $w_t \in \{w_L, w_H\}$, in which $w_L < w_H$. The wage follows an independent and uniform process, and we denote the wage history up to session t by $\boldsymbol{w}^t$. At t the worker knows $\boldsymbol{w}^t$, but not the wage for any subsequent section. Third, each agent draws a daily-level labor supply cost shock represented by the absolutely continuous random variable ε, distributed according to the cumulative distributive function F. Importantly, this shock is observed by the agent before any decision they make, but it is not observed by the econometrician, unlike the other state variables.

Instantaneous preferences. The agent's preferences on each sessions are represented by the instantaneous utility function $u(t, w_t, \varepsilon)$. We normalize the utility of stopping to be zero at all sessions. We assume that u is a differentiable function with partial derivatives given by $u_t < 0, u_w > 0$, and $u_\varepsilon < 0$.

The agent's utility is decreasing in the time of the session to capture the idea that it is more costly to work when you are tired, as would be the case later on in the day. The agent utility is naturally increasing on the session's wage, implying that the workers are less likely to quit in a session with high wage than in a session with low wage.[4] Finally, since we called ε a labor supply cost shock, we determine that $u_\varepsilon < 0$.

Inter-temporal preferences Let $\boldsymbol{d}^t$ be a state contingent-vector of stopping decisions decisions from t onward. The utility of the agent from session t until the end of

[4] Although this is a natural assumption, we verify empirically that this is the case in our setting. Furthermore, Chen and Sheldon (2015) also provides evidence that Uber drivers are less likely to quit when dynamic price surges are in place.

the day is given by

$$U(\boldsymbol{d}^t, t, \varepsilon) = \sum_{t'=t}^{\infty} \mathbb{E}_{w_t} \left[d_t(t, w_t) u(t, w_t, \varepsilon) \right]$$

The corresponding value function is given by $V(t, w_t \varepsilon)$ which can be represented recursively as

$$V(t, w_t, \varepsilon) = \max_{d_t \in \Delta_t} \left\{ d_t u(t, w_t, \varepsilon) + d_t \mathbb{E}_{w_{t+1}} \left[V(t+1, w_{t+1}, \varepsilon) \right] \right\} \tag{3.3}$$

Going forward we adopt the following notation for the expected value over wages: $V^e(t, \varepsilon) = \mathbb{E}_w \left[V(t, w_t, \varepsilon) \right]$

3.3.1 Main results

Under the assumptions we made, the optimal stopping rule for the agent follows a simple monotonic strategy, which we formalize in Lemma 3.1. This and all the other proofs are found in the appendix.

Lemma 3.1. *For every wage w and time period t, there exists a unique value $\overline{\varepsilon}(t, w)$ such that the agent keeps working at t (i.e., $d^*(t, w, \varepsilon) = 1$) iff $\varepsilon \leq \overline{\varepsilon}(t, w)$. Moreover, $\overline{\varepsilon}(t, w)$ is increasing on w and decreasing on t.*

This characterization result is important to understand where the dynamic selection issue appears in the model. It says that from the econometrician point of view, in which we do not observe ε, an individual who is still working in a large t or after receiving a low wage w, will tend to have a smaller labor supply cost shock ε. We formalize this result on Lemma 3.2.

Lemma 3.2. *Assume that at t, the econometrician observes an individual with wage history $\boldsymbol{w}^t$ with at least one instance of w_H and one instance of w_L and that chose to work at t (i.e., $d_t = 1$).*

Let $T^L = \max_{t'} \{t' | t' \leq t, w_{t'} = w_L\}$ and $T^H = \max_{t'} \{t' | t' \leq t, w_{t'} = w_H\}$.

(i) The distribution of ε conditional on the econometrician's information set at t is given by

$$F(\varepsilon | d_t = 1, \boldsymbol{w}^t) = F(\varepsilon | \varepsilon < \hat{\varepsilon}(\boldsymbol{w}^t)) \ , \ \textit{where} \ \ \hat{\varepsilon}(\boldsymbol{w}^t) = \min \left\{ \overline{\varepsilon}(T^H, w_H), \overline{\varepsilon}(T^L, w_L) \right\} \tag{3.4}$$

(ii) Moreover, if $T^L > T^H$ then $\hat{\varepsilon}(\boldsymbol{w}^t) = \overline{\varepsilon}(t, w_L)$

(iii) If $w_t = w_L$ or $w_t = w_H$ for every t, then $\hat{\varepsilon}(\boldsymbol{w}^t) = \overline{\varepsilon}(t, w_t)$.

This lemma shows two important results. First, we show that from the econometrician's perspective, all that matters to account for selection from the cost shock ε is the function $\hat{\varepsilon}(\boldsymbol{w}^t)$. Moreover, this function is quite simple, depending only on the decision thresholds defined on Lemma 3.1 in the latest periods in which we observe the high and the low piece rates. If the last wage we observe is high, it is not clear whether the lowest utility from work, and therefore the smallest value of $\overline{\varepsilon}(t', w_{t'})$, comes from the last section, or in the latest previous session in which $w_{t'} = w_L$. This will depend on the dynamics of the disutility of work over time (u_t), the utility from higher piece rates u_w, and the interaction between the two. The second important point is that whenever the last observed wage is w_L, then the last session of work will be the one with the smallest utility and, therefore, the smallest $\overline{\varepsilon}(t', w_{t'})$. Finally, the third item in the proposition in included for completeness.

Now we are ready to show the two main results of this section. From Lemma 3.2, we know that, under this model, the probability that an agent keeps working at any given section t conditional on the wage history and not having stopped before t is given by

$$Pr(d_t = 1|d_{t-1} = 1, \boldsymbol{w}^t) = F\left(\hat{\varepsilon}(\boldsymbol{w}^t)|\varepsilon < \hat{\varepsilon}(\boldsymbol{w}^{t-1})\right) \tag{3.5}$$

Proposition 3.1 proves a case for which, even in a model without income effects, the dynamic selection issue would lead the econometrician to find that individuals with higher income early in the day are more likely to quit working at any given session.

Proposition 3.1. *Take an observation at t for which $d_{t-1} = 1$ and that $w_t = w_H$. Then,*

$$Pr(d_t = 1|d_{t-1} = 1, w_t, w_{t-1} = w_H) < Pr(d_t = 1|d_{t-1} = 1, w_t, w_{t-1} = w_L) \tag{3.6}$$

The proof of this statement follows directly from Lemma 3.2, since an agents who faced $w_{t-1} = w_L$ has, on average, a smaller value of ε than an agent who faced $w_{t-1} = w_H$. Lemma 3.2 also helps to develop a strategy to control for the dynamic selection issue. Notice that at t, the distribution of the cost shock ε only depends on the last session wage when $w_{t-1} = w_L$. Therefore, the follow results follow

Proposition 3.2. *Take a session t for which $d_{t-1} = 1$ and with wage history $\boldsymbol{w}^t$ such that $w_{t-1} = w_L$. Then*

$$Pr(d_t = 1|d_{t-1} = 1, w_t, w_{t-1} = w_L, \boldsymbol{w}^{t-2}) = Pr(d_t = 1|d_{t-1} = 1, w_t, w_{t-1} = w_L)$$

Proposition 3.2 states that any wage prior to session $t-1$ has no predictive effect on the stopping decision once the econometrician conditions on $w_{t-1} = w_L$. Thus, if the econometrician finds evidence that past wages affect stopping decision even after conditioning on $w_{t-1} = w_L$, this model is rejected in favor of a model of labor supply with income effects.[5]

In the next section, we formally present the dynamic selection problem in our context and show how it can generate "spurious" correlations between cumulative earnings and the stopping decision. Furthermore, we propose a fix for this problem that can easily be implemented in our setting.

[5] It is important to point that we still cannot identify with this approach any income effect coming exclusively from income earned at the immediately previous time period. Thakral and Tô (2019) points out that in the context of New York City cab drivers, the negative income effect seems to come mostly recent income, although not exclusively from the past hour.

4 Results

In this section, we discuss the results of our empirical specifications using piece-rate variation and monetary windfalls. We find that High piece-rates earlier in the day decrease subsequent output, and show that these effects are driven exclusively by dynamic selection. In addition, monetary windfalls do not significantly affect labor supply. We then discuss the implications of our findings to the income targeting literature.

4.1 Piece-rate variation

We start our investigation by ignoring the dynamic selection issue. In columns 1-3 of Table 2, we use data from long shifts, when participants had more freedom to choose what time they would like to leave and, therefore, dynamic selection is a potential problem to the identification. In column 1, we can see that while a contemporaneous High piece-rate increased output by 283 units (19.2% in comparison to the excluded group's mean), having a high piece-rate in the previous session decreased contemporaneous output by 55 units or 19.4% of the effect of the contemporaneous piece-rate. Moreover, all the other lags of High piece-rate dummies are negative and statistically significant at least at the 5% level, in accordance to theories of income targeting. We find a similar pattern for work breaks and the stopping decision. Contemporaneous High piece-rates increase labor supply provision by decreasing breaks (56% in comparison to the excluded group mean, $p < 0.01$) and the probability of quitting (57%), while the lags work in the opposite direction. Interestingly, for all the outcome variables, the coefficients associated with second to fourth lags are significantly smaller (in magnitude) than that of the first lag. These findings echo Thakral and Tô (2019), who find that recent income shocks have a larger effect on labor supply than less recent ones on the same day.

However, the interpretation of these findings as income targeting is complicated by dynamic selection. In columns 4 to 6 of Table 2, we consider only short shifts, when participants have a high incentive to leave precisely at 5 PM. The differences are staggering. While the contemporaneous effect of the High piece-rate on output and work breaks is still positive and substantial, the estimates of almost all lags of piece-rates decrease in magnitude or change sign. This suggests that the negative income effects estimated on long days are capturing dynamic selection rather than income targeting behavior.

One downside of focusing on short shifts is that the participants work for a shorter period of time, and the extensive margin is shut down. This could be driving the differences in results in comparison to long days. In columns 7 to 9, we focus again on long shifts, but control for dynamic selection using the strategy devised in Section 3.3, focusing in the sub-sample that received a low piece-rate in the immediately previous section. This specification reveals that, if anything, having a High piece-rate two sessions ago induces a *positive* effect on the output of roughly 1% ($p = 0.019$). This result further supports the idea that negative income effects in columns 1 to 3 and in the literature as a whole may be a result of a failure to control for dynamic selection.

In conclusion, the results from this section indicate a limited role for negative income effects on labor supply. This is true both in different margins of labor supply once we control for the issue of dynamic selection. Next, we explore whether the same result holds when income shocks are induced by monetary windfalls, which do not present the issue of dynamic selection.

4.2 Monetary windfall

We corroborate the results from the previous subsection by using monetary windfalls as an exogenous source of income variation. In comparison to variations in wages, monetary windfalls have the advantage of affecting earnings without directly affecting the stopping probability since substitution effects are null. As such, dynamic selection is not an issue for this type of variation.

To estimate income effects from monetary windfalls, we estimate the following regression:

$$\text{Output}_{idt} = \beta_t \text{DoublePay}_{id} + \gamma X_{idt} + \varepsilon_{idt} \tag{4.1}$$

where i represents a participant, d a day, and t a set of 30-minutes windows after the monetary windfall.[1]

[1] We center every time windows such that $t = 1$ captures the first 30 minutes after the cognitive task (i.e., the monetary windfall) has been completed, and so on.

The monetary windfall is captured by the dummy variable DoublePay_{id}, which indicates whether the participant received a High payment on any of the two cognitive tasks. This implies that the participants had a roughly 12% income shock on average in comparison to the excluded group, considering the income accumulated until the cognitive task. X_{idt} is a vector of covariates capturing participant, date, day in study, and time window fixed effects.

This specification is not conditional on being at work: we set $\text{Output}_{idt} = 0$ if the participant is not present at the office during working hours. This way, we capture the total effect on output, including both intensive (productivity) and extensive margins (quitting time).

Figure 2a shows the average cumulative output at different t for the participants that received, and that did not receive a monetary windfall. Both curves behave similarly before and after the monetary windfall. Surprisingly, the cumulative output by the end of the day of the group that received a monetary windfall is slightly higher (23 units, $p = 0.81$) than the output of the group that did not.[2] Figure 2b shows the estimates from equation 4.1, which include control variables and analyze the effect of the monetary windfall on output at different moments after the monetary shock. Considering the first three windows, we estimate a *positive* income effect on output of 20.6 ($p = 0.081$), 17.5 ($p = 0.207$), and 20.8 ($p = 0.289$) units, respectively. The remaining windows are much closer to zero and also insignificant.

We interpret this result as strong evidence against income targeting in the labor supply. The income shock is relatively large, about 12% of average income up to that point of the day, and our estimates are precise. For example, when looking at the aggregate output after the monetary windfall, we can reject income effects below -168 units (2.5% of average output in that period) at the 95% level. Moreover, this specification has the key advantage of not being affected by dynamic selection or fatigue confounders, which afflict the specifications that use wage variation as a shock to daily income.

Our results suggest that dynamic selection issues could be a problem for the identification of income effects in part of the literature. In the next subsection, we analyze what our results mean for the income targeting literature.

[2] This number refers to a regression at the daily level including participant and date fixed effects. Errors are clustered for calculating p-values.

4.3 Implications to the literature

In our setting, we show that dynamic selection is driving results that were seemingly in accordance to the predictions of income targeting theory. There are reasons to worry that dynamic selection could also be driving the findings of income targeting in other settings. There are two conditions necessary for dynamic selection to become an issue, and these are likely to arise in most settings. First, an increase in contemporaneous wages should reduce the likelihood of ending the shift. This condition could be true even in the presence of income targeting Köszegi and Rabin (2006) and it is true in our paper and in Chen and Sheldon (2015) in the context of Uber drivers. Second, there must be a component of the marginal cost of work on any given day that is observed by the worker but not by the econometrician. It is hard to imagine any worker not subject to variations on unobserved factors such as sleep, health, eating, or general energy that affect the marginal cost of work. As long as we cannot control for a meaningful amount of this variation, dynamic selection will be a concern.

Take for example the cab drivers like in Thakral and Tô (2019) and Farber (2015). Even though cab drivers do not have an explicit instantaneous wage rate, there are factors that a driver could use to predict whether their instantaneous wage would be relatively high or low. For example, drivers may understand that traffic or unusual events in the city may increase the demand for cabs, ultimately increasing their (latent) instantaneous wage. If taxi drivers delay ending the shift when they perceive to be in a high wage period, it would not be possible to separate income targeting from dynamic selection.

The dynamic selection problem is even more salient in settings with the presence of dynamic pricing, such as ride-hailing drivers or delivery couriers, since nominal wages fluctuate explicitly and unexpectedly. Because drivers are likely to extend their shifts in response to dynamic pricing, and because dynamic pricing is very salient, unobservable labor supply shocks may induce selection in the same fashion as described before. A recent study of Uber drivers labor supply (Chen and Sheldon, 2015) suggests that drivers are indeed less likely to quit under dynamic pricing. They estimate that the odds of quitting is roughly halved in periods subject to the price multiplier, a magnitude similar to how High piece-rates affect quitting probabilities in our setting. They also found that cumulative fares increases the probability of quitting for the day, although there is reason to suspect this result can be potentially driven by dynamic selection. Because such settings are potential avenues for future research on flexible labor supply due to large data availability, the dynamic selection problem must be carefully taken into consideration when interpreting results.

To illustrate our argument in the context of the income targeting literature, we re-estimate our results for the stopping decision using a very similar specification to that of Thakral and Tô (2019) and Farber (2015). We present the results in Figure 3, which shows the impact of a 10% increase of cumulative income on the stopping probability at different moments in the shift, as estimated by Equation 3.2. For ease of interpretation, we divide the estimates by the probability of stopping at the corresponding moment in the shift.

Not accounting for dynamic selection, Figure 3a shows that an increase in cumulative income earlier in the day is associated with an increase in the probability of stop working at any moment of the shift. For example, if we take the moment of the day representing the median shift duration (7.5 working hours), a 10% increase in cumulative earnings is associated with an increase of roughly 9.5% in the stopping probability ($p = 0.011$).[3] Relative estimates are stable and marginally significant throughout the windows with an increase in magnitude towards advanced hours of the shift. In comparison, in a sample of NYC cab drivers, Thakral and Tô (2019) report a 3.3% increase in the stopping probability in the median shift duration, while Farber (2015) reports a 9% increase.

However, these results do not correct for the negative bias from dynamic selection we observed in Section 4.1. To address that, we re-estimate Equation 3.2 using the strategy of conditioning the sample to those participants with a Low piece-rate in the immediately previous session and show the results in Figure 3b. Similarly to the results using output as the outcome variable, the evidence for negative income effects is much less compelling when controlling for dynamic selection. Virtually all coefficients are close to zero in magnitude and not statistically significant, although we lose precision by reducing the sample size. The only exception is the coefficient associated with the last moment of the day, when only 10% of participants are still working, which is still large and significant. Therefore, in the most generous interpretation of our results, we only find evidence of negative income effects in the extensive margin in a small fraction of our observations.

[3] The coefficient of the regression is 0.000095, which multiplied by 10% of the average cumulative income, Rs. 42.8, and divided by the probability of stopping at a given page after 7.5 hours of work, roughly 0.042, yields 0.095.

A limitation of our approach to dealing with dynamic selection in the extensive margin is that we have to sacrifice the income variation from the latest past incentive session. Indeed, Thakral and Tô (2019) argues that daily reference points are adaptive throughout the day, implying that recently earned money should be more relevant for the stopping decision. Under their theory, ignoring the most recent source of income variation would naturally decrease the estimates of the income effect. Although an income targeting theory with adaptive targets is not separably identifiable from the dynamic selection issue, the results from Table 2 at least reject a theory of adaptive reference points in which income shocks that happen more than 30 minutes in the past still affects contemporaneous labor supply.

A second strand of the literature investigate the existence of income effects on labor supply using exogenous variation from monetary windfalls (Dupas et al., 2019; Andersen et al., 2018). Unlike the papers that use intra-day wage variation, these papers find no evidence of income effect in labor supply.

Although these two strands are different in many ways, a key difference between them is that monetary windfalls are not vulnerable to dynamic selection issues, which might explain why they find no evidence of negative income effect. On the other hand, an alternative explanation for the null effect is that monetary windfalls do not enter the same mental account as earned income. The leading theories of negative income effects on labor supply explain it using reference-based preferences (Camerer et al., 1997; Köszegi and Rabin, 2006). How the reference points are formed is, however, an open question. In light of the large body of evidence showing that people tend to have separate mental accounts for different types of earnings and spending (Hastings and Shapiro, 2013), it is conceivable that monetary windfalls do not enter the same mental account as earned income. In this case, the authors could have found negative income effects if they had a variation on earned income.

In our setting, we replicate the findings from both strands of literature when ignoring dynamic selection. When we do that, we find evidence of negative income effects from earned income when ignoring the dynamic selection issue, but not income effect from a monetary windfall. However, once we take dynamic selection into account, the supposed negative income effect from earned income goes away. Thus, we interpret this result as suggestive evidence that the discrepancy in results from these two strands of literature might come from the failure in accounting for dynamic selection, instead of because monetary windfalls do not enter the same mental account as earned income.

5 Conclusion

We showed that the existing evidence of income targeting in labor supply could be caused by a dynamic selection issue, which has not been noted in the literature. Leveraging data from an RCT with exogenous intra-day variation on wages and using the state-of-the-art techniques from the income targeting literature (Thakral and Tô, 2019; Farber, 2015), we find evidence of negative income targeting in different margins of labor supply. However, this result is driven entirely by the dynamic selection issue. When taking dynamic selection into account or estimating income effects from monetary windfalls, which are not affected by dynamic selection, there is no evidence of negative income effect.

Moreover, we also bridge a gap between two strands of the income targeting literature. While most papers using intra-day variation on wages find evidence of negative income targeting Thakral and Tô (2019); Farber (2015); Crawford and Meng (2011), the papers using monetary windfalls find no evidence of negative income effect Dupas et al. (2019); Andersen et al. (2018). We replicate the latter findings in our setting, using exogenous monetary windfalls in our setting. This result suggests that the difference in results from these strands of the literature might be explained by failure in the former to account for dynamic selection effects.

The theoretical and empirical results from this paper point out that the estimation of income targeting using wage variation should address the issue of dynamic selection. We propose two different strategies to deal with the issue. First, we show that focusing on the participants who received a very low instantaneous wage in the immediate past allows for estimation of income effect from variations on wages further in the past. This solution is hard to apply directly in settings in which the instantaneous wages are not observable, but could be used in the context of ride-sharing, in which drivers face different multipliers. A second solution is to focus on instances of the data for which the extensive margin is shut-off and investigate income targeting in other margins of labor supply, such as effort and breaks during the shift.

Admittedly, these solutions will not always be feasible in real-world settings. As an alternative, future research using wage variation to study income effects could deal with the dynamic selection issue using a structural model that incorporates dynamic selection explicitly. Such an approach has the usual disadvantages associated with structural estimation, such as being more opaque, harder to implement, and identified out of functional form. However, they still must be employed if there is no other way to address the dynamic selection issue.

Bibliography

Andersen, S., Brandon, A., Gneezy, U., and List, J. A. (2018). Toward an Understanding of Reference-Dependent Labor Supply: Theory and Evidence From a Field Experiment. *NBER Working Paper*, 60637(No.20695):1–38.

Bessone, P., Rao, G., Schilbach, F., Schofield, H., and Toma, M. (2019). Sleepless in Chennai: The Consequences of Sleep Deprivation Among the Urban Poor. *Working Paper*, (14):1–55.

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.

Card, D. and Hyslop, D. R. (2005). Estimating the effects of a time-limited earnings subsidy for welfare-leavers. *Econometrica*, 73(6):1723–1770.

Chang, T. and Gross, T. (2014). How many pears would a pear packer pack if a pear packer could pack pears at quasi-exogenously varying piece rates? *Journal of Economic Behavior and Organization*, 99:1–17.

Chen, M. K. and Sheldon, M. (2015). Dynamic Pricing in a Labor Market: Surge Pricing and the Supply of Uber Driver-Partners. *Working Paper*, pages 1–18.

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):1912–1932.

Dupas, P., Robinson, J., and Saavedra, S. (2019). The Daily Grind: Cash Needs, Labor Supply and Self-Control. *Working Paper*.

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):1069–1082.

Farber, H. S. (2015). Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons From Cab Drivers. *The Quarterly Journal of Economics*, 130(4):1975–2026.

Fehr, E. and Goette, L. (2007). Do Work More if Wages Are High? Evidence from a Do Workers Randomized Field Experiment. *American Economic Review*, 97(1):298–317.

Ham, J. C. and Lalonde, R. J. (1996). The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training. *Econometrica*, 64(1):175.

Hastings, J. E. and Shapiro, J. M. (2013). Fungibility and consumer choice: evidence from commodity price shocks. *Quarterly Journal of Economics*, 128(November):1449–1498.

Jonason, D. and Wållgren, J. (2013). Labor Supply of Stockholm Cab Drivers : Revisiting the Evidence. *Working Paper*, (May).

Köszegi, B. and Rabin, M. (2006). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics*, CXII(August):1133–1165.

O'Donoghue, T. and Sprenger, C. (2018). *Reference-Dependent Preferences*, volume 1. Elsevier B.V.

Oettinger, G. S. (1999). An Empirical Analysis of the Daily Labor Supply of Stadium Vendors. *Journal of Political Economy*, 107(2):360–392.

Read, D., Loewenstein, G., and Rabin, M. (1999). Choice Bracketing. *Journal of Risk and Uncertainty*, 19(1-3):171–197.

Thakral, N. and Tô, L. T. (2019). Daily Labor Supply and Adaptive Reference Points. *American Economic Review*, Forthcomin.

Figures and Tables

Figure 1: Example of a typing screen

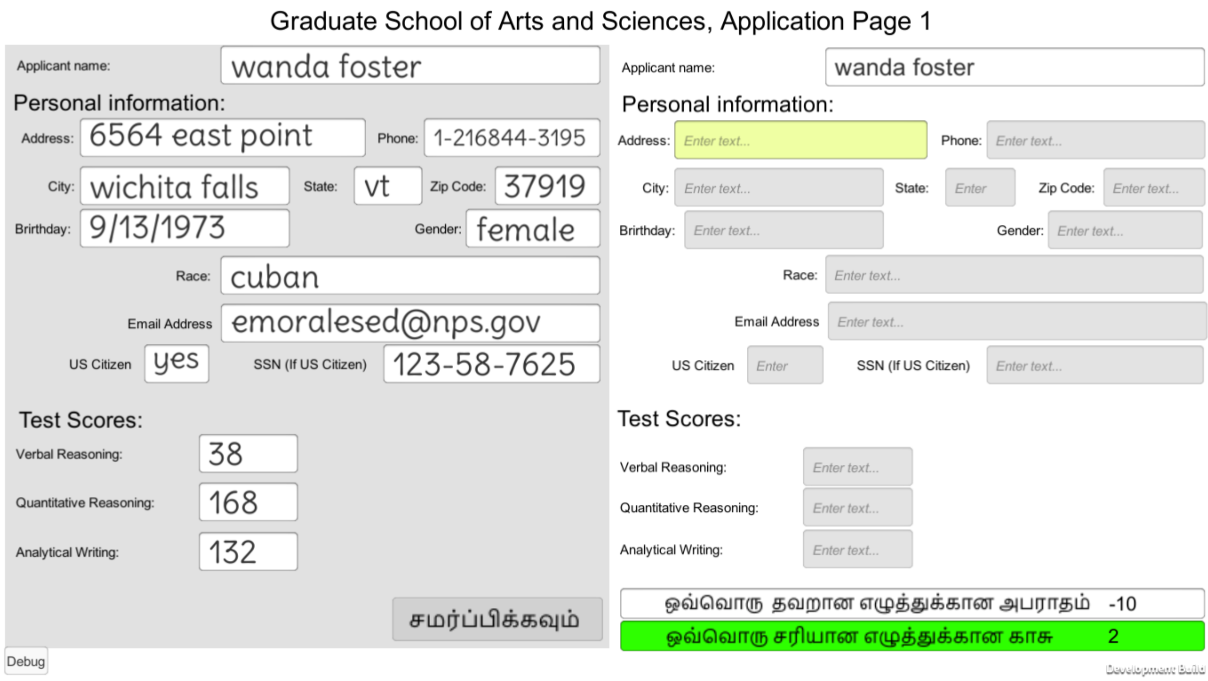
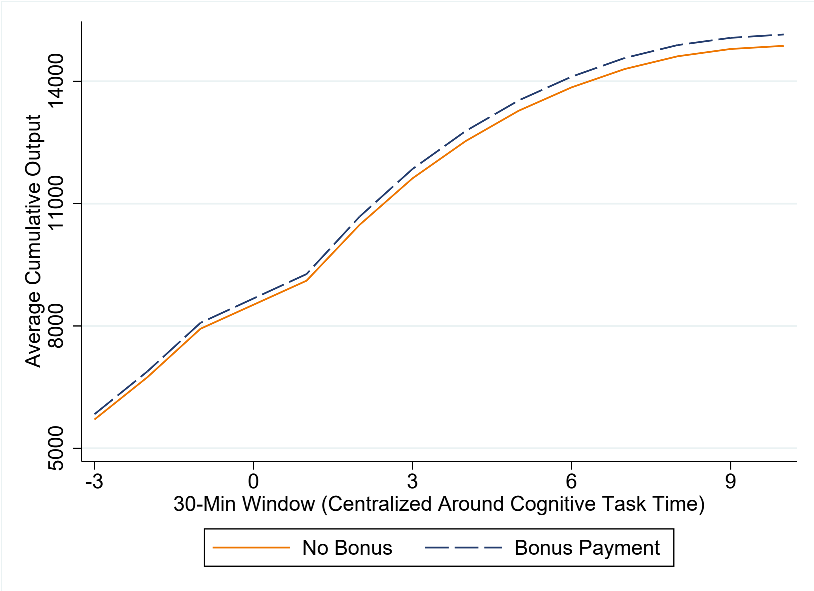
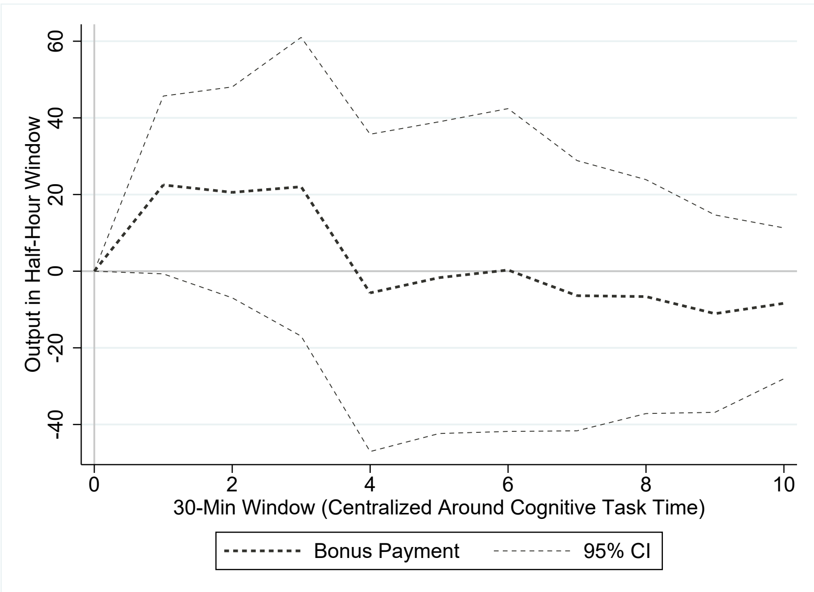


Figure 2: Effects of monetary windfalls - doubling task payment

(a) Raw average output over day



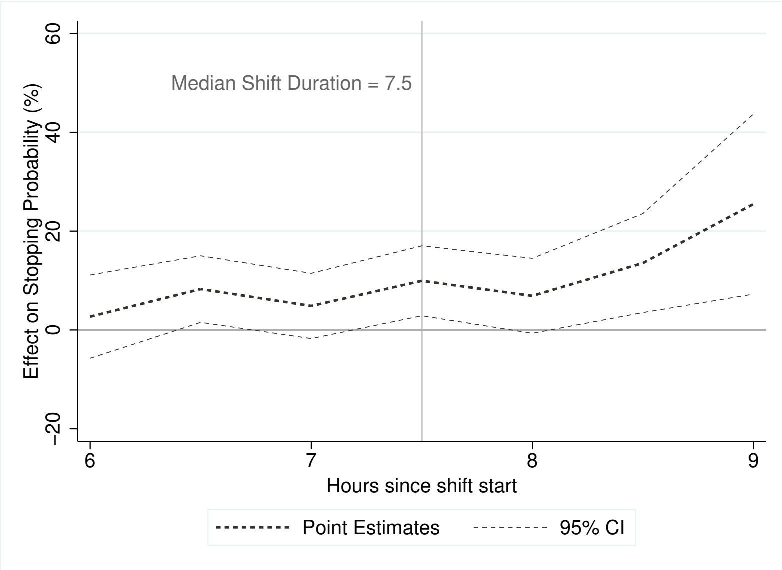
(b) Effect on output by half-hour window



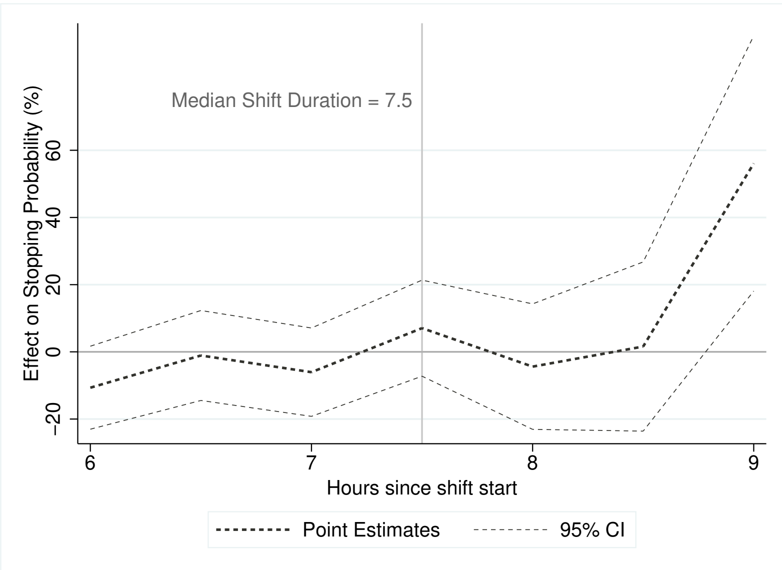
Notes: These figures compares the output of participants who received a monetary windfall during their shift to those who did not. In figure (a), we present the raw average data-entry output of each group. In figure (b), we present estimates from equation (4.1), in which each point represents the impact of the monetary windfall on output in 30 minutes windows after the time of the monetary windfall.

Figure 3: Effect of cumulative income on worker's stopping probability

(a) Without correcting for dynamic selection



(b) Correcting for dynamic selection



Notes: These figures present the effect in the probability of stopping in response to a 10% increase in cumulative earnings according to equation (3.2). The figure in (a) uses all the data from long shifts, while the figure in (b) corrects for dynamic selection by conditioning the sample to the participants who received a Low piece rate in the immediately previous time period. Each point-estimate is taken over a 30-minute window centralized at {6, ..., 9} hours since the shift began. Errors are clustered at the participant level.

Table 1: Summary statistics at daily and session level

Panel A: Participant-level

	(1)						
	Mean	Std. Deviation	Median	p5	p25	p75	p95
Age	34.95	7.22	33.00	27.00	29.00	40.00	50.00
Years of Education	10.18	2.88	10.00	5.00	9.00	12.00	15.00
Number of Children	1.35	1.06	2.00	0.00	0.00	2.00	3.00
Female	0.66	0.47	-	-	-	-	-
Used Computer Before	0.28	0.45	-	-	-	-	-
Observations	452						

Panel B: Shift-level (long days)

	Mean	Std. Deviation	Median	p5	p25	p75	p95
Output	18328.79	11776.47	15490.00	5297.00	10229.50	23135.50	41755.00
Typing Time (Hr.)	5.29	1.24	5.31	3.21	4.48	6.15	7.30
Productivity (Output/Hr)	3410.75	1834.51	3034.62	1201.66	2079.21	4310.80	7134.88
Voluntary Pauses (Min.)	15.95	22.88	9.12	0.00	3.67	19.26	54.50
Scheduled Pauses (Min.)	111.18	22.41	115.00	80.00	95.00	125.00	154.80
Typing Earnings	343.10	175.96	304.01	142.22	224.30	418.64	692.28
Number of Incentive Sessions	10.89	2.48	11.00	7.00	9.00	13.00	15.00
Number of Pages	90.18	57.31	76.00	27.00	51.00	113.00	205.00
Time in Office	8.05	1.12	8.05	6.20	7.40	8.75	9.83
Observations	4968						

Panel C: Session-level (long days)

	Mean	Std. Deviation	Median	p5	p25	p75	p95
Output	1588.02	977.87	1404.00	343.00	895.00	2055.00	3558.00
Productivity (Output/Min)	55.20	32.24	48.46	16.50	32.00	70.35	120.81
Typing Time (Min.)	29.13	7.26	29.92	13.75	29.58	30.77	34.53
Voluntary Pauses (Min.)	1.47	5.97	0.00	0.00	0.00	0.07	7.80
Scheduled Pauses (Min.)	10.20	17.10	0.00	0.00	0.00	15.00	40.00
Typing Earnings	31.42	20.80	24.37	9.28	16.81	42.30	73.09
Performance Earnings	20.81	20.29	13.32	1.26	5.68	31.22	62.12
Attendance Earnings	10.61	2.50	10.84	5.00	10.76	11.20	12.55
Observations	54286						

Notes: This table presents summary statistics for three different units of observations: participants, shifts, and incentive sessions. The panels in the shift and incentive session levels only include long shifts (see Section 2).

Table 2: Effects of past High sessions on output

	Long Days			Short Days			Long, conditional on past wage Low		
	(1) Output	(2) Work Breaks	(3) Quit	(4) Output	(5) Work Breaks	(6) Quit	(7) Output	(8) Work Breaks	(9) Quit
High	284.71	-1.06	-0.08	156.60	-0.65	-0.00	261.39	-0.84	-0.08
	(12.13)	(0.10)	(0.00)	(12.11)	(0.09)	(0.01)	(12.23)	(0.09)	(0.00)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.57]	[0.00]	[0.00]	[0.00]
High Lag 1	-56.29	0.37	0.03	9.17	0.22	-0.01	0.00	0.00	0.00
	(6.17)	(0.08)	(0.00)	(9.54)	(0.08)	(0.01)	(.)	(.)	(.)
	[0.00]	[0.00]	[0.00]	[0.34]	[0.00]	[0.08]	[.]	[.]	[.]
High Lag 2	-18.22	0.22	0.01	-1.17	-0.01	0.00	15.47	-0.07	-0.00
	(5.12)	(0.07)	(0.00)	(9.08)	(0.08)	(0.01)	(6.40)	(0.08)	(0.00)
	[0.00]	[0.00]	[0.01]	[0.90]	[0.93]	[0.44]	[0.02]	[0.36]	[0.28]
High Lag 3	-11.45	0.13	0.00	-3.23	0.05	0.00	-6.98	0.02	-0.00
	(4.62)	(0.07)	(0.00)	(8.17)	(0.08)	(0.01)	(6.28)	(0.07)	(0.00)
	[0.01]	[0.06]	[0.94]	[0.69]	[0.55]	[0.71]	[0.27]	[0.76]	[0.79]
High Lag 4	-12.48	0.17	0.00	14.58	-0.22	-0.01	-5.85	0.08	-0.01
	(5.61)	(0.07)	(0.00)	(9.02)	(0.07)	(0.01)	(6.16)	(0.07)	(0.00)
	[0.03]	[0.02]	[0.31]	[0.11]	[0.00]	[0.24]	[0.34]	[0.23]	[0.15]
Excluded Group Mean	1470.75	1.87	0.14	1460.20	1.64	0.16	1512.75	1.70	0.11
Excluded Group Std.	941.48	6.24	0.34	922.50	4.57	0.37	925.67	5.00	0.31
Observations	33716	33716	33716	9628	9628	9628	16181	16181	16181
Participants	452	452	452	452	452	452	452	452	452

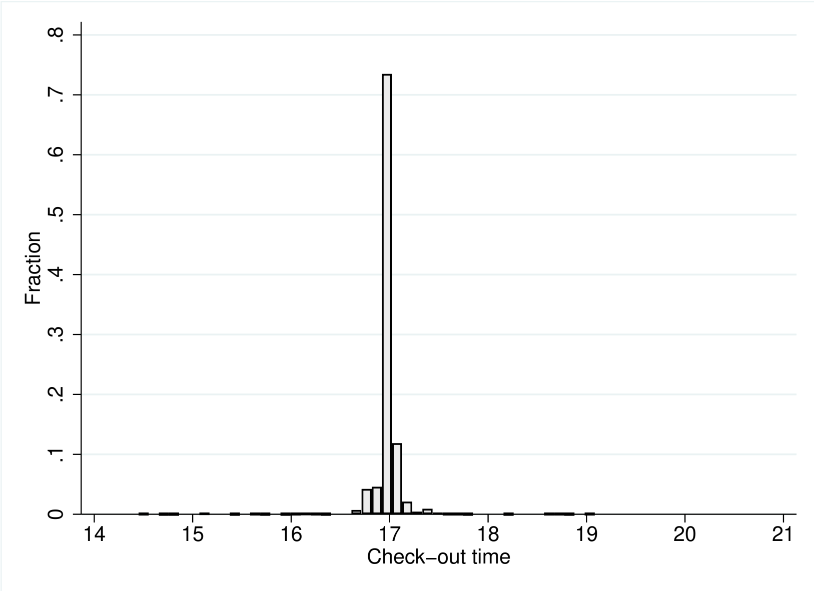
Notes: This table presents the results of equation (3.1), where we estimate the effect of present and past High piece rates on labor supply outcomes at the session-level. Columns 1 to 3 only uses observations from long workdays, while Columns 4 to 6 uses observations from short workdays. Columns 7 to 9 implement our solution to the dynamic selection problem by conditioning the sample to those who have received a Low piece-rate in the preceding session. Standard errors are clustered at the participant-level and are displayed in parenthesis, while p-values are displayed in brackets.

Appendix

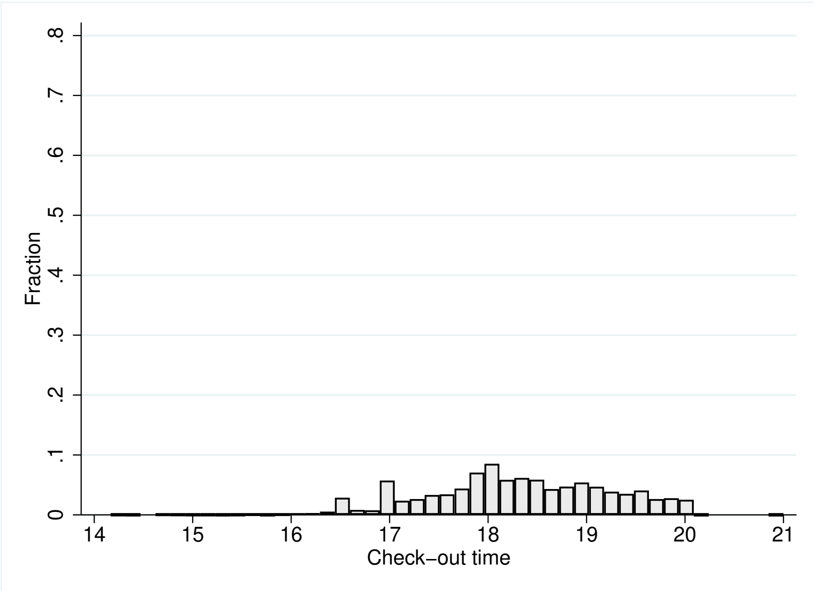
APPENDIX A – Extra Tables and Figures

Figure A1: Distribution of check-out times of typing task

(a) Short days



(b) Long days



Notes: These figures present histograms of check-out times, defined as the last recorded keystroke in the typing task software. In figure (a), we consider only short workdays, while figure (b) considers only long workdays.

APPENDIX B – Proofs

Proof Lemma 3.1:

Let $U^1(t,w,\varepsilon) \equiv U(t,w,\varepsilon,d_t=1) = u(t,w,\varepsilon)+V^e(t+1,\varepsilon)$. The agent chooses $d^*(t,w,\varepsilon) = 1$ iff

$$U^1(t,w,\varepsilon) \equiv u(t,w,\varepsilon) + V^e(t+1,\varepsilon) \geq 0$$

Note that by assumption $u_\varepsilon(t,w,\varepsilon) < 0$ for any (t,w). This implies that $V^e(t+1,\varepsilon)$ is also (weakly) decreasing on ε, since any agent facing two different ε can always choose exactly the same $(d(t,w,\varepsilon))$ path and receive a higher inter-temporal utility under the smaller value of ε. Thus, $U^1(t,w,\varepsilon)$ is also decreasing on ε. This implies that there is at most one point that $U^1(t,w,\varepsilon)$ equals zero, $\overline{\varepsilon}(t,w)$, and that $U^1(t,w,\varepsilon) \geq 0$ iff $\varepsilon \leq \overline{\varepsilon}(t,w)$.

We now prove the comparative static results. Take two wages such that $w' > w$. We have that

$$U^1(t,w,\overline{\varepsilon}(t,w)) \equiv u(t,w,\overline{\varepsilon}(t,w)) + V^e(t+1,\overline{\varepsilon}(t,w)) = 0 \tag{B.1}$$

$$U^1(t,w',\overline{\varepsilon}(t,w')) \equiv u(t,w',\overline{\varepsilon}(t,w')) + V^e(t+1,\overline{\varepsilon}(t,w)) = 0 \tag{B.2}$$

Because $u_w > 0$, $u(t,w',\varepsilon) > u(t,w,\varepsilon)$ for any t,ε. Since, $u(t,w,\overline{\varepsilon}(t,w)) = u(t,w',\overline{\varepsilon}(t,w'))$ and $u_\varepsilon < 0$, this means that $\overline{\varepsilon}(t,w') > \overline{\varepsilon}(t,w)$.

Take now two time periods such that $t' > t$. Again,

$$U^1(t,w,\overline{\varepsilon}(t,w)) \equiv u(t,w,\overline{\varepsilon}(t,w)) + V^e(t+1,\overline{\varepsilon}(t,w)) = 0 \tag{B.3}$$

$$U^1(t',w,\overline{\varepsilon}(t',w)) \equiv u(t',w,\overline{\varepsilon}(t',w)) + V^e(t'+1,\overline{\varepsilon}(t',w)) = 0 \tag{B.4}$$

We have that both $u_t(t,w,\varepsilon) < 0$ and $V_t^e(t,\varepsilon) \leq 0$ for any t,w,ε, implying that $U^1(t',w,\varepsilon) < U^1(t,w,\varepsilon)$ for any ε. Since, (i) $U^1(t,w,\overline{\varepsilon}(t,w)) = U^1(t',w,\overline{\varepsilon}(t',w))$ and (ii) $V_t^\varepsilon(t,\varepsilon) \leq 0$, it must be the case that $\overline{\varepsilon}(t,w) > \overline{\varepsilon}(t',w)$. □

Proof of Lemma 3.2

Assume that we observe the wage history $\boldsymbol{w}^t$ and that $d_t = 1$, implying that $d_{t'} = 1$ for all $t' < t$. This information set is equivalent to stating that $\varepsilon \leq \hat{\varepsilon}(\boldsymbol{w}^t)$ for all t, where

$$\hat{\varepsilon} \equiv \min_{t' \in \{1,2,\ldots,t\}} \{\bar{\varepsilon}(t', w_{t'})\}$$

If $w_{t'} = w_H$ or $w_{t'} = w_L$ for every $t' \leq t$, then $\hat{\varepsilon} = \bar{\varepsilon}(t, w_t)$, since by Lemma 3.1, $\bar{\varepsilon}(t, w)$ is decreasing on t and w_t is fixed. If $w_t = w_L$, then again $\hat{\varepsilon} = \bar{\varepsilon}(t, w_t) = \bar{\varepsilon}(t, w_L)$, since $w_t \leq w_{t'}$ for $t' \leq t$. Finally, if $w_t = w_H$ and T^L is well-defined, then by the same arguments, $\bar{\varepsilon}(t, w_H) < \bar{\varepsilon}(t', w_H)$ for any $t' < t$ and $\bar{\varepsilon}(t^L, w_L) < \bar{\varepsilon}(t', w_L)$ for any $t' < t^L$. The only ambiguous expression is then whether $\bar{\varepsilon}(t^L, w_L) < \bar{\varepsilon}(t, w_H)$. $\square$.

Proof of Proposition 3.1

Fix an agent at a given period t such that $d_{t-1} = 1$. Consider now the decision to stay conditional on $d_{t-1} = 1$. By Lemma 3.1, $d_{t-1} = 1$ implies that, for every $t' \leq t-1$, $\varepsilon \leq \bar{\varepsilon}(t', w_{t'})$. This implies that when we observe this agent at period t, it must be that her daily shock ε is at least smaller than $\hat{\varepsilon}(\mathbf{w}^{t-1})$, which is summarized by Lemma 3.2. We then have that conditional on the information set available at t, probability of continuining to work is:

$$Pr(d_t = 1|d_{t-1}, w_t) = F\left(\bar{\varepsilon}(t, w)|\varepsilon \leq \hat{\varepsilon}(\mathbf{w}^{t-1})\right)$$

We now consider $\hat{\varepsilon}(\mathbf{w}^{t-1})$ under $w_t = w_L$ and $w_t = w_H$. Let $T^S = \max_{t'} \{t'|t' \leq t-1, w_{t'} = w_S\}$ for $S = \{H, L\}$. Following Lemma 3.2, we have that:

$$\hat{\varepsilon}(\mathbf{w}^{t-1}) = \begin{cases} \bar{\varepsilon}(t-1, w_L), \text{ if } w_{t-1} = w_L \\ \min\left\{\bar{\varepsilon}(t-1, w_H), \bar{\varepsilon}(T^L, w_L)\right\} \text{ if } w_{t-1} = w_H \end{cases}$$

We then have that:

$$Pr(d_t = 1|d_{t-1}, w_t, w_{t-1} = w_H) = F\Big(\bar{\varepsilon}(t, w_t)|\varepsilon \leq \min\left\{\bar{\varepsilon}(t-1, w_H), \bar{\varepsilon}(T^L, w_L)\right\}\Big)$$
$$Pr(d_t = 1|d_{t-1}, w_t, w_{t-1} = w_L) = F\Big(\bar{\varepsilon}(t, w_t)|\varepsilon \leq \bar{\varepsilon}(t-1, w_L)\Big)$$

We now show that $\bar{\varepsilon}(t-1, w_L) < \min\left\{\bar{\varepsilon}(t-1, w_H), \bar{\varepsilon}(T^L, w_L)\right\}$. Note that $\bar{\varepsilon}(t-1, w_L) < \bar{\varepsilon}(t', w_L)$ for $t' < t$ since $\bar{\varepsilon}$ is decreasing in t. Moreover, since $\bar{\varepsilon}$ is increasing in w, we have $\bar{\varepsilon}(t-1, w_L) < \bar{\varepsilon}(t-1, w_H)$. We complete the proof by noting a property of conditional probabilities, $F(x|x \leq a) < F(x|x \leq b)$ for $a > b$. $\square$.