Bachelor thesis
Course 659
Stockholm School of Economics
Supervisor: Juanna Schrøter Joensen
15th of May 2013

Presented on the 29th of May 2013
Discussants: Pontus Liljeblad and Jasmin Moshfegh
Examiner: Örjan Sjöberg

Labor Supply of Stockholm Cab Drivers: Revisiting the Evidence

David Jonason* and Joel Wållgren**

ABSTRACT: In this paper, panel data on the hours worked and wage rates of taxicab drivers in the city of Stockholm are used to test two competing theories of labor supply: the standard neoclassical model, which predicts positive wage elasticity, and the target income model, which predicts negative wage elasticity. Particularly comprehensive and precise data sets allow us to revisit the evidence presented in recent literature that focuses on professions in which workers are free to set their own hours. In contrast to Camerer et al. (1997), this paper identifies significant positive autocorrelation in the wage across days and positive wage elasticity estimates for a number of specifications, implying that the labor supply behavior of Stockholm taxicab drivers is inconsistent with a one-day target income hypothesis. This conclusion demonstrates that further attention must be given to factors influencing the decision-making time horizon of labor supply, as well as to the effect of the extensive margin on wage elasticity estimates.

Keywords: labor supply, income targeting, life-cycle model, wage elasticity

JEL Classification: J22, C23, J30

*22046 **22218

Acknowledgments

A special gratitude is owed to the Market Manager and Technical Manager of the data-providing taxi franchise company for their patience and enthusiasm for this project, as well as the Stockholm cab drivers that patiently answered our strangely specific questions on many late nights. We are most thankful for the assistance and guidance of our supervisor Juanna without whose econometric expertise and pedagogical skills this paper would never have seen the light of day. We would also like to thank Professor Kurt Weyland for his valuable comments and suggestions. Last but not least, special thanks to Katie, Patricia, and Lotta.

Table of Contents

1. Introduction

The extent to which remuneration influences labor supply is a central topic in labor economics and public policy discourse. This issue is of crucial importance for the evaluation of such government policies as tax and transfer programs, which depend on reliable estimates of the sensitivity of labor supply to changes in income levels. In order to avoid sub-optimal practices, individuals and organizations alike would benefit from a better understanding of the factors that influence our labor supply decisions. Empirical research on this topic has focused on measuring the wage elasticity of labor supply, which captures the extent to which labor supply responds to a change in the wage rate in a given time period. In this paper, panel data on hours worked and wage rates of taxicab drivers in the city of Stockholm are used to test two competing theories of labor supply: the standard neoclassical model and the target income model.

According to the neoclassical life-cycle model of labor supply, individuals derive utility from total lifetime consumption and hours of labor. Since individuals are free to reallocate resources over time through borrowing and lending, the budget constraint incorporates income and expenditure over an entire lifetime. By this logic of intertemporal substitution, current labor supply depends on all past and expected future wage rates. In a multi-period maximization problem, a transitory change in the wage rate has negligible impact on the life-cycle wealth. However, if the substitution effect of the wage change on hours worked is positive, a transitory wage increase should lead to an increase in labor supply (Lucas and Rapping 1969). Simply put, when wages are temporarily high the opportunity cost of leisure rises, in turn making individuals more willing to work. This provides a solid test of the neoclassical model of labor supply: the wage elasticity estimate, obtained by using the daily number of hours as the primary dependent variable and the average wage the driver received during that day as the main explanatory variable, is predicted to be positive.

In contrast, the target income labor supply hypothesis is based on fundamentally different assumptions about worker preferences. It stems from prospect theory, a model of risk attitudes first developed by Kahneman and Tversky (1979, 1991) according to which individuals diverge from the strict rationality implied by the neoclassical theory and instead use simple heuristics in daily decision-making. In the target income model, workers set a fixed income target over a short time horizon, and adjust their working hours to meet this target. With utilities being "reference-dependent", gains (outcomes above the target) and losses (outcomes below the target) might be treated differently. Indeed, significant experimental evidence suggests that people experience more displeasure from a loss than they derive pleasure from equally large gains – a phenomenon commonly referred to as loss-aversion. By this logic, workers determine how much they would like to earn on a given day, their reference point, and are inclined to stop working once this level has been reached. The consequence is the somewhat counterintuitive idea that individuals work fewer hours when wages are subject to temporary increases and more hours when wages experience temporary decreases. If workers do in fact make labor supply decisions based on an income reference point, the target income hypothesis predicts negative wage elasticity.

Because driving a cab is one of the few professions that fulfills the assumption inherent in most models of labor supply that workers are free to choose their own hours of work, much of the research on the topic uses taxicab drivers as subjects (Barberis

2013). While a focus on such a narrow segment of the labor market might seem odd, the taxicab driving profession has provided an ideal "laboratory" to conduct unusually clean tests of the validity of the two competing models of labor supply, since they are also regularly exposed to transitory changes in wage and free to set their own working hours.

Previous research on the subject – most notably the pioneering 1997 study *Labor Supply of New York City Cabdrivers: One Day at the Time* by Camerer, Babcock, Loewenstein and Thaler – have found negative wage elasticities in data sets covering the hours and earnings of taxicab drivers in cities such as New York City (NYC), and Singapore (Chou 2002). These results contrast heavily with the predictions of the standard neoclassical model of labor supply, and suggest that drivers quit early on days when it is easy to make money, whereas they work longer hours on days in which fares are scarce. Camerer et al. (1997), finding the wages of NYC taxicab drivers to be correlated within days but uncorrelated between days, attribute their findings to the idea that taxicab drivers make labor supply decisions "one day at a time" by setting loose daily income targets. In fact, explaining negative wage elasticity estimates requires the assumption that workers have a decision-making horizon no longer than one day: as explained further in this paper, targeting at even a two-day decision-making horizon would allow workers to substitute their labor supply across days. We can therefore make the distinction between a general income target model and the one-day income target hypothesis that is the focus of this literature. Some argue that a number of econometric issues in these studies could potentially bias the estimated wage elasticity downward (Stafford 2013; Farber 2005; Oettinger 1999), and that the negative estimates might therefore be the unfortunate consequence of flawed estimation methodology. After fifteen years of research on the topic, wide disagreement still remains over the basic mechanisms explaining the labor supply decision of taxicab drivers. It is of great importance to resolve this source of discord for continued progress in the field of labor supply decision-making (Farber 2005).

A decisive factor as well as a prime challenge in the field has been the access to comprehensive and reliable empirical data. Previous studies have relied on self-reported hours and fares by participating subject, data sets of questionable panel character and very limited time horizons. We therefore believe that revisiting the evidence from the taxicab driver labor market with extensive data now made possible by modern digital log systems can provide additional evidence and refinement to the present understanding of wage elasticity. By analyzing data from 47 Stockholm taxicab drivers' daily shifts over a period of 3 months, as well as trip-by-trip data of 22 drivers over a period of ten days, this study hopes to take a closer look at the empirical evidence to hopefully help settle the debate. Indeed, compared to most previous data used in the field, our Stockholm data set not only leaves less room for measurement error, but also allows us to examine the daily participation decision by observing the days during which individual taxicab drivers do not work. To our knowledge, it is the first study in its kind to have access to data that enables looking at the participation decision of taxicab drivers. Additionally, the set of cultural and market specifications associated with the city of Stockholm could add some nuance to previous analysis and potentially help control for certain variables such as large amounts of customer tips or daily liquidity constraints.

Building on previous studies, the purpose of this paper is to revisit the evidence on labor supply decision theory on extensive and previously inaccessible data in order to test if it supports either neoclassical or a one-day reference-based models. Thus, the central research question can be formulated as follows:

Is the labor supply behavior of Stockholm taxicab drivers consistent with a one-day target income hypothesis?

Several hypotheses are inherent in this question. A one-day target income model implicates that the taxicab driver's wage per hour exhibits positive autocorrelation within days but is uncorrelated between days. Most importantly, it implicates that Stockholm taxicab drivers display negative wage elasticity.

Our results suggest that there is a positive relationship between hourly wage within days but also a significantly positive autocorrelation between days. The labor supply of Stockholm taxicab drivers does not display negative wage elasticity in various specifications, but instead positive estimates in line with the standard neoclassical model of labor supply. In addition, the analysis suggests that the stopping behavior of Stockholm taxicab drivers depends less on an income target than on reaching a cumulative hours target. Based on these results, this paper questions the one-day time horizon assumption made in much of the previous research on the topic and concludes that the labor supply of taxicab drivers might not after all be such a clear-cut example of prospect theory "in the wild". However, the results do not indicate that the general target income hypothesis does not hold, only that it might be more difficult to test the theory on the labor supply decision of taxicab drivers than previously assumed.

The paper is organized as follows: Section 2 covers the previous research on labor supply decisions and outlines a mathematical model describing the foundation of the target income hypothesis. Section 3 contains a review of the econometrical methods used to address the research question, while section 4 describes the Stockholm taxicab driver data collected for the study. Thereafter, section 5 presents the empirical results, then discussed in section 6 where specific factors in the Stockholm taxi market are examined before taking on a more universalistic approach by reexamining the data. Lastly, in section 7 and 8, we present concluding remarks as well as a summary of the contributions of this paper.

2. Previous Research

In this section, we will examine the previous empirical research, focusing on studies of the labor supply decisions of taxicab drivers. Secondly, a mathematical modeling of the labor supply decision facing taxicab drivers explains the predicted outcome of negative wage elasticity. Finally, criticism to studies displaying negative wage elasticity as well as further empirical and theoretical developments are presented.

Evidence of Negative Wage Elasticity

The standard theory of labor supply makes a straightforward prediction: hours of labor should be positively related to transitory fluctuations in wages. However, despite the existence of significant literature estimating the wage elasticity of labor supply, this prediction has proved difficult to verify. Early studies (surveyed by Pencavel 1986; Killingsworth and Heckman 1986; Blundell and MaCurdy 1999) related annual changes in hours worked to annual changes in the average hourly wage, only to find wage elasticity estimations that were small, often statistically insignificant, and in some cases even slightly negative. Such studies have used a range of different types of data such as aggregate data (Mankiw et al. 1985), cohort data (Browning et al. 1985), and panel data (Altonji 1986). Estimates of the intertemporal substitution elasticity range from -0.07 to 0.45 for men, with the central tendency of 0.20. Estimates of women's labor supply elasticity have been somewhat larger, though still considerably less than one (Killingsworth and Heckman 1986).

However, this literature confronts substantial obstacles and often falls prey to inherent difficulties in avoiding measurement errors in wage and labor supply variables. Wage changes are rarely purely transitory, often serially correlated, and nominal wages suffer from downward rigidity. Furthermore, a main criticism of this literature is that while the standard neoclassical model assumes workers to be free to set their hours in response to changes in the wage, substantial evidence points to the fact that is rarely the case (Farber 2005). Empirical data consequently shows "lumpy" distributions of hours, with substantial fractions of workers reporting weeks of precisely 40 hours (Farber 2005). If workers are not fully capable of adjusting their hours worked in response to annual wage changes, estimates are biased towards zero.

In their now prominent paper Camerer et al. (1997) find that it is possible to circumvent many of these issues by analyzing the labor supply decisions of taxicab drivers. New York City cab drivers, the object of Camerer et al.'s (1997) research, typically lease their cabs for a fixed fee over a specified period under which they are free to work as much or as little as they want. While they are responsible for fuel and some maintenance costs, drivers keep the full fare income for themselves. The fares themselves are set by the NYC Taxi and Limousine Commission (TLC). In their study, Camerer et al. collect three separate data sets based on NYC cab driver daily trip sheets; sequential lists of each fare as well as start and end time as reported by the drivers themselves. The first data set ("TRIP") consists of 70 trip sheets from 13 drivers (after screening for incomplete trip sheets in an original sample of 192 sheets) who rent their cabs for twelve-hour shifts. While "TRIP" is collected directly from a fleet company, the other two sets, "TLC1" and "TLC2", originate from summary daily statistics collected (and previously screened) by the TLC. TLC1 includes 1044 trip sheets of 484 drivers and TLC2 includes 712 trip sheets by the same number of drivers. Drivers in the TLC samples either lease (on a daily, weekly or monthly basis) or own their cars (the latter being prone to rent out their vehicle which can limit their freedom of choice concerning driving hours).

Using the trip-by-trip data of the TRIP set, Camerer et al. (1997) investigate how the hourly rate varies within the day by regressing the median hourly wage (across drivers working during that hour) on the previous hour's median wage. Their analysis suggests that wages within days are strongly and positively serially correlated: they report autocorrelation of 0.493 (se = 0.092), second-order correlation of 0.578, as well as positive and significant third and fourth-order autocorrelation. Thus, they can rule out the potential spurious consistency with the target income theory that would arise from negatively autocorrelated intraday hourly wage, namely if workers quit early on days with high early wages because they expect the wage to fall. In such a scenario, it would be it impossible to distinguish between the two theories: the neoclassical model would predict that workers stop working since the opportunity cost of leisure has fallen, while the target hypothesis predicts that drivers stop working because they reach their target for the day.

Camerer et al. (1997) then regress the logarithm of daily hours on the logarithm of daily wage rate (the ratio of daily income to daily hours). They control for weather condition and integrate a fixed effect variable to account for the possibility of heterogeneity among drivers. Contrary to the neoclassical model's predictions, the authors find significant and substantial negative wage elasticity in two out of their three data sets. While they reject the hypothesis that the elasticity of hours worked with respect to changes in the wage rate is -1, their data reveal consistent and statistically significant wage elasticity estimates of around -0.5.

In light of their results, Camerer et al. (1997) suggest that, rather than intertemporally substituting leisure for labor across multiple days when wages are temporarily high, cab drivers make labor supply decisions "one day at a time" by setting loose daily income targets. Although the authors do not present a formal model, they argue that their data is consistent with a framework in which the drivers derive utility from the difference between the daily income and a target (or reference) level of income. This is the target income hypothesis formulated in prospect theory, a model of risk attitudes developed by Kahneman and Tversky (1979, 1991). According to prospect theory, supported by considerable experimental evidence, individuals diverge from the strict rationality implied by the neoclassical theory and instead use simple heuristics when making decisions under risk. Reference dependence, a central idea in prospect theory, is the idea that people derive utility from gains and losses, measured relative to some reference point, rather than from absolute levels of wealth. Another central element of prospect theory, loss aversion, is the idea that people are much more sensitive to losses, however small, than to gains of equal magnitude. Graphically, loss aversion results in a value function that is steeper in the region of losses than in the region of gains. This implies, for instance, that earning $10 less than one's reference target is considerably more painful than earning $10 more is pleasurable. According to this logic, workers determine (consciously or not) how much they would like to earn on a given day, their reference point, and are inclined to stop working once this level has been reached. In complete contrast to the stipulations of the standard neoclassical model, an actor who experiences a transitory wage per hour increase will thus work fewer hours since the reference point is reached faster.

However, as noted by the authors, explaining negative wage elasticity estimates implies adopting a model where workers make labor supply decisions based solely on a one-day time horizon. Even under a two-day decision-making horizon, estimated elasticities would be positive for a wide range of plausible specifications (Camerer et al. 1997). Under that scenario, drivers would substitute labor between the two days, working long hours on the first day if it has high wage level, or short hours if it turns out to be a low wage day. For Camerer et al. (1997), such short time horizons are consistent with significant research in both psychology and economics suggesting that people simplify decisions by isolating them from the larger context in which they are embedded, a phenomenon commonly referred to as narrow bracketing (Read et al. 1999). Empirical evidence supporting different aspects of prospect theory has also been found in various settings of decision-making, such as stock markets (Benartzi and Thaler 1995), consumer purchasing behavior (Hardie et al. 1993), consumer choice (Samuelson and Zeckhauser 1988) and insurance markets (Cicchetti and Dubin 1994).

The Camerer et al. (1997) study and its unexpected results have since spawned a succession of studies on the subject. Arguably most true to the original NYC study, Chou (2002) carries out an analysis of the labor supply of taxicab drivers in Singapore based on self-collected survey data. In return for a payment of about \$14, participants were asked to provide personal particulars, answer questions on their driving habits, and fill a form with all dates, shifts, starting and ending times, breaks taken, and total fares collected for five consecutive days. As in the NYC paper, the results indicate significant negative relationship between log hours worked and the log wage rate (the ratio of daily income to daily hours): Chou's results include highly significant OLS wage elasticity estimate of -0.40, an IV estimate of -0.56, and a fixed effect estimate of -0.51. However, since it is not disaggregated into hours within a shift, Chou's data does not allow testing for the autocorrelation of daily wages.

A Mathematical Framework

Gaining an intuitive understanding of why negative wage elasticity can be associated with reference dependent labor supply theory can shed some light on the theoretical assessment. We here adopt the simplistic framework suggested by Chou (2000) and set up a one period conventional additive utility function, with the purpose of understanding $\frac{dh}{dw} < 0$.

Say an agent maximizes the function:

$$u(C) + u(L) \tag{1}$$

subject to,
$$C = y = wh \tag{2}$$

where C is total consumption, y is income, L are number of hours of leisure and w is hourly wage. To impose some structure, let $h = 16 - L$ designate hours worked (incorporating the assumption of 8 hours of sleep).

The first-order condition yields:
$$w = \frac{u\prime(L)}{u\prime(y)} \tag{3}$$

which in turn, when differentiated in respect to L, gives:

$$\frac{dw}{dL} = \frac{u\prime\prime(L)}{u\prime(y)} - u'(L)u''(y)\frac{\frac{dy}{dL}}{(u'(y))^2} \tag{4}$$

Using the structure imposed above, $h = 16 - L$, we can rearrange the above expression using $\frac{dy}{dL} = \frac{dw}{dL}(16 - L) - w$, and finally arrive at:

$$\frac{dh}{dw} * \frac{w}{h} = \frac{1 - yK_y}{yK_y + hK_L} \qquad (5)$$

where, $$K_y = -\frac{u''(y)}{u'(y)} \text{ and } K_L = -\frac{u''(L)}{u'(L)} \qquad (6)$$

By imposing the assumption of concavity in respect to $u(L)$ and $u(y)$ this implies that $K_y, K_L > 0$. Hence we arrive at:

$$\frac{dh}{dw} < 0 \quad if \quad K_y > \frac{1}{y} \qquad (7)$$

As previously discussed, a core aspect of reference dependent choice theory is that agents are risk-adverse in the domain of gains and risk-seeking in the domain of losses, thus $u(y)$ would be kinked. This can be incorporated in the model examined above by adding a reference point: $u(y - t)$ where y designates income and t is the target. When the target is reached, no more utility is added by increasing income: $u(0) = 0$. The function exhibits loss-aversion given that a loss of a given value is more distressful than a gain of the same amount is pleasurable. To examine this, let $u(y, t)$ exhibit the traits associated with prospect theory: $u'(y, t)$ at $y < t$ is greater than $u'(y, t)$ at $y > t$.
Returning to the last stage of the utility function, a kink in $u(y)$ implies that K_y is infinite in the area around the reference point, leaving the wage elasticity to -1.

Hence, $\frac{dh}{dw} < 0$.

This is in contrast to the standard neoclassical model which predicts a wage elasticity of 1. Following this framework, it becomes clear why a short time horizon is an essential requisite to explain negative wage elasticity estimates. Let us now assume instead that the agent maximize the expected utility over a period of two time units:

$$E_1[\textstyle\sum_{t=1}^{2} u(C_t) + u(L_t)] = u(C_1) + E\big(u(C_2)\big) + u(L_1) + E\big(u(L_2)\big) \qquad (8)$$

subject to, $$C_1 + E(C_2) = w_1 h_1 + E(w_2) h_2 \qquad (9)$$

Assuming a discount rate of zero, in addition to utility functions form, first order condition implies:

$$\frac{u'(L_1)}{E(u'(L_2))} = \frac{E(w_2)}{w_1} \qquad (10)$$

The expression reveals that if the expected wage tomorrow is higher than the current wage, agents will substitute supplied labor by taking the afternoon off and instead work more tomorrow. Agents will make decisions based on expected relative wage and consequently the underlying logic behind the drivers stopping behavior cannot be distinguished since one cannot know if they stop because they've reached their target or because they are anticipating higher wages the following day.

Criticism and Econometric Issues

Considering the fame and influence achieved by the Camerer et al. (1997) study, it is important to note that the paper has also been received with skepticism in some quarters. Several alternative explanations for the Camerer et al. (1997) and Chou (2002) results, besides a flawed standard model of labor supply, have been suggested. Some, including Farber (2005) and Oettinger (1999), call the validity of these studies' elasticity estimates into question by arguing that there are a number of econometric issues that have the potential to bias the estimated wage elasticity downward. The negative estimates may then, instead of reflecting true behavior, be the unfortunate consequence of flawed estimation methodology.

First of all, such a bias can be generated if the hourly wage suffers from endogeneity (Stafford 2013), either if it is being affected by the labor supply or if an omitted variable affects both wage and labor supply. Additionally, without access to complete panel data, it is impossible to control for self-selection into participation. The implication of this is that if daily wage fluctuations affect both hours worked and the participation probability in the same direction, it would induce negative correlation between the wage and the error term in the hours equation and, ultimately, create further negative bias to the wage elasticity estimate (Stafford 2013). Finally, by using the observed wage as opposed to a proxy, both Camerer et al. (1997) and Chou (2002) run the risk of introducing measurement error that in turn could induce a negative bias on the elasticity estimates (Stafford 2013; Farber 2005). There is still an ongoing argument whether these issues are significant enough to create negative and significant wage elasticity estimates if the true elasticity is in fact positive.

Farber (2005) identifies yet another point of concern. Based on their finding of substantial positive autocorrelations in the hourly wage within a given day, Camerer et al. (1997, p.408) argue that taxi drivers' wages are "relatively constant within a day". However, Farber (2005) does not find such significant within-day autocorrelations in his data set. Instead, according to his analysis, fares opportunities vary dramatically and unpredictably over the course of the day. In a neoclassical model, if earnings are uncorrelated within a day, the question whether earnings are unexpectedly high or low early in the day becomes irrelevant since the income effect is negligible and expected earnings later in the day become unpredictable (Crawford and Meng 2011). In that case, drivers cannot form rational expectations of the opportunity cost of leisure; hence the intra-day substitution effect will be reduced, resulting in a decrease in the estimated wage elasticity. This would have wide-ranging implications for the validity of Camerer et al.'s (1997) support for the target income hypothesis.

Further Empirical and Theoretical Developments

Although Camerer et al.'s (1997) analysis has inspired a number of empirical studies on the labor supply decisions of workers with flexible hours, the literature has not yet converged on the extent to which the evidence supports reference dependence (Crawford and Meng 2011). Notably, a number of studies investigating labor supply responses in other settings in which workers are free to set their schedule have found evidence for positive wage elasticity estimates.

In a field study examining the daily participation decisions of stadium baseball match vendors, Oettinger (1999) finds evidence for substantial positive intertemporal labor supply elasticity. In a randomized field experiment, Fehr and Goette (2007) examine the effect of a fully anticipated temporary wage increase on the participation rate of bicycle messengers who can freely choose their hours and effort of work. The experimental wage raise led to an increase in labor supply in the form of a greater number of days worked during the month. However, that effect was partially offset by a decrease in labor supply on any particular day. The experiment thus reveals a large positive elasticity of overall labor supply and an even larger elasticity of hours, implying a negative elasticity of effort per hour. The authors argue that the explanation for this effect lies in that messengers are loss-averse relative to a fixed daily target, combined with the lower likelihood of failing to reach the daily benchmark when the wage per hour is high. For Farber (2005), this does not however indicate an inconsistency with the standard neoclassical model: the study identifies large positive intertemporal elasticity of labor supply, and the reduction in daily hours may be due to the possibility that working more but shorter days is the most efficient way for the bike messengers to supply more labor during the high-paying period.

Stafford (2013), in an effort to address a number of methodological issues with the original NYC taxicab study, uses a remarkably extensive panel data set (almost 1,000 individuals over circa 300 days) over daily labor supply decisions of Florida lobster fishermen. Her analysis suggests a small but statistically significant wage elasticity estimate (0.07) and a larger and significant wage elasticity of participation (ranging from 1.29 to 1.42) – results consistent with the neoclassical model. Notably, Stafford's (2013) study shows that methods that do not control for wage endogeneity, measurement error, and participation decisions can generate spurious negative estimates of labor supply elasticity.

Farber (2005) in a study of both his own NYC taxi cab data (593 trip sheets for 22 drivers that lease their cab on a weekly basis over a time period of 13 months) and Camerer et al.'s (1997) data, finds that daily income effects are small and that the decision to stop work at a particular point on a given day is primarily related to cumulative daily hours at that point. Thus, his results are consistent with the conventional neoclassical intertemporal labor supply model. Farber (2005) attributes these sharply contrasting results not to differences in data, but to differences in empirical methods as well as in the conception and measurement of the daily wage rate. Returning again to NYC's taxicab drivers, Farber (2008) reassess the importance of the reference point. He argues that while there may be a reference level of income on a given day such that there is a discrete increase in the probability of stopping when that income level is reached, each particular driver's target varies substantially from day to day and many stop before it is even reached. He continues to argue that cumulative hours might be important, but concludes that the reference point, while not as prominent as Camerer et al. (1997) first suggested, remains an important factor in labor supply models.

A key difficulty in providing further evidence for the target income hypothesis in the labor supply behavior is that the factors involved in determining a driver's reference income point remain unclear. As Barberis (2013) notes, Khaneman and Tversky offered little guidance on this particular aspect of how people think about gains and losses. Köszegi and Rabin (2006) make an attempt to clarify this and propose that targets are based on the driver's expectations. According to their model, workers derive utility from absolute levels of income and hours worked, but also derive prospect theory utility from the difference between daily income and expected income, as well as from the difference between daily number of hours worked and expected number of hours worked.

Crawford and Meng (2011) build on this refined notion of reference-dependent preferences to propose a more advanced model of cab drivers' labor supply in an attempt to reconcile Camerer et al.'s (1997) negative wage elasticity of hours and Farber's (2005, 2008) finding that stopping probabilities primarily and significantly relate to hours. As suggested by Köszegi and Rabin (2006), their framework includes proxy targets for both hours and income as determined by rational expectations using the driver's history of income earned and hours worked on each day of the week. Their analysis of the Farber (2005) data indicates that drivers appear to stop when they reach the second of the two targets: if earnings early in the shift are lower than expected, the income target becomes the determining factor, and if the early earnings are higher than expected, it is instead the hours target that becomes decisive.

To conclude, the extensive research on workers with flexible hours, spawned from the pioneering Camerer et al. (1997) study on NYC taxicab drivers, has typically built on the same basic premises and methods as well as similarly limited and self-reported data (such as Chou 2002; Farber 2005, 2008; Crawford and Meng 2011; Oettinger 1999). Although the debate has gained sophistication and much effort has been put in modeling taxicab drivers' stopping behavior, the basic question of whether the labor supply decisions of taxicab drivers are consistent with a one-day target income hypothesis remains unclear. In order to develop the debate surrounding labor supply decisions, we believe it is necessary to strengthen its foundation by revisiting the evidence from the original study with new comprehensive and precise data. This is precisely what we seek to achieve in the following study.

3. Method

Based on a deductive approach, two competing theories with contrasting predictions are tested against each other with new empirical data. With a heavy focus on providing a transparent and clear analysis, and adopting methods used by previous studies, we make an effort to promote the intersubjectivity and comparability of our findings. Additional alternative procedures are used to provide novel and more detailed understanding of the data. With a critical discussion on the implications of the particular factors of the Stockholm setting, we hope to frame our results in a larger context in order to contribute to a better understanding of labor supply decisions on a general basis. In this section we describe the methods used in this study, before turning to the collected data in section 4.

In order to address the research question, we will estimate the labor supply curves inherent in our data. For this purpose, we must examine the relationship between hours worked and hourly wage, and in addition include control variables, control for measurement errors, and examine the autocorrelation in hourly wage between and within days. Furthermore, we will also look at the participation decision and consider the alternative target of cumulative hours worked.

The Hours of Work Equation

We begin by estimating the following Frisch (or λ-constant) hours of work equation proposed by MaCurdy (1981). The model, adopted by Camerer et al. (1997) and formalized by Farber (2005), reads as follows:

$$\ln H_{it} = \eta * \ln W_{it} + X_{it}\beta + \epsilon_{it} \tag{11}$$

H_{it} is defined as hours worked during a specific day, W_{it} is the hourly wage earned during that day calculated by multiplying the taxicab driver i's daily income at day t by 0.37 and then dividing it with hours worked during that day.[1] X_{it} represents other factors and ϵ_{it} represents the error term.

A regression analysis will be carried out using the data set containing summaries of each driver's daily fares. Acknowledging that demand could have an impact on the hours supplied by taxicab drivers during a given workday, we include different proxy variables to control for demand, most linked to weather conditions (as done by Camerer et al. 1997; Chou 2002; Farber 2005). When it is raining or cold, the logic goes, people are more likely to take a taxi, and thus the driver's hourly wage increases with this increase in demand. From the supply side, later studies have shown that adverse weather conditions have a positive impact on the number of hours one chooses to work (Connolly 2008), yet Camerer et al. (1997) also argue that driving a taxicab is incrementally harder in the rain.

One weakness of the equation (11) model is that it is possible that there are unobservable factors that affect the hours supplied during a given day. However, due to the panel character of the Stockholm data, it is possible to include a fixed effects (FE) component that allows for driver-specific heterogeneity, resolving the problem with omitted variables that differ between individuals but are constant over time. Adding this element (α_i) to the model accounts for time-constant factors such as personal wealth, motivation, and personality that remain relatively unchanged under the examined time period. The FE estimated model can thus be formulated as:

$$\ln H_{it} = \eta * \ln W_{it} + X_{it}\beta + \alpha_i + \epsilon_{it} \tag{12}$$

The main assumption when applying FE is strict exogeneity between the explanatory variables, as well as a time-constant unobserved individual effect. The strict exogeneity assumption (neglecting logs) in the purposed model can be presented as follows;

$$E(H_{it}|x_{i1}, \ldots, x_{iT}, a_i) = E(H_{it}|x_{it}, a_i) = x_{it}\beta + a_i \tag{13}$$

If equation (13) holds, $\{x_{it}: t = 1,2, \ldots, T\}$ are strictly exogenous conditional on the unobserved effect a_i. Note that this assumption is more likely to hold than the assumption on strict exogeneity without conditioning on the unobserved effect.[2] Indeed, the method is commonly used when looking at individuals' discrete choices, as well as the previous research on taxicab drivers' labor supply decisions (Camerer et al. 1997; Chou 2002). Examining equation (12) using observed income to derive hourly wage and hours worked (conditioned on hourly wage) is likely to be jointly determined. If some factors are unobserved, and not constant over time, they will remain as a part of ϵ_{it}. Since both have an effect on wage and hours supplied, this will cause endogeneity and bias the coefficient estimate of hourly wage.

---

[1] 0.37 represents the percentage of each fare that Stockholm drivers get to keep as reimbursement.

[2] The strict exogenous assumption restricts how the expected value of H_{it} can depend on the explanatory variables in other periods. Without conditioning on the unobserved effect the assumption

[2] The strict exogenous assumption restricts how the expected value of H_{it} can depend on the explanatory variables in other periods. Without conditioning on the unobserved effect the assumption is: $E(H_{it}|x_{i1}, \ldots, x_{iT}) = E(H_{it}|x_{it}) = x_{it}\beta$. This assumption will fail whenever the conditional assumption holds and the expected value of a_i depends on $(x_{i1}, \ldots, x_{iT})$. For a more complete argumentation, see Wooldridge (2002, p.287).

An Instrumental Variable Approach

Measurement error is a pressing concern in practically all studies on labor supply. However the observations in our data do not, as in previous studies, originate from self-reported trip sheets, rather they are obtained from an automated log system. This dramatically reduces the likelihood of errors in reporting that could otherwise lead to biased estimates. However we cannot completely rule out that our data may suffer from recording errors. For example, if the hours were overstated, estimated hourly wage would be lowered just as understated hours would artificially raise the estimated wage.

According to Camerer at al. (1997), measurement error can be solved if an instrument for wage is available, on the condition that the instrument is uncorrelated with the measurement error in hours. The authors suggest, following Maddala (1992), that other drivers' reported wage from the same day (the 25th, 50th, and 75th percentiles) could be used as an instrument for individual wage. The instrument is appropriate if it is uncorrelated with the measurement error of a specific driver. However, Farber (2005) argues that this instrument fails to remove the bias if the calendar date both affects wages and correlates with the labor supply at a given wage. But Farber (2005) does not offer a new instrument, therefore we will use the instrument introduced by Camerer et al. (1997) and Chou (2002). Additionally, measurement error may also arise in income, for example due to the omission of tips. As will be discussed further, the relative lack of regular tipping in Stockholm provides us with an opportunity to minimize this particular effect in our analysis.

Testing for Autocorrelation of Wage

To carry out the proposed study, Camerer et al. (1997) introduced two key assumptions that must hold in order to be able to distinguish between the two models of labor supply: no correlation of wage between days, and positive autocorrelation on hourly wage within days. Both are fairly easy to test.
To assess if wage is correlated across days, we regress the mean or median wage on day t on the mean or median wage on day $t-1$. If we find autocorrelation across days, we can also empirically test the equation (10) by looking at two specific weekdays across our sample, estimating whether hours worked are higher on high-wage days, and if drivers work for shorter hours the day before. This would, in a simplistic way, test if drivers actually choose to work shorter hours if they expect higher wages tomorrow.

Trip-by-trip data is essential to test the second assumption, since our primary data set only contains summary statistics: a second data set was collected for this purpose. To test for autocorrelation within days, the days in our data set were broken into hours, and the median hourly wage for all drivers during that day was calculated. We then regress median hourly wage of drivers working during that hour, on the previous hour's median wage. The fundamental characteristic of the life-cycle theory is that when wage is transitorily high, the opportunity cost of leisure rises. But if there is negative autocorrelation between hours, the summary statistic would not capture the fact that the opportunity cost of leisure has decreased, making it impossible to determine which theory explains labor supply. In the opposite case, if there is positive or zero autocorrelation between hours, the summary statistics will accurately predict the opportunity cost of leisure and this will in turn give ground for a clear test of the theories.

The Extensive Margin

Furthermore, we need to account for the fact that the drivers in our sample have the choice whether to work or to take the day off on any given day. As Heckman (1979) argues, neglecting this would lead to a selection problem. To only account for the intensive margin (daily hours supplied) and disregard the extensive margin (daily participation) has been found to bias the wage elasticity estimate at macro (Coleman 1989; Mankiw et al. 1985; Alogoskoufis and Manning 1988) and micro levels (Heckman 1974; Oettinger 1999). The suggested solution to this problem (a topic to which we will return later) is to implement a Heckman model, also known as the Type II Tobit. While this model is straightforward, getting accurate predictions with it is difficult since it requires an instrument variable that has an effect on the participation decision, $(H > 0)\ or\ (H = 0)$ on a given day, but that has no effect on income. Previous studies on taxicab drivers (Camerer et al. 1997; Farber 2005) have not been able to analyze the effect of the extensive margin. While the data of this study allows for such an analysis, finding an instrument was not possible.[3] However it is reasonable to assume that drivers indeed make an active participation decision both on the extensive and intensive margin. Since we cannot use the Heckman model, we suggest an alternative approach.

We regress the active drivers on any given day on two dummy variables: whether the aggregated wage from the same day and the aggregated wage from the previous day are above the median or not. By including a control variable, we can get some estimate of participation increases if wages are high, resulting in a positive dummy variable estimate. If this is true, neglecting the extensive margin will cause our predictions to be biased in the original models (equation (11) and (12)). The bias would be the opposite sign of the correlation between the error terms in hours and participation decision. Thus, if the unobserved shocks to participation and hours are positively correlated, elasticity would be downward-biased.

The proposed approach will not correct for any possible bias but allows us to predict in which direction the bias would appear and to consider the magnitude of the potential selection problem. We are not arguing that the proposed model would serve as empirical evidence of a casual relationship between the number of drivers that participate on a given day and median wage, even if we will include control variables.

To do so would necessitate making the very liberal assumptions either that the proposed model is fully correctly specified, or if there is an omitted variable, that it does not have any effect on participation decision or is uncorrelated with all the independent variables. This is not likely to be the case. Instead, our objective is simply to get a better understanding of the extensive margin.

Heteroscedasticity

Based on preliminary analysis, it seemed clear that our data could suffer from heteroscedasticity. This has not been addressed in the previous research (Camerer et al. 1997; Chou 2002). To be conservative in our predictions and stay true to these studies, we will report results based on homoscedasticity and accompanied by robust

[3] We have attempted using several different alternative instruments, including whether it was raining on the previous day, bank holidays, Christian holidays, as well as days when schools were closed during workweeks. None had an effect on participation without affecting income, hence the Heckman model could not be used. Notably, this is regarded as the main drawback of the Heckman model, and has caused econometricians to search for other methods that could correct the selection problem while being more accessible to real-world conditions (Puhani 2000).

and cluster standard errors at the individual level and p-value in footnotes.[4]

Proxy Targets

In addition, progress in the field has led researchers to question the applicability of using income (in terms of hourly wage) as a target (Farber 2005, 2008; Crawford and Meng 2011). In order to address this question, we adopt a stopping probability model that attempts to reconcile target income hypotheses with an alternative target: cumulative hours worked. In doing so, Crawford and Meng (2011) proxy drivers' targets, both in terms of hours worked and income, by looking at specific driver day-of-the-week targets. Simply put, the proxy for the second week is the week before, while the proxy for the third week it is the average over the relevant days in the two previous weeks. They do not include the actual day when calculating the target and ignore sampling variation for simplicity. By including both targets it is possible to see which has had an impact on the probability of stopping. Due to data and time limitations, we cannot compute the exact same model, but nearly.[5] We first calculate the empirical hazard of stopping and use this as dependent variable, in respect to hours and income. As independent variables we use cumulative hours from that day, all income earned during that day, and whether the driver reaches the proxy income target or the hourly target. The proposed model is then:

$$P(quitting) = \beta_0 + \beta_1 CH_{it} + \beta_2 CI_{it} + \delta RHT_{it} + \delta RIT_{it} + \delta RBothT_{it} + \epsilon_{it} \quad (14)$$

where CH_{it} is cumulative hours, CI_{it} is cumulative income, RHT_{it} and RIT_{it} are dummy variables that take the value of 1 if the hourly target is reached and income target are reached (respectively), and $RBothT_{it} = 1$ if both targets are reached. The last term $RBothT_{it}$ differs from the original Crawford model due to the fact that we can only observe whether or not the targets are reached, as opposed to which target is reached first. Another consequence of this is that we use a simple OLS regression. The model circumvents many of the obstacles encountered in the equations (11) and (12): wage is not calculated by dividing by hours, thus wage and hours are not jointly determined. In addition, we try to model a target, even if based on a noisy proxy, that shouldn't be systematically biased and, since it is predetermined, should not cause endogeneity. In summary, this is a combination of Farber's (2005) and Crawford and Meng's (2011) models, with some changes to fit the nature of our data.

If any of the dummy variables are significantly positive, this will indicate consistency with the reference dependence model, since the neoclassical model would predict no "jumps" in the probability of stopping their shift when a proxy target is reached.

---

[4] Although our data set is not far from satisfying the general rule-of-thumb requirement of $panels \approx 50$ for using clustered standard errors, results should still be taken with a grain of salt. Due to computation obstacles, we do not adopt the possibility to cluster both in respect to individuals and time, which could have an impact on our results.

[5] The model is based on that at any point τ during a shift, a driver with rational expectations can predict an optimal stopping point τ^*. A driver will stop at τ if $\tau^* \leq \tau$. In our data we only observe τ^*, therefore we adopt a reduced form (Farber 2005).

4. Data

In this section, the origin and nature of the data sets collected for this study are discussed. We begin by describing the data sets, and the Stockholm market from which they originate (in part based on informal interviews with representatives of taxi franchise companies, the transportation workers union *Svenska Transportarbetareförbundet*, the Swedish Transport Administration *Trafikverket*, as well as several taxicab drivers). We then describe the process of screening this data for inconsistency before providing some summary statistics of the data sets.

Stockholm's Taxi Market

As previously described, much of the progress in research on labor supply decisions has suffered from the fact that extensive and reliable data rarely is available to researchers. Previous studies on taxicab drivers have been based on trip sheets that taxi drivers are supposed to fill in during their shifts. The reliability of such data depends heavily on the sincerity and diligence of the participants. Taxicab drivers are arguably less prone to fill out such sheets correctly during busy days (especially in a profession when hands stay on the wheel and eyes should be kept on the road), and it is fairly simple and risk-free for them to simulate the trip sheets in the end of the shift. Additionally, it is very difficult for researchers to post hoc identify when such dishonesty or otherwise inaccurate reporting takes place. Since these early studies, important technological advances have been made, and modern systems now allow doing away with the hassle of self-reported trip sheets.

In Stockholm, where the market is dominated by a number of large-scale taxi franchise companies that rely on central call centers, there is a strong incentive for developing advanced log systems. We set out to take advantage of the existence of such log systems to collect much more reliable and comprehensive panel data that does not require effort or any participation decision from the drivers. There are two main obstacles to collecting such real world data: gaining access to the data, and turning the data into useful data sets. After contacting many taxi franchise companies over a period of several months and many meetings, one company finally agreed to grant us access to their log system. The software automatically computed summaries of the days of all the drivers, but made it difficult to examine individual drivers' trip-by-trip driving behavior. With help from the taxi franchise company, data was eventually successfully extracted in the form of individual image files. Converting these to a usable data set took considerable effort and patience.

The company in question has an advanced log system that requires drivers to identify themselves with a personal identification card during the shift. The drivers start and end their shift by inserting and extracting their card from the taximeter. There is a clear incentive to accurately carry out this procedure since it is the only way to get connection to the booking central and receive orders from the taxicab franchise. Indeed, there is in Stockholm a lesser reliance on picking up passengers (so called "cruising") than in for instance NYC. For the larger firms, about 60% of fares go through the central call center.

The Swedish taxi market has been largely unregulated since 1990: fares are not set by the government but by individual taxi firms themselves (Forssén 2008). The standard employment arrangement in our sample is one in which drivers are lent a vehicle from a fleet company for a given period of time (days, weeks or months) in exchange for which they are expected to pay a part of the earned income to the fleet company. Although the car could technically be taken back if the driver does not reach a minimum target, drivers are in high demand in Stockholm. The minimum amount is therefore often relatively low and drivers only very rarely get "fired" by the fleet company. Interestingly, drivers generally have very little loyalty to their fleet or taxi franchise company, and often change between firms. As one of the interviewees put it: "It's a driver's market." For these reasons, it does not seem that this minimum level of income imposed by the lease company would be a significant factor in drivers' labor supply decisions or their response to transitory wage fluctuations.

Although the market is described as chaotic in terms of employment conditions (this was a reoccurring complaint in our interviews), the typical agreement is that the fleet company takes a certain cut of each fare, in exchange for which it provides the car and pays fuel and maintenance costs. Drivers earn commission on the fare income, typically about 37%.[6] Two key factors are central to our analysis: drivers in our samples earn only commission and are free to choose their own work hours.

The fleet companies are generally tied to a specific taxi franchise company that trains the drivers, set the fares, and operates the booking centers. The taxi franchise company who supplied us with the data is one of the largest firms on the Stockholm market with approximately 1,200 vehicles generating about 10,000 cab trips a day.

Data Sets

This paper is based on two different sets of data. The first main data set includes trip sheets of 48 cab drivers over a period of 3 months, from April the 1st to the 30th of June 2012. This particular time period has the advantage of offering the smooth but still variable weather of the brisk Stockholm summer while avoiding disruptions such as a large drop in taxi demand associated with the vacation hiatus of July.

The drivers in the sample do not share their taxi vehicle with any other driver, which makes them completely free to set their own work schedule since they do not need to take any another driver's preferences into account. The data set includes daily income, hours worked, and occasionally breaks. Income is calculated by adding together all trips during one shift, not by calendar date, which allows for looking at actual shift-income and not disrupting the analysis by relying on the artificial boundary imposed by calendar dates which very often does not correspond to the actual labor supply behavior of taxi cab drivers. Indeed, night shifts are very popular. The number of hours worked is calculated as the difference between the time of insertion and time of extraction of the driver's identification card. Breaks are only observed when drivers remove their identification card.

In addition, we collected a data set that allows us to observe driving behavior within days, namely daily trip-by-trip data. This data set contains 22 of the drivers included in the previously described data set, and covers 10 days from the 1st to the 10th of

[6] The national transportation worker's union *Transportarbetareförbundet* informed us that drivers, according to collective agreements, should be paid a minimum wage in case they do not reach a minimum of an average hourly wage of 106 SEK. However, they told us this is very rarely the case. Indeed very few of the fleet companies use this collective agreement and less than 10% of taxi drivers are members of the union. All taxi drivers we have spoken to confirmed this: none of them earned anything but a commission on fares. The taxi franchise company who provided us with data told us that only a tiny fraction of their drivers were part of such a collective agreement, but nonetheless honored our request to restrict the samples to drivers who solely earn commission. We have no reason to believe that the resulting potential selection bias should be anything but negligible.

April 2012. It includes specific fare income, where the trip started, at what time the trip started, and how the consumer paid (card or cash), and, for a majority of observations, the type of passenger (government-sponsored, corporate clients, pick-ups, or booked through the call center).

All observations were attributed corresponding daily weather conditions, (rainfall and temperature) as well as other variables. The weather data was collected from Sweden's Meteorological and Hydrological Institute (SMHI). Although the data sets include the taxi license number, this number does not indicate when the licenses were issued and the taxi franchise company does not keep records of when the drivers started working in the franchise, thus we have no indicator of the experience of the drivers.

Screening Process

Below follows an outline of the logic and rules adopted for the handling and screening of the two main data sets.

The first data set (DAY#) contains daily summary statistics of 48 drivers over a period of three months; a total of 4,215 observations. Regularity checks showed that one driver in the sample drove excessively long hours, even going so far as working back-to-back 30-hour shifts. We therefore removed this individual from the data set and were left with 47 drivers. The minimum number of active drivers on any of the 91 days included in the sample is 14, and the maximum is 47. Overall participation for the drivers ranges from 27 to 86 days, with a central tendency of approximately 68.
The screening procedure suggested by Farber (2005) to ensure that the sheets are internally consistent (including making sure that shifts start no earlier than the end of a previous shift, and start before they end) was adopted. As expected, since the data originates directly from the computer system of the taxi franchise company, no observation displayed such internal inconsistency. However, 7 observations out of nearly 4,215 were excluded due to suspiciously long (5 observations), or short and lucrative (2 observations) shifts. Excluding these observations has no effect on the central tendencies of our results, but increases the variance, in turn resulting in lower significance levels. According to our contact at the taxi franchise company, these outlying observations can be a result of either software malfunction, drivers keeping the taximeter on when using the car off-duty in an effort to evade taxation of fringe benefits[7], or drivers picking up single fares outside of their "working hours" (resulting in very short, and unusually lucrative "shifts").
Drivers, having complete discretion in setting their working hours, sometimes work several separate shifts within a same day. They also typically take short breaks during shifts, and these intra-shift breaks are recorded in the data set whenever drivers also physically remove their card from the taximeter. Both of these phenomena contribute to the existence of duplicate observations per day and driver (amounting to 678 observations). Unfortunately, only one observation per day and driver can be accommodated within the panel data analysis. To address this issue, as well as to attempt to differentiate between shorter intra-shift breaks and more significant inter-shift pauses, we have chosen to create three data sets based on different assumptions about the gaps between such duplicate shifts.

[7] Since Swedish tax authorities consider the unreported use of taxicab vehicle for personal purposes as tax evasion, drivers might keep their taximeter on when they are not actually working.

- In DAY1, the time attributed to all breaks and gaps between recorded shifts is removed, no matter how long or short. 94 observations were altered, leaving 3530 observations. In this manner, the average hourly wage is as close to reflecting actual hours worked as our data allows for. Therefore, DAY1 is used for the main analysis in section 5.

- In DAY2, breaks are only removed when longer than 150 minutes. If shorter, they are accounted for as part of the shift (time worked). This results in treating such shorter breaks in a similar manner as the unrecorded intra-day breaks taken without the removal of the driver's identification card. Again, 94 observations were altered, leaving 3530 observations. DAY2 will be used to test the implications of including breaks as working hours in section 6.

- In DAY3, hours supplied by each driver on a given day are computed by neglecting the breaks and including the complete time between the first insertion and the last extraction of the identification card for any given driver and day. Again, 94 observations were altered, leaving 3530 observations. This results in a data set that resembles the summary statistics of the TLC samples used by Camerer et al. (1997) and will be examined in section 6. Two additional data sets, SUB1 and SUB2, represent randomly chosen smaller subsamples (18 drivers over 12 days and 15 drivers over 13 days respectively) of DAY3. SUB1 contains 173 observations from 1st to 12th of April and SUB2 contains 178 observations from 18th to 30th of June. These smaller data sets will be examined in section 6. It is important to note that these were the first and only subsamples extracted.

The second main data set, TRIP1, contains trip-by-trip data of 22 drivers over 10 consecutive days beginning on the 1st of April 2012. Since no internal inconsistency in the data was identified, the data set was not altered in any way. TRIP1 contains 1716 individual trips.

All in all, the data screening process necessitated very little screening of observations compared to previous studies (Camerer et al. 1997; Chou 2002; Farber 2005, 2008).

Sample Characteristics

TABLE 1: SAMPLE CHARACTERISTICS

	Median	Mean	Std. Dev.
DAY1 (n=3530)			
Hours worked	11.52	11.60	3.14
Average wage (SEK)	93.67	94.89	28.78
Total revenue (SEK)	2979.00	2961.50	1101.53
Rain (mm)	0	2.65	5.99
DAY2 (n=3530)			
Hours worked	11.83	11.65	3.14
Average wage (SEK)	93.37	94.51	28.71
Total revenue (SEK)	2979.00	2961.50	1101.53
Rain (mm)	0	2.65	5.99
DAY3 (n=3530)			
Hours worked	11.95	11.82	3.47
Average wage (SEK)	92.61	93.68	28.61
Total revenue (SEK)	2979.00	2961.50	1101.53
Rain (mm)	0	2.65	5.99
SUB1 (n=173)			
Hours worked	11.70	11.53	3.53
Average wage (SEK)	74.86	76.84	22.91
Total revenue (SEK)	2380	2347.97	866.058
Rain (mm)	0	1.18	1.51
SUB2 (n=178)			
Hours worked	12.33	12.41	3.77
Average wage (SEK)	85.55	86.63	28.06
Total revenue (SEK)	2758.00	2824.62	967.56
Rain (mm)	0.8	3.57	5.24

Table 1: Sample Characteristics – The table contains the key variables from the five different data sets.

Table 1 exhibits the mean, median, and standard deviation of the key variables in the five data sets DAY1, DAY2, DAY3, SUB1 and SUB2. In DAY1 and DAY2, a taxi driver works on average about 11.5 hours per day, and collects approximately 95 SEK per hour. The hourly wage is marginally lower in DAY2, due to hours being slightly longer since some breaks are included. DAY3 has the longest reported hours, consequently driving down the hourly wage. As expected the total revenue is the same as in DAY1 and DAY2. The subsamples SUB1 and SUB2 have longer reported hours, resulting in lower wage. SUB1 exhibits substantially lower earned income, which seems attributable to the specific time period (beginning of April) in which all drivers have lower reported income. TRIP1 contains 1716 trips with an average fare of 268 SEK and a median of 214 SEK.

The reported average daily temperature during the examined time period ranges from 0.2 to 18.6 degrees Celsius with an average of 10.29, and the amount of daily rainfall ranges between 0 and 33.5 millimeters with an average of 2.65. Distributional graphs and diagrams can be found in Appendix 1.

5. Results

In this section, we will first examine the two assumptions regarding autocorrelation within and between days, before turning to our main focus, namely wage elasticity estimates using both OLS and IV approaches. Lastly, we look into the participation decision at the extensive margin and implement a simple model incorporating both hours and income targets.

Correlation of Wage Within and Between Days

To examine autocorrelation within days, we regress the median hourly wage on the wage in the previous hour, and estimate a positive autocorrelation of 0.162 (se = 0.0641) with a p-value<0.05. The second-order autocorrelation and third-order autocorrelation both go in the same direction. Weighting the median by the number of drivers also gives the same result. This implies that if a day turns out to have early high wages, it is likely to continue to have high wages later in the day, and is similar to what Camerer et al. (1997) find. Appendix 2 contains tables and figures.

Turning to autocorrelation between days, we run the regression of the median or mean wage of day *t* on the median or mean wage of day *t*-1. We find substantial correlation between days when computing the first regression, with an estimated autocorrelation of 0.432 (se = 0.0972) and a p-value<0.01 (Table 3). This means that if wages were high yesterday, drivers can also expect wages to be high today. Since this is an unanticipated result, a closer look at the evidence is justified. Even when checking for second and third-order autocorrelation, the result remains positive and statistically significant with p-values<0.01 (see model 2 and 3 in Table 3). Controlling for weekdays still results in a positive and significant autocorrelation (model 4, Table 3).[8] Judging from our data, it is clear that some days are more likely to have high wages than others. Unsurprisingly, Fridays and Saturdays are the most likely candidates to be high wage days, with statistically significant (p-value<0.01) estimated increases in earnings of 18.43 SEK (se = 3.793) and 19.81 SEK (se = 3.793) respectively (as compared to Mondays). In model 5 from Table 3, we include median wage from the two and three previous days, second-order and third-order autocorrelation, leading the estimate to decrease slightly, although it still remains largely positive and statistically significant.

[8] Since including all the days of the week would lead to perfect multicollinearity, Monday is excluded but still represented in the constant.

TABLE 3: CORRELATION BETWEEN MEDIAN HOURLY WAGE ACROSS DAYS – DAY1

	1	2	3	4	5
Median wage *t*-1	0.432***	0.509***	0.491***	0.422***	0.359***
	(0.0972)	(0.106)	(0.110)	(0.100)	(0.114)
Median wage *t*-2		-0.122	-0.0752		0.153
		(0.106)	(0.121)		(0.119)
Median wage *t*-3			-0.0790		0.0240
			(0.108)		(0.112)
Rain				-0.492	-0.673
				(2.149)	(2.159)
Tuesday				7.493*	7.652*
				(3.903)	(4.223)
Wednesday				6.067	7.746*
				(3.837)	(4.359)
Thursday				7.171*	8.737**
				(3.876)	(4.277)
Friday				18.43***	20.10***
				(3.996)	(4.362)
Saturday				19.81***	22.03***
				(3.793)	(4.451)
Sunday				7.132*	7.929*
				(3.903)	(4.311)
Observations	90	89	88	90	88
Adjusted R-squared	0.183	0.217	0.213	0.465	0.478
Constant	53.72***	58.27***	63.04***	45.50***	33.77**
	(9.267)	(10.84)	(12.69)	(10.25)	(13.68)
Number of drivers	47	47	47	47	47

Table 3: Correlation between median hourly wage across days – The table contains autocorrelation between days by regressing the median wage at day t on the median wage at day *t*-1. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1

Consequently, our analysis suggests that wages are correlated across days, and fairly stable within days. These results differ from those of the Camerer et al. (1997) study, in which wages were found to be uncorrelated across days and correlated within days. This will be discussed further in section 6.

Wage Elasticity

For a graphical representation of the data, scatterplots of (log) hours – (log) wage relationship can be found in Appendix 3, while regression estimates can be found in Table 5.

To account for the panel nature of the data, we indicate if the estimates include a driver fixed effect at the bottom of Table 5. We include two dummy variables to control for shifts in labor supply: whether it is a high temperature day (above 10 degrees Celsius) and whether it rains during the day. Including linear rain and temperature variables has a negligible effect the estimated wage elasticity. The resulting estimated wage elasticity of our six models ranges from 0.132 to 0.144 and, with p-values<0.01, are statistically different from zero. Both control variables are insignificant and the fixed effect estimates are virtually the same as in the standard model (comparing estimates on equation 1, 3, and 5 to 2, 4 and 6 in Table 5).[9] We also conduct several sensitivity analyses, as well as procedures that reduce outlier influence (i.e. median regression), suggesting that the results are not sensitive to potential outliers. The results imply non-negative wage elasticity and indicate that there is little support for the reference income target hypothesis in our data. In contrast to the prediction of the income target hypothesis, the drivers work longer hours when the wage is transitorily high. However, due to the observed correlation across days, the implications are not straightforward. This will be further discussed in section 6.

In addition, we compute an empirical test for the proposed model in equation (10) that evaluates whether higher expected wage tomorrow reduces the amount of hours worked today. As discussed, our previous analysis indicates that Fridays are high-wage days. Computing a simple test of hours worked on Thursdays compared to Fridays, supplied work-hours on Fridays is significantly larger than on Thursdays with a p-value<0.01. These results suggest that a driver works longer hours on Fridays while cutting back on hours worked on Thursdays, which gives some empirical grounds from equation (10). This is consistent with the multi-period substitution effect that could result from a framework with, for instance, a two-day decision-making horizon.

TABLE 5: OLS LOG HOURS WORKED EQUATION – DAY1

	1	2	3	4	5	6
Log hourly wage	0.132***	0.144***	0.131***	0.144***	0.130***	0.142***
	(0.0166)	(0.0167)	(0.0166)	(0.0168)	(0.0167)	(0.0168)
Rain			0.00373	0.00344	0.00438	0.00407
			(0.0102)	(0.0102)	(0.0103)	(0.0102)
High temperature					0.0104	0.0100
					(0.0104)	(0.0104)
Fixed effect	No	Yes	No	Yes	No	Yes
Observations	3,530	3,530	3,530	3,530	3,530	3,530
Adjusted R-squared	0.2060	0.2060	0.2066	0.2065	0.2075	0.2074
Constant	1.813***	1.758***	1.814***	1.759***	1.812***	1.758***
	(0.0761)	(0.0755)	(0.0761)	(0.0755)	(0.0761)	(0.0755)
Number of drivers	47	47	47	47	47	47

Table 5: OLS log hours worked equation DAY1 – The table contains regression results estimating hours worked equation including control variables and estimates with fixed effects. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1

Controlling for Measurement Error

As previously discussed, measurement error is a reoccurring problem in labor supply research. We run the same regression as presented in Table 5, but in addition use other drivers' reported wage on the given day (25th, 50th, and 75th percentiles) as an instrument for individual wage. If the measurement error is significant, the results

[9] Robust and cluster standard errors lead to practically the same estimates, with p-values<0.01.

displayed in Table 5 will be biased towards zero.[10] Consequently, when controlling for measurement error, the estimates are expected to be even more positive. The results can be found in Table 6.

Again, to account for the panel nature of the data, we indicate whether the estimates include a driver fixed effect at the bottom of Table 6. As with OLS, we also include two dummies: whether it was a high temperature day (above 10 degrees Celsius) and whether it rained during the day. First stage regression indicates that other drivers reported wage is a strong instrument for individual wage. The joint test of the null hypothesis that all coefficients are zero can be easily rejected.

TABLE 6: IV LOG HOURS WORKED EQUATION – DAY1

	1	2	3	4	5	6
Log hourly wage	0.138***	0.149***	0.137***	0.148***	0.134***	0.142***
	(0.0209)	(0.0211)	(0.0210)	(0.0212)	(0.0210)	(0.0168)
Rain			0.00344	0.00326	0.00416	0.00407
			(0.0103)	(0.0102)	(0.0103)	(0.0102)
High temperature					0.0102	0.0100
					(0.0105)	(0.0104)
Fixed effect	No	Yes	No	Yes	No	Yes
Observations	3,530	3,530	3,530	3,530	3,530	3,530
Constant	1.784***	1.738***	1.787***	1.740***	1.794***	1.758***
	(0.0952)	(0.0953)	(0.0952)	(0.0954)	(0.0951)	(0.0755)
Number of drivers	47	47	47	47	47	47
		FIRST STAGE				
		1	2			
Reported wage		1.319***	1.286***			
		(0.0169)	(0.0169)			
Adjusted R-squared		0.6543	0.6543			
P-value on F-test instrument of wage		0.00	0.00			
Fixed Effects		No	Yes			

Dependent variable is the log of average hourly wage.
Regressions also include weather characteristics as explanatory variables.

Table 6: IV log hours worked equation DAY1 – The table contains IV regression result estimating hours worked equation including control variables and estimates with fixed effects. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1

The resulting estimated wage elasticity of our six models range from 0.138 to 0.148 and, with p-values<0.01, are again statistically different from zero.[11] Both control

[10] The derivation can be presented using a simple regression model $y = \beta_0 + \beta_1 x_1^* + u$, where x_1^* is unobserved. Instead, we observe x_1, implying that $e_1 = x_1 - x_1^*$ is the measurement error given that the actual estimated model is $y = \beta_0 + \beta_1 x_1 + (u - \beta_1 e_1)$.
The classical error-in-variables assumption is $Cov(x_1^*, e_1) = 0$, so e_1 and x_1 must be correlated.
Hence, Cov$(x_1 e_1)$= $E(x_1 e_1)$= $E(x_1^* e_1) + E(e_1^2) = 0 + \sigma_{e_1}^2 = \sigma_{e_1}^2$
This results in $Cov(x_1, u - \beta_1 e_1) = -\beta_1 \sigma_{e_1}^2$. Deriving the bias leads to $(\hat{\beta}_1) = \beta_1 (\frac{\sigma_{x_1^*}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2})$.
The expression shows that OLS regression will suffer from attenuation bias.

variables are insignificant and the fixed effect estimates are virtually the same as in the standard model. The IV estimates are all greater than the estimates obtained with the standard OLS procedure. This is anticipated, and reflects that when measurement error is present in some of the independent variables, it creates a predictable bias towards zero. The results imply even more non-negative wage elasticity than obtained by OLS and reinforces that there is little support in our data for the reference income target hypothesis over a one-day time horizon.

Driver Participation: the Extensive Margin

Earlier results presented in this section imply that wages in our data fluctuate substantially across days, but in contrast to what previous studies have found, seem to fluctuate in a systematic manner. The estimate of autocorrelation across days, and the variability of wage given specific weekdays, indicate that there is a pattern in wage over time. This is noticeable in the graphical representation of wage over time in Figure 6 and 7 in Appendix 2, which clearly exhibits some recurrent high and lows. We can test if drivers acknowledge this pattern by choosing to drive when the expected wage is high and take the day off when expected wage is low, by conducting the suggested method described in section 3. Results are found in Table 7.
Model 1 shows no statistical significant relationship. Model 2 include a dummy variable that indicate if the median wage from the previous day is above the general median across the 91 days. From earlier results we know that wage is correlated with workdays, which could have an effect on the participation rate. We therefore run Model 3 that looks only at the participation decision and includes the days of the week.[12] The estimates are statistically significant but do not vary much across days. Contrary to model 1, model 4, that includes the dummy variable of whether the day's median wage is above the median wage across all days, finds support that drivers indeed are more likely to participate on high wage days. The estimate suggests that, controlling for daily fluctuation in participation rates, 4.477 more drivers choose to work (se = 0.964) on high wage days. The estimate is statistically significant with a p-value<0.01.

[11] Robust and cluster standard errors is not possible with IV estimates.
[12] In order to avoid perfect multicollinearity, we omit Sunday during which the participation rate fluctuate most.

TABLE 7: ACTIVE DRIVERS WHEN MEDIAN WAGE IS HIGHER THAN THE MEDIAN

	1	2	3	4
Higher than median	2.529	3.819**		4.477***
	(1.823)	(1.825)		(0.9989)
Higher than median *t*-1		-2.545		-0.3639
		(1.825)		(0.9989)
Monday			17.15***	17.75***
			(1.630)	(1.526)
Tuesday			17.85***	17.98***
			(1.630)	(1.505)
Wednesday			18.23***	17.65***
			(1.630)	(1.486)
Thursday			18.31***	18.02***
			(1.630)	(1.471)
Friday			20.31***	19.076***
			(1.630)	(1.522)
Saturday			16.38***	15.03***
			(1.630)	(1.476)
Rain				0.7224
				(0.857)
Observations	91	90	91	90
Adjusted R-squared	0.021	0.0529	0.720	0.773
Constant	36.96***	38.17***	23.38***	19.66***
	(1.576)	(1.710)	(1.153)	(1.416)
Number of drivers	47	47	47	47

Table 7: Active drivers when median wage is higher than the median across the sample – The table contains regression result on participation decisions. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1

The results presented in Table 5 implies that drivers are more likely to work on high wage days, which causes the estimates in Table 5 (OLS) and Table 6 (IV) to suffer from a selection problem. Potential implications will be discussed in section 6.

Targets: Income versus Cumulative Hours

Table 8 describes the result of the estimated equation (14) that estimate the probability of stopping based on if either cumulative hours or aggregated income, or both, has been reached on any given day. Calculating the proxy target for each driver, the number of observations decreases to 2990.

The model's estimates imply that hours have a large impact on the probability of drivers ending their shift. For each additional hour driven on a given shift, the probability of quitting increases by approximately 6 percentage points. Notably, when the hour target is reached, the probability of stopping increases by approximately 20 percentage points. Both estimates have p-values<0.01.

TABLE 8: EFFECTS ON PROBABILITY OF STOPPING

	1
Cumulative hours	0.0589***
	(0.000787)
Aggregated income	8.04e-06***
	(2.11e-06)
Aggregated income > income target	0.00420
	(0.00512)
Cumulative hours > hours target	0.199***
	(0.00600)
Both targets reached	0.0125*
	(0.00646)
Observations	2,990
Adjusted R-squared	0.9169
Constant	-0.320***
	(0.00696)
Number of drivers	47

Table 8: Effects on probability of stopping – The table contains a simplified version of Crawford and Meng's (2011) reduced model of stopping probability. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Neither cumulative income nor whether the income target is reached is statistically significant on any conventional significance level. Moreover, if both targets have been reached, the probability of quitting increases, but it is a minor effect and with a p-value<0.1.

The result implies that reference-dependence might indeed be a factor in taxicab drivers labor supply decision. However, it questions the validity of the claim that the reference point is an income target, and suggests that a target based on cumulative hours worked could be an even stronger factor. The neoclassical model predicts that hours would influence the quitting decision but vary smoothly with realized income. The exhibited "jumps" when targets are reached are inconsistent with that prediction.[13]

6. Discussion

This section analyzes how our results differ from previous studies that have found evidence supporting the one-day target income hypothesis in the labor supply decision of taxicab drivers. We then try to assess why this is the case, starting with examining the specific cultural and market factors of Stockholm that could have an effect on the estimates. Lastly, taking on a more universalistic approach, we proceed by reexamining the evidence from the data, which suggest that much of the differences in the results can arguably be attributed to shortcomings in previous studies' data.

[13] We are not able to use a probit model, so the probability of quitting can take negative values and values that exceed 1. However, this is a common problem and affects very few of our actual observations.

Findings

To summarize the results, we find widely different outcomes compared to those of both Camerer et al. (1997) and Chou (2002). The only finding consistent with these earlier studies is the positive autocorrelation within days. Apart from that, the data exhibits substantial autocorrelation across days, positive wage elasticity across numerous models, and some effect on the extensive margin given high wage days. We also find some evidence that questions the relevance of the income target itself. As opposed to Farber (2005) these contrasting results are not due to differences in methodology, as they have been found using the same methods as in the original studies. Interestingly, they are similar to Stafford's (2013) study of comprehensive panel data on Florida lobster fishermen.

Intuitively, our results imply that drivers do not in fact set one-day income targets, and drive for longer hours when the wage per hour is high. Indeed, positive wage elasticity estimates are consistent with the standard intertemporal model of labor supply that predicts that drivers, seeing the wage being temporarily high, decide to continue working since the opportunity cost of leisure has risen (Lucas and Rapping 1969). Our results regarding autocorrelation between days suggest that wage over time indeed exhibits some pattern, and that wage varies across days. Looking at 91 days, we estimated positive autocorrelation even when controlling for the days of the week. It seems reasonable to use the flip side of Camerer et al.'s (1997) argumentation. They claim that a lack of autocorrelation would make it impossible for drivers to predict today's wage, which in turn would inhibit a substitution effect between days inherent in a longer decision-making time horizon. If we follow this argumentation, a positive autocorrelation would imply that it is possible that drivers have a two or three day (or even longer) target, making such a substitution effect both likely and strong. Our results thus question the assumption at the core of Camerer et al.'s (1997) argumentation that taxicab drivers make narrowly bracketed labor supply decisions. As previously described (see especially equation (10)), autocorrelation between days makes it impossible to distinguish whether taxicab drivers do in fact practice targeting over a longer time horizon than one day, or if the neoclassical theory holds.
Additionally, our simple reduced proxy target model using both aggregated income and cumulative hour targets, inspired by Farber (2005) and Crawford and Meng (2011), is consistent with their finding that cumulative hours, and not income targeting, is the main determinant of driver's stopping behavior. Indeed, although our model does not allow identifying which of either the proxy income or cumulative hours target is reached first, the drivers' probability of ending their shift increases sharply once the cumulative hours target is reached. The effect is much smaller for the income target. This is in line with Crawford and Meng's (2011) argument that there is more to the labor supply decision facing taxicab drivers than, as assumed in the original NYC study, income.

All in all, our analysis suggests that the labor supply behavior of Stockholm taxi cab drivers is inconsistent with a one-day target income hypothesis. Although the significantly positive wage elasticity is consistent with the standard neoclassical model of labor supply, we cannot rule out the possibility that a longer decision-making time horizon (as made possible by positive autocorrelation of wage across days) or targeting based on cumulative hours might implicate that a general reference-based model remains a plausible factor.

A Consequence of Market and Cultural Factors?

As has been previously touched upon, and although this paper's title might suggest otherwise, the Stockholm taxi market is not perfectly similar to the NYC market. By this effect, it could be argued that cultural and market differences between for instance the United States and Sweden might, to some degree, explain our contrasting findings.

First, one could argue that the incentive structure for drivers is somewhat different. Indeed, the NYC taxicab driver leasing arrangement, in which drivers pay a fixed fee for the cab plus fuel and maintenance costs but keep all the fare income, is described by Farber (2005, p.49) as being "close to the incentive theorist's first-best solution to the firm-worker principal-agent problem of selling the firm to the worker" since it internalizes the costs and benefits of working for the driver. In our sample however, the taxicab drivers do not pay a daily fee for access to a taxi vehicle, but are granted the cab over longer periods of time in exchange for a cut of the total fares. Additionally, Stockholm drivers are not paid on a day-by-day (or rather fare-by-fare basis) such as in the US or Singapore, but instead typically receive a paycheck at the end of the month. This combination of factors arguably makes both the driver's work input and output less cognitively distinct since they are both more abstract (a percentage of the income rather than a fixed fee) and distributed over a longer time horizon (pay at the end of the month instead of as a direct and distinct consequence of the daily output). This could potentially push Stockholm drivers to adopt a longer decision-making time horizon.

Secondly, as previously mentioned, the nature of demand is somewhat different on the Stockholm taxi market. The demand is relatively low, possibly because taking a taxi in Stockholm is expensive compared to a lot of other cities worldwide. This could have direct implications on our estimates. For instance, one of our contacts at the taxi franchise company told us that targeting might make sense when clients are abundant, but seems like an unlikely strategy in the Swedish market with its relatively low demand; his intuition was that when Stockholm drivers get "on a roll" and find many clients, they keep driving. The taxi industry also relies much less heavily on "cruising" and instead lends a much larger importance to booked trips (distributed equally between government-sponsored trips, individual customers and corporate clients). For such fares, the call center automatically chooses a driver depending on his location (how close he is to the customer) and how long he has been waiting in that particular area. This creates an incentive for drivers to wait for the next trip even if the previous hours have been slow. The implication is that drivers will work longer hours than they would in a "cruiser's market" such as New York or Singapore. While we cannot measure the effort drivers put in their work, Camerer et al. (1997) argue that it is harder work for NYC taxicab drivers to search for a passenger (since "cruising" requires both attention and skill) than it is to carry them, and therefore reject the notion that increasing disutility of effort explains why drivers might quit early on high-wage days. Arguably, the opposite is true in Stockholm where taxicab drivers can simply wait until a booked trip is allocated to them. In this line of thought, driving during high wage days requires more effort than driving during slow days, which could theoretically decrease the wage elasticity. However, this logic does not correspond to the narrative of the people in the taxi industry we have spoken to who seem to agree that the boredom of waiting for a call is just as wearying as being on the road looking for passengers. The implication of the utility function of effort is unclear and beyond the scope of this paper.

Thirdly, while tipping is an expected and important component of taxi payment in many parts of the world (and not least NYC), it is not in line with Swedish culture and norms. According to the taxi franchise that granted us access to the data, tipping is rare and typically averages about 3% percent of earned income per day and driver, a claim confirmed by several taxicab drivers. The implication is that probable measurement error in income due to the omission of tips in studies such as Camerer et al.'s (1997) does not affect our data.

Fourthly, there could be a difference in the number of undocumented trips missing from the data as a result of dishonest drivers. Because of the heavy tax burden and income cut taken by the fleet company, there are arguably larger incentives for Swedish drivers to take passengers without registering it in the taximeter. This could have direct implications on our estimates. On the one hand, if we assume that such illegal driving is a systematic behavior across all drivers and independent of hours worked across all days, this would lead to an overall slightly lower reported wage than what is actually the case and not cause bias in our estimates. However, this is a liberal assumption. For instance, if drivers who are working shorter hours are more prone to engage in illegal driving, their earned income would be understated compared to the rest of the drivers, which would lead to overstated wage elasticity estimates. Although the taxi franchise company monitors individual driving patterns in order to identify and dissuade illegal driving behavior (a system they themselves consider highly successful), we still cannot capture or rule out this effect in our data.

Fifthly, differences in liquidity constraints between markets can have an effect on exhibited labor supply behavior. On a macro level, ignoring borrowing constraints has been shown to decrease wage elasticity estimates by as much as 50% as compared to the underlying elasticity (Domeij and Flodén 2006). The fact that some drivers in NYC pay a daily fixed fee of 80 dollars to gain access to a vehicle could make the liquidity constraint a valid concern, since drivers might be forced to work long hours on bad days in order to collect the required amount to pay the next day's fee as well as fuel and maintenance costs. Drivers stopping behavior would then be determined by liquidity constraints and not due to income targeting, resulting in overstated negative wage elasticity. Arguably, while the pressure of liquidity constraints might play an important role in previous studies on taxicab drivers, it has no significant effect in our sample since Stockholm drivers pay for neither for fuel nor any leasing fee and do not receive their wage at the end of the day.

Lastly, one can imagine that cultural differences might have an effect on the behavioral strategies adopted by taxicab drivers. However, since Stockholm cab drivers often are not Swedish by birth but originate from countries such as Iraq, Iran or Turkey, it is difficult to make broader comments about the impact of specific aspects of Swedish culture. Interestingly, this also applies to taxicab drivers in New York City who often originate from India, Pakistan, Russia, Haiti or West Africa (Hodges 2007). The effects on the wage elasticity estimates are unclear, but the implications for the theories of labor supply merit further attention. According to Henrich et al. (2010), behavioral research (such as Tversky and Kahneman's) makes claims about human psychology and behavior based on samples drawn entirely from Western, Educated, Industrialized, Rich, and Democratic ("WEIRD") societies, assuming that there is little variation across human populations. The authors argue that there is in fact considerable variation "in the extent to which people value choice and in the range of behaviors over which they feel that they are making choices."

By this logic, it is not certain than one theory can uniformly explain the behavior of drivers the world over. However, this is a very broad issue, and something that cannot be addressed extensively in this paper.

In summary, while our list is probably not exhaustive, we can identify several factors that could affect the stopping behavior of Stockholm taxicab drivers. Whether the cumulative effect of these factors is strong enough to bias the estimates to the extent of making negative estimates appear positive is unclear, although arguably unlikely.
The incentive structure of the Swedish taxi market may somehow lead Stockholm drivers to adopt longer decision-making time horizons, which in turn might play a role in explaining the lack of evidence for a one-day time horizon in our data. On the other hand, if this is the case it questions whether the NYC taxicab driver profession, with its very specific incentive structure, really is an ideal laboratory for understanding general labor supply theory. Having now discussed the cultural and market differences that may affect our results, our comprehensive data also grants taking a more universalistic approach.

A Consequence of Comprehensive Data?

Reexamining the evidence from the data suggests that much of the differences in the results may be due to the comprehensive and precise nature of this study's data sets. We will consider the effect of breaks, the selection problem due to the participation decision, and the consequence of using smaller samples to estimate wage elasticity.

Breaks

Unaccounted breaks lead to exaggerated observed working hours, which in turn diminishes hourly wage. This leads to a downward bias in the estimated wage elasticity. Just as in Camerer et al. (1997), our data does generally not indicate when drivers are on a break; something they argue has a negligible effect on the estimates.
However, the fact that the taxicab drivers sometimes take out their identification card from the taximeter during such intra-shift breaks gives us an opportunity to produce an elegant test of the actual effect on the estimated wage elasticity.

In DAY2 the recorded intra-shift breaks (shorter than 150 minutes) are removed from the data. The result is that the positive wage elasticity drops to approximately 0.04 with a p-value<0.01. If we instead use DAY3, in which all breaks are ignored, the wage elasticity drops to practically zero. The regression results are displayed in Appendix 4. Clearly, neglecting the possibility of breaks is a very liberal assumption, and will have a significant impact on the resulting estimates. It is likely that the problem is even larger if drivers, for instance, take more or longer breaks on slow days. If all breaks were accurately displayed in our data, the estimated wage elasticity would be even more positive, a strong argument in favor of our findings.

Selection Problem

As reviewed in section 3, since taxicab drivers have the everyday choice of whether to work or take the day off, there is a distinct possibility of encountering a selection problem (Heckman 1979). If an unobserved factor affect drivers' decision to participate (work) on any given day and these factors also affect the decision of hours worked, the wage elasticity will be biased in the opposite sign of the correlation between the error terms in hours and participation decision. This implies that if unobserved shocks to participation and hours are positively correlated, wage elasticity will be downward-biased (Camerer et al. 1997). Camerer et al. (1997) circumvent this obstacle with the assumption of one-day decision-making time horizons and ultimately reject the notion that this factor plays enough of a role so as to alter the sign of the observed wage elasticity. However, the effect of the extensive margin has been reviewed in several studies (Oettinger 1999; Stafford 2013) that find that neglecting the participation decision leads to biased estimates. Similar to Fehr and Goette's (2007) experimental data on bicycle messengers, the Stockholm data allows for observing the days that taxicab drivers do not work, i.e. the daily participation decision.[14]

We have shown that there is autocorrelation across days, and that daily wage fluctuations follow some pattern. This implies that drivers, to some extent, can make active participation choices. Our results suggest that this is likely to be the case: more drivers choose to participate when expected wage is high. Still, the fact that we have neglected the extensive margin in our regression analysis does not exaggerate our estimates, since it would lead to a downward bias. We have attempted to apply a Heckman model but as discussed, finding an appropriate instrument for participation that does not affect income proved difficult (see footnote 3). Had we found such an instrument, the estimates of wage elasticity in our study would likely have been even more positive. Crucially, and unlike our study, Camerer et al.'s (1997) wage elasticity estimates are prone to be overstated since they are negative.

Subsample Analysis

As previously stated, the randomly selected subsamples SUB1 and SUB2 are created with the goal of replicating the data sets for which Camerer et al. (1997) find negative wage elasticity estimates (namely TLC1 and TLC2, which span 8 and 3 days respectively). For this purpose, hours worked are calculated as the time difference between the first insertion and last extraction of the driver's identification card on a given day. Using the method proposed in section 3, we can examine SUB1 and SUB2 in order to identify the effect that such smaller and less precise samples can have on the estimates. The results are presented in Appendix 5.

We begin by reexamining the assumption of autocorrelation across days. The regression, presented in Table 11, exhibits no correlation between days and hourly wage in the subsamples. Even though the estimated effect on today's wage given the previous day's wage is positive (0.253 and 0.142) the t-statistic is practically zero, implying statistical insignificance. This is not the case in the complete data set, as DAY3 exhibits equally significant positive autocorrelation across days (Table 12). Thus, by adopting the limited nature of the Camerer et al. (1997) study, the data becomes consistent with their findings of insignificant autocorrelation.

But what about the wage elasticity? Figure 9 in Appendix 5 displays a slightly negative log hour - log wage relationship in SUB1. Results for the OLS estimates are shown in Table 13 and estimates controlling for measurement errors can be found in Table 15. Surprisingly, we do in fact find substantially negative wage elasticity, with SUB 1 estimates ranging from -0.119 to -0.194 with p-values$<$0.05. For SUB2, the estimates range between -0.03 and -0.11, but with p-values ranging from 0.08 to 0.25 (Table 14). For increased clarity, we continue our analysis with SUB1 (we will return to the SUB2 estimates shortly). Including weather conditions (but neglecting the high temperature variable since it affects none of the days in the sample), the estimate becomes more significant and negative. In addition, using FE has a much larger effect

[14] Oettinger (1999), in his study on stadium vendors, look only at day-to-day participation and not intra-daily labor supply decisions.

than on DAY1, implying that individual heterogeneity is more prominent in SUB1. To control for measurement error, we run the same regressions using the reported hourly wage of other drivers working during that particular day as an instrument. First stage regression indicates that other drivers' reported wage is a strong instrument for individual wage and the joint test of the null hypothesis that all coefficients are zero can easily be rejected. As predicted, the reported estimates in Table 15 are even more negative: the wage elasticity now ranges from -0.158 to -0.272, and all estimates have p-values<0.10 (p-value<0.01 for model 4).
To summarize, using the randomly selected subsamples of our data leads to estimates that differ from our general findings but are consistent with the results of Camerer et al. (1997) and Chou (2002). No evidence for autocorrelation is found and one of the two subsamples exhibit significant negative wage elasticity.

Why then do DAY1 and SUB1 differ so substantially? Three main reasons come to mind. First, when testing for autocorrelation, data set DAY1 uses a timeframe of 91 days, and therefore wage fluctuations over time can be estimated with much more precision and it is possible to identify patterns. One can reasonably assume that drivers will act upon such patterns to substitute work and leisure in somewhat efficient ways. Only looking at a limited range of days, such as under two weeks, makes precise and statistically significant approximations less likely. Secondly, for drivers to work multiple shifts during a single calendar date is relatively common in our data; by ignoring the inter-shift breaks (not captured in such summary statistics data as TLC1 and TLC2) results in overstating hours worked. Camerer et al. (1997) only find negative wage elasticity in two out of three samples, and the two samples that showed negative wage elasticity were summary statistics, computed not by the authors themselves but by the TLC. It is possible that their data set exhibit the same traits as ours, before correcting for multiple shifts. Lastly, when relying on a limited number of observations from a short time period (especially if the observations are of questionable panel data character), the resulting estimates are prone to be influenced by chance.[15] While it is very possible that further subsampling could lead to very different results (the estimates in the second sample SUB2, for instance, are much less negative and not statically significant), this only demonstrates the importance of using comprehensive data.

Previous literature on the labor supply of taxicab drivers relies on few observations for each driver distributed over short periods of time. The results from our two randomly selected subsamples are clear indications that such limited data sets can lead to misleading conclusions. Additionally, the analysis suggests that neglecting breaks, and not accounting for the effect of the extensive margin, might lead to downward biased estimates.

[15] The data sets SUB1 and SUB2 are smaller then the ones used by Camerer et al. (1997) and Chou (2002) in terms of number of observations, although not in terms of days. This reflects the panel nature of our data.

7. Conclusion

Our analysis suggests that the labor supply behavior of Stockholm taxicab drivers is inconsistent with a one-day target income hypothesis. Based on extensive and precise log system data, we find significantly positive wage elasticity estimates ranging from 0.13 to 0.15. While the data exhibits a positive autocorrelation in hourly wage within days consistent with Camerer et al. (1997), it also reveals significantly positive autocorrelation in the wage across days ranging from 0.36 to 0.51. This implies that drivers can make rational predictions of future wage levels. Additionally, examining the influence of the participation decision suggests that taxicab drivers work more on high wage days, which might lead to downward biased wage elasticity estimates. An audit of the effect of breaks also suggests that our wage elasticity estimates are downward biased. Although the significantly positive wage elasticity is consistent with the standard neoclassical model of labor supply, we cannot rule out the possibility that a longer decision-making time horizon (as made possible by positive autocorrelation of wage across days) or targeting based on cumulative hours (which a simple proxy target analysis finds support for) might implicate that a general reference-based model remains a plausible factor.

While differences in the incentive and market structures might have an impact on the driving behavior and decision-making time horizon of Stockholm taxicab drivers, the paper suggests that the contrasting results compared to previous studies might mainly be due to more comprehensive data. Furthermore, if negative wage elasticity only appears in the specific context of daily leasing fees and fare-by-fare reimbursement, the validity and generalizability of such findings are limited since the incentive structure of NYC taxicab drivers arguably applies to a very restricted segment of the general labor market. While prospect theory may offer accurate descriptions of risk attitudes in experimental settings, some question whether its predictions remain accurate outside the laboratory (Barberis 2013). In the real world, people may have considerable experience making the decision at hand and stakes can be significantly higher. Additionally, behavioral research often makes claims about human psychology assuming that there is little variation across human populations (Henrich et al. 2010). The fact that taxicab drivers in the West often originate from less industrialized countries might implicate difficulties in making cross-cultural comparisons of labor supply decisions in the research focusing on this specific profession.

The above findings, as well as the limited scope of this paper, highlight several areas that could benefit from closer examination. Future research can, for instance, examine decisions between days and weeks to determine whether targets are applied to longer time horizons, and whether this is in turn affected by different incentive structures. Additionally, identifying an instrument that has an effect on the participation decision but no effect on income could help adjust for the selection problem inherent in studies on the labor supply of taxicab drivers. While we weren't able to compute Crawford and Meng's (2011) structural model, our simplified reduced form suggests that the stopping behavior of Stockholm taxicab drivers depends more on reaching a cumulative hours target rather than an aggregate income target. Applying the complete model with comprehensive data could further contribute to a better understanding of reference-based preferences.

8. Summary

The comprehensive and precise data on Stockholm taxicab drivers allow us to revisit the evidence supporting the one-day target income hypothesis of Camerer et al. (1997) as opposed to the standard neoclassical model of labor supply. The contributions can be summarized as follows:

1) Hourly wage is significantly positively autocorrelated across days, implying that taxicab drivers can rationally make predictions about future days' wage levels and thus substitute work and leisure across days. This questions the one-day time horizon assumption.

2) Additionally, we find that Stockholm taxicab drivers do not in fact display negative, but significantly positive wage elasticity estimates. This is consistent with the neoclassical model of labor supply, but without support for the one-day time horizon assumption, the general target income hypothesis cannot be rejected.

3) As our data enables us to look at the participation decision of taxicab drivers, we examine the possible implications of the extensive margin. Although we do not incorporate this factor into our general estimates, we can demonstrate that taxicab drivers seem to participate more on higher wage days, which might lead to downward biased wage elasticity estimates.

4) By using proxy targets for taxicab drivers' stopping behavior based on both aggregated income and cumulative hours, we find that that income has a minor effect relative to worked hours. This is in line with Crawford and Meng's (2011) findings.

Based on these findings, we conclude that the labor supply behavior of Stockholm taxicab drivers is inconsistent with a one-day target income hypothesis. While factors specific to the Stockholm taxi market might partially explain the contrasting findings compared to previous studies, we show that their results can arguably be attributed to shortcomings in previous data. Furthermore, if the one-day target income hypothesis only holds in the specific NYC taxicab driver's incentive structure, this would imply questionable generalizability of the theory.

9. References

Altonji, J. G. 1986. Intertemporal Substitution in Labor Supply: Evidence from Micro Data. *Journal of Political Economy*, 94: 176–215.

Alogoskoufis, G. and Manning, A. 1988. Wage Setting and Unemployment Persistence in Europe, Japan and the USA. *European Economic Review*, 32: 698–706.

Barberis, N. C. 2013. Thirty Years of Prospect Theory in Economics: A Review and Assessment. *Journal of Economic Perspectives*, 27(1): 173–196.

Benartzi, S., and Thaler, R. 1995. Myopic Loss Aversion and the Equity Premium Puzzle. *Quarterly Journal of Economics*, 110: 73–92.

Blundell, R., and Macurdy, T. 1999. Labor Supply: A Review of Alternative Approaches. In: O. Ashenfelter and D. Card, ed. 1999. *Handbook of Labor Economics*, Vol 3. Amsterdam: Elsevier Science, pp. 1559–1695.

Browning, M., Deaton, A. and Irish, M. 1985. A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle. *Econometrica*, 53: 503–543.

Camerer, C., Babcock, L., Loewenstein, G. and Thaler, R. 1997. Labor Supply of New York City Cab Drivers: One Day at a Time. *Quarterly Journal of Economics*, 112(2): 407–442.

Chou, Y. K. 2002. Testing Alternative Models of Labour supply: Evidence from Taxi Drivers in Singapore. *Singapore Economic Review*, 47(1): 17–42.

Cicchetti, C. and Dubin, J. 1994. A Microeconometric Analysis of Risk Aversion and the Decision to Self-Insure. *Journal of Political Economy*, 102: 169–186.

Colman, T. S. 1989. Unemployment Behavior: Evidence from the CPS Work Experience Survey. *Journal of Human Resources,* 24(1): 1–38.

Connolly, M. 2008. Here Comes the Rain Again: Weather and the Intertemporal Substitution of Leisure. *Journal of Labor Economics*, 26(1): 74–100.

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.

Domeij, D. and Flodén, M. 2006. The Labor-Supply Elasticity and Borrowing Constraints: Why Estimates are Biased. *Review of Economic Dynamics*, 9(2): 242–262.

Farber, H. S. 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.

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.

Forssén, K. 2008. Taxinäringen i tiden. *Fokus på Näringsliv och Arbetsmarknad.* Örebro: Statistiska Centralbyråns Avdelning för Ekonomisk Statistik.

Hardie, B. G. S., Johnson, E. J., and Fader P. S. 1993. Modeling Loss Aversion and Reference-Dependence Effects on Brand Choice. *Marketing Science*, 12: 378–394.

Heckman, J. J. 1974. Life Cycle Consumption and Labor Supply: An Explanation of the Relationship Between Income and Consumption Over the Life Cycle. *American Economic Review*, 64(1): 188–194.

Heckman, J. J. 1979. Sample Selection Bias as a Specification Error. *Econometrica*, 47(1): 153–161.

Henrich, J., Heine, J. S. and Norenzayan, A. 2010. The Weirdest People in the World? *Behavioral And Brain Sciences*, 33(2-3): 1–75.

Hodges, G. R. 2007. *Taxi!: A Social History of the New York City Cabdriver*. Baltimore: Johns Hopkins University Press.

Kahneman, D., Knetsch, J. and Thaler, R. 1990. Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy*, 98: 1325–1348.

Kahneman, D., and Tversky A. 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47: 263–291.

Kahneman, D. and Tversky, A. 1991. Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics,* 106: 1039–1061.

Killingsworth, M. R. and Heckman J. J. 1986. Female Labor Supply: A Survey. In: O. Ashenfelter and R. Layard, ed. 1986. *Handbook of Labor Economics*, Vol. 1. Amsterdam: Elsevier Science, Chapter 2.

Kőszegi, B. and Rabin, M. 2006. A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics* 121(4): 1133–1165.

Lucas, R. E. Jr. and Rapping, L. A. 1969. Real Wages, Employment, and Inflation. *Journal of Political Economy*, 77(5): 721–754.

MaCurdy, T. 1981. An Empirical Model of Labor Supply in a Life-Cycle Setting. *Journal of Political Economy*, 89(6): 1059–1085.

Maddala, G. S. 1992. *Introduction to Econometrics*. 2nd ed. New York, NY: Macmillan Publishing Company: pp. 461–464.

Mankiw, G. N., Rotemberg, J. J. and Summers, L. H. 1986. Intertemporal Substitution in Macroeconomics. *Quarterly Journal of Economics*, 100: 225-251.

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

Pencavel, J. 1986. Labor Supply of Men: A Survey. In: O. Ashenfelter and R. Layard, ed. 1986. *Handbook of Labor Economics*. Vol. 1. Amsterdam: Elsevier Science, Chapter 1.

Puhani, P. A. 2000. The Heckman Correction For Sample Selection And Its Critique. *Journal Of Economic Surveys*, 14(1): 53–67.

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

Samuelson, W. and Zeckhauser, R. 1988. Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty*, 1: 39–60.

Stafford, T. 2012. Labor Supply of Fishermen: An Empirical Analysis. In: Australian Agricultural and Resource Economics Society, *56th Conference.* Freemantle, Australia. 7-10 February 2012.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, Massachusetts: MIT Press, pp. 89–715.

Wooldridge, J. M. 2009. *Introductory Econometrics: A Modern Approach*. 4th ed. Mason, OH: South-Western Cengage Learning.

Appendix 1: Sample Characteristics

FIGURE 1: DISTRIBUTION IN HOURLY WAGE – DAY1

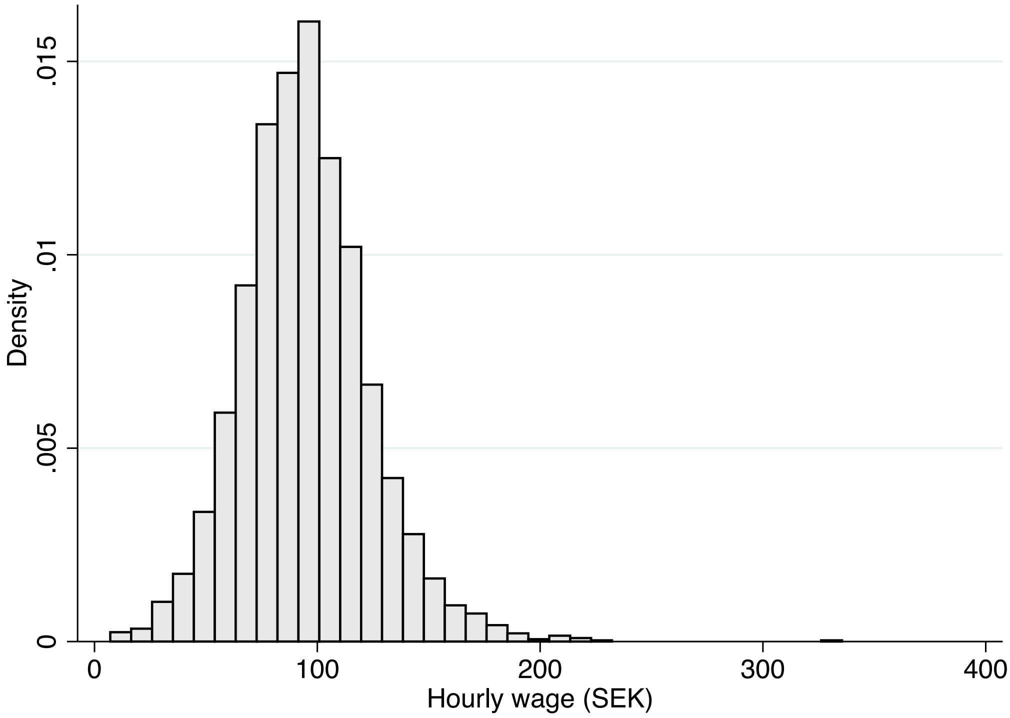

Figure 1: Distribution in hourly wage DAY1

FIGURE 2: DISTRIBUTION IN HOURS WORKED – DAY1

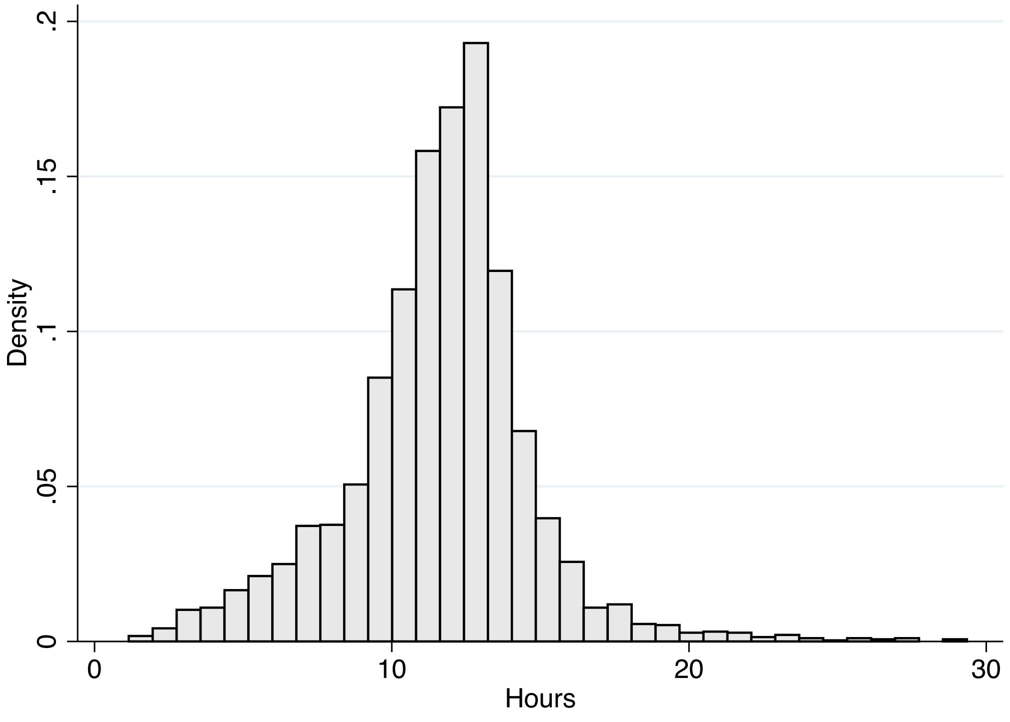

Figure 2: Distribution in hours worked DAY1

FIGURE 3: DISTRIBUTION IN INCOME EARNED ON A GIVEN DAY – DAY1

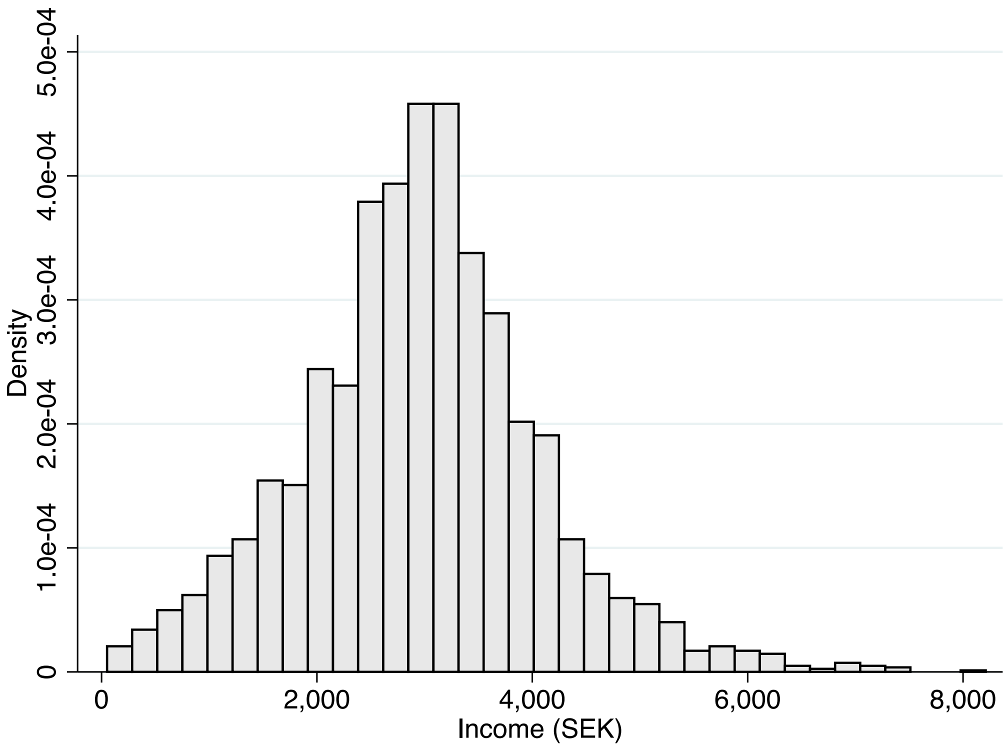

Figure 3: Distribution in income earned on a given day DAY1

FIGURE 4: KERNEL DENSITY ESTIMATE OF LOG HOURLY WAGE – DAY1

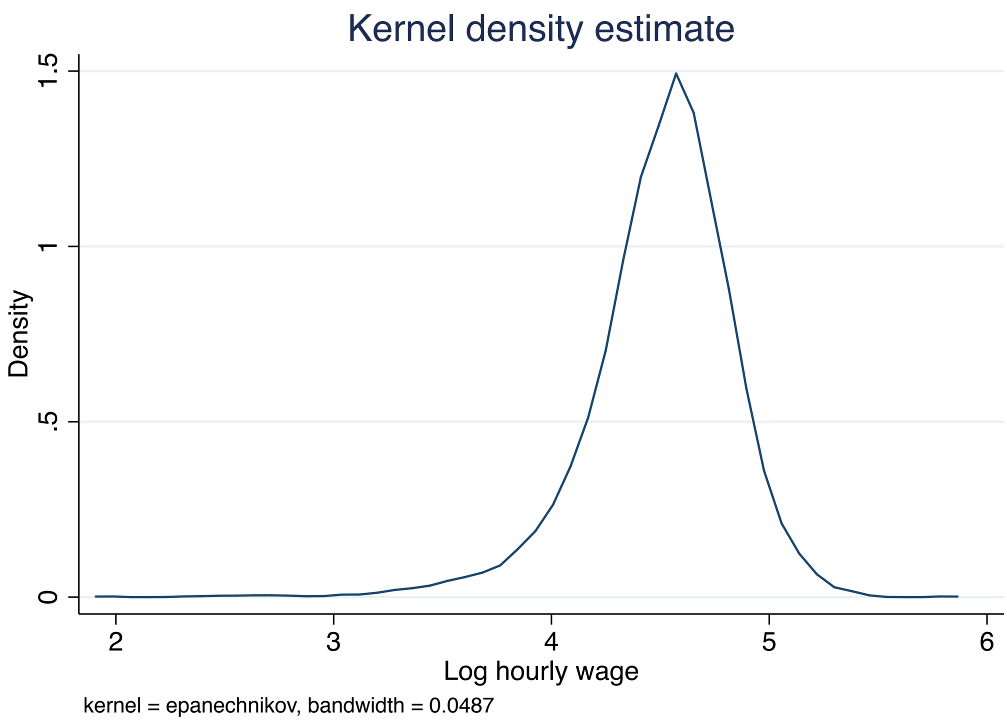

Figure 4: Kernel density estimate of log hourly wage DAY1

FIGURE 5: KERNEL DENSITY ESTIMATE OF LOG HOURS WORKED – DAY1

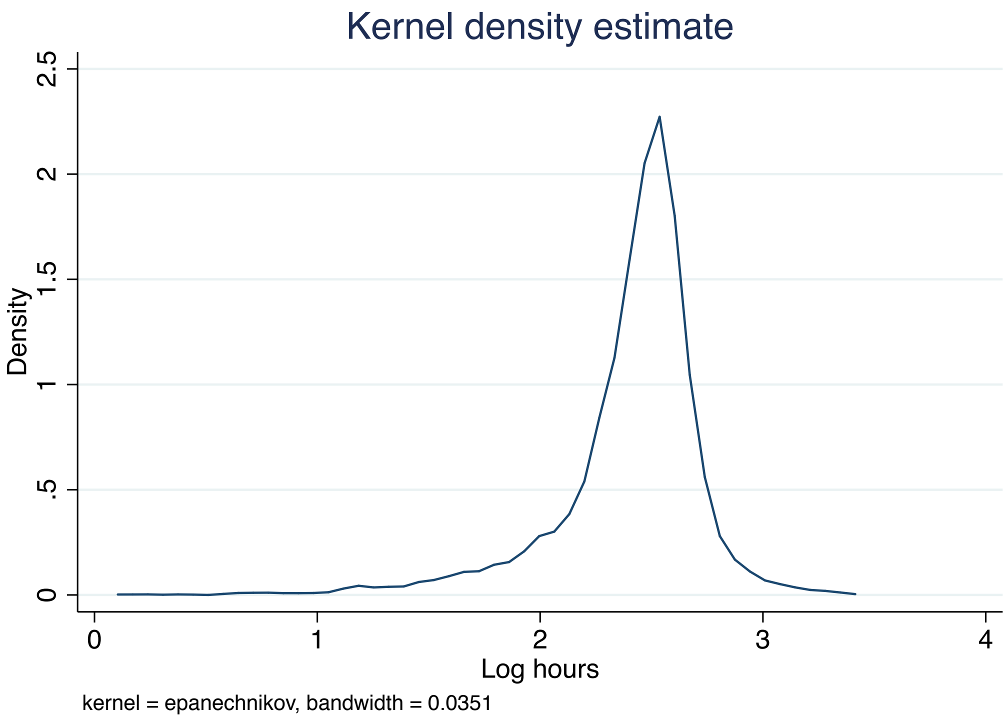

Figure 5: Kernel density estimate of log hours worked DAY1

Appendix 2: Correlation of Wage Within and Between Days

TABLE 2: CORRELATION BETWEEN MEDIAN HOURLY WAGE WITHIN DAYS – TRIP1

	1	2	3
Median wage *h*-1	0.162**	0.131**	0.128*
	(0.0641)	(0.0637)	(0.0654)
Median wage *h*-2		0.176***	0.178***
		(0.0637)	(0.0645)
Median wage *h*-3			-0.0165
			(0.0649)
Observations	239	238	237
Adjusted R-squared	0.026	0.057	0.055
Constant	217.0***	178.1***	182.1***
	(19.50)	(23.62)	(26.41)
Number of drivers	22	22	22

Table 2: Correlation between median hourly wage within days TRIP1 – The table contains autocorrelation within days by regressing the median wage at hour h on the median wage at hour *h*-1. Standard errors in parentheses and *** $p<0.01$, ** $p<0.05$, * $p<0.1$

FIGURE 6: MEDIAN HOURLY WAGE DAYS – DAY1

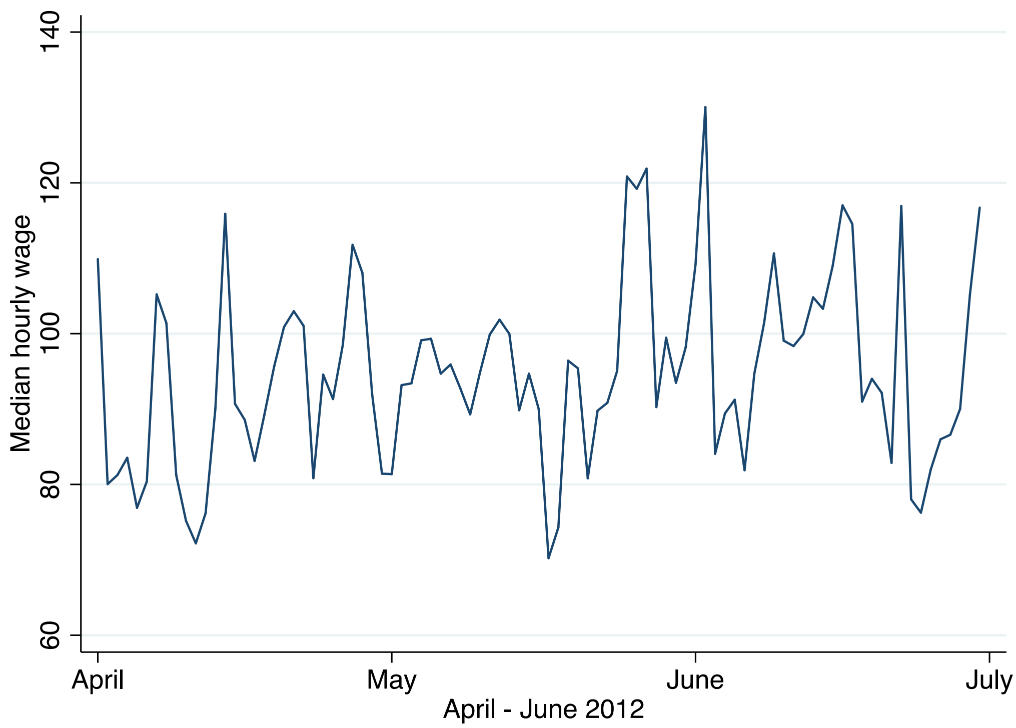

Figure 6: Median hourly wage – The figure exhibits median hourly wage during April-June 2012.

FIGURE 7: AVERAGE HOURLY WAGE – DAY1

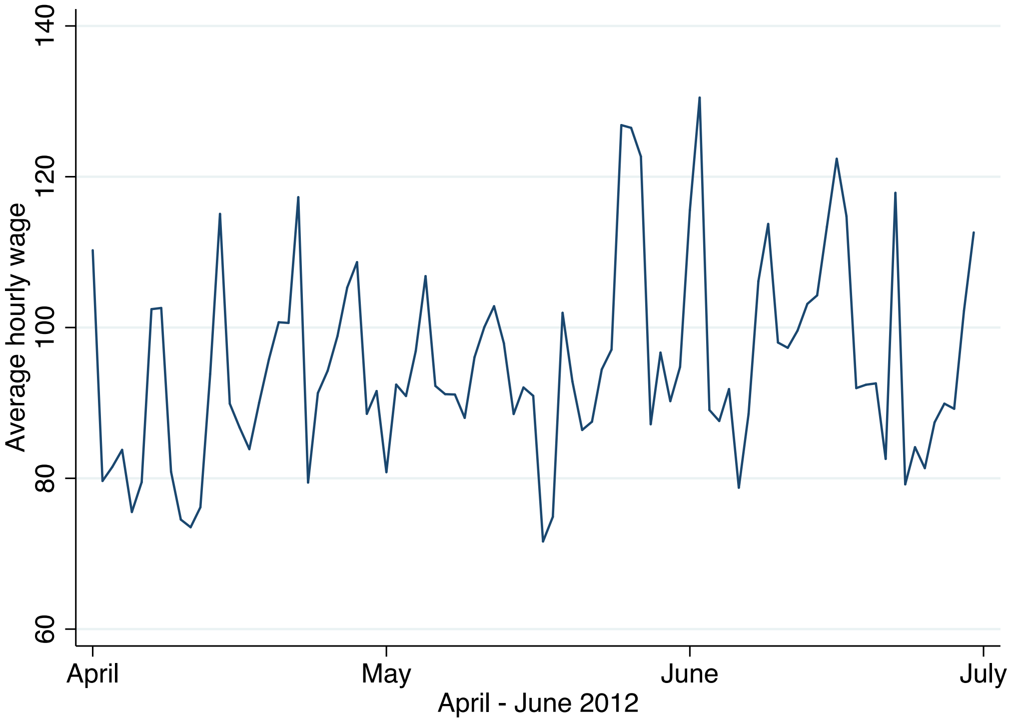

Figure 7: Average hourly wage – The figure exhibits average hourly wage during April-June 2012.

TABLE 4: CORRELATION BETWEEN AVERAGE HOURLY WAGE ACROSS DAYS – DAY1

Average wage *t*-1	0.417***	0.492***	0.471***	0.392***	0.308***
	(0.0972)	(0.106)	(0.109)	(0.101)	(0.114)
Average wage *t*-2		-0.125	-0.0688	-	0.194
		(0.105)	(0.119)		(0.116)
Average wage *t*-3			-0.102	-	0.0475
			(0.107)		(0.112)
Rain				0.455	0.451
				(2.216)	(2.215)
Tuesday				6.750	6.802
				(4.066)	(4.397)
Wednesday				6.962*	9.579**
				(4.020)	(4.563)
Thursday				6.495	9.499**
				(4.024)	(4.506)
Friday				19.59***	22.34***
				(4.168)	(4.547)
Saterday				21.45***	25.29***
				(3.910)	(4.663)
Sunday				8.597**	10.44**
				(4.038)	(4.513)
Observations	90	89	88	90	88
Adjusted R-squared	0.173	0.206	0.205	0.486	0.505
Constant	55.37***	60.40***	66.76***	47.52***	30.81**
	(9.315)	(10.91)	(12.80)	(10.53)	(14.07)
Number of drivers	47	47	47	47	47

Table 4: Correlation between average hourly wage across days – The table contains autocorrelation between days by regressing the median wage at day t on the median wage at day *t*-1. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1

Appendix 3: Wage Elasticity

FIGURE 9: LOG HOURS - LOG WAGE RELATIONSHIP – DAY1

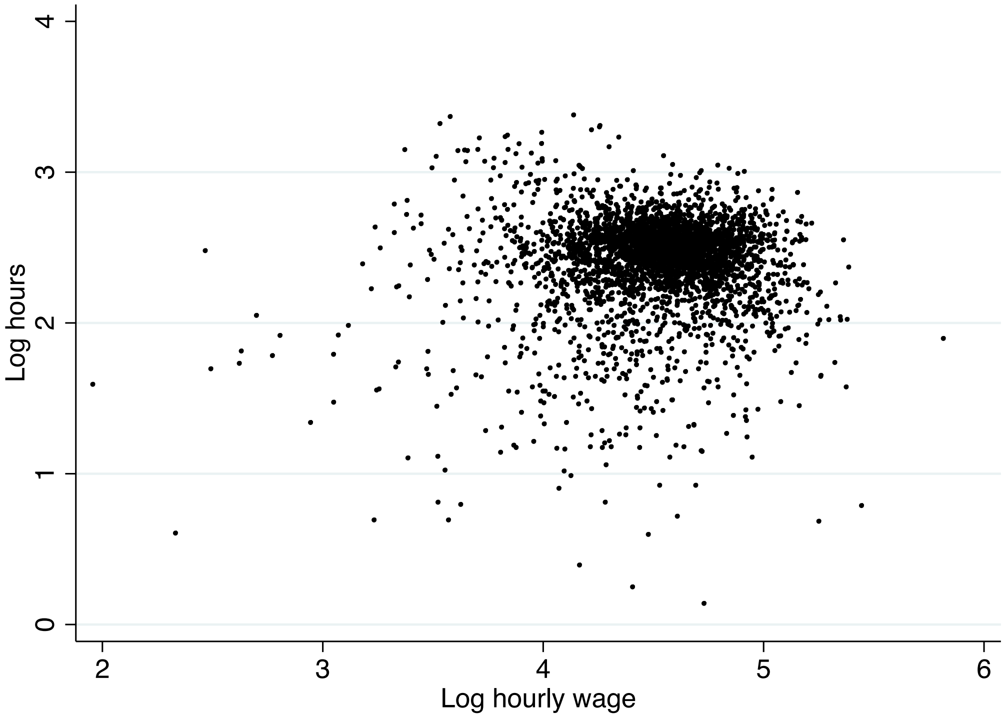

Figure 8: Hours-wage relationship DAY1 – Scatterplot containing the (log) hours - log wage relationship.

Appendix 4: Breaks

TABLE 9: OLS LOG HOURS WORKED EQUATION – DAY2

	1	2	3	4	5	6
Log hourly wage	0.0406**	0.0551***	0.0393**	0.0538***	0.0380**	0.0522***
	(0.0183)	(0.0185)	(0.0183)	(0.0186)	(0.0184)	(0.0186)
Rain			0.0119	0.0115	0.0128	0.0124
			(0.0116)	(0.0116)	(0.0116)	(0.0116)
High temperature					0.0145	0.0140
					(0.0118)	(0.0118)
Fixed Effect	No	Yes	No	Yes	No	Yes
Observations	3,530	3,530	3,530	3,530	3,530	3,530
Adjusted R-squared	0.01	0.01	0.01	0.01	0.01	0.01
Constant	2.222***	2.158***	2.222***	2.158***	2.219***	2.157***
	(0.0836)	(0.0836)	(0.0836)	(0.0836)	(0.0837)	(0.0836)
Number of drivers	47	47	47	47	47	47

Table 9: OLS log hours worked equation DAY1 – The table contains regression results estimating hours worked equation including control variables and estimates with fixed effects. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1

TABLE 10: OLS LOG HOURS WORKED EQUATION – DAY3

	1	2	3	4	5	6
Log hourly wage	0.00611	0.0178	0.00429	0.0160	0.00253	0.0140
	(0.0184)	(0.0186)	(0.0184)	(0.0187)	(0.0185)	(0.0187)
Rain			0.0183	0.0181	0.0196*	0.0194
			(0.0118)	(0.0118)	(0.0118)	(0.0118)
High temperature					0.0203*	0.0200*
					(0.0120)	(0.0120)
Fixed Effect	No	Yes	No	Yes	No	Yes
Observations	3,530	3,530	3,530	3,530	3,530	3,530
Adjusted R-squared	0.01	0.01	0.01	0.01	0.01	0.01
Constant	2.384***	2.333***	2.383***	2.332***	2.379***	2.329***
	(0.0840)	(0.0840)	(0.0840)	(0.0839)	(0.0840)	(0.0839)
Number of drivers	47	47	47	47	47	47

Table 10: OLS log hours worked equation DAY3 – The table contains regression results estimating hours worked equation including control variables and estimates with fixed effects. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1

Appendix 5: Subsamples

TABLE 11: CORRELATION BETWEEN MEDIAN HOURLY WAGES ACROSS DAYS – SUB1 AND SUB2

	SUB1	SUB2
Median wage *t*-1	0.253	0.142
	(0.258)	(0.378)
Observations	11	12
Adjusted R-squared	0.097	0.014
Constant	61.27**	79.39**
	(22.42)	(34.31)
Number of drivers	18	15

Table 11: Correlation between median hourly wage across days SUB1 and SUB2 – The table contains autocorrelation between days by regressing the median wage at day t on the median wage at day *t*-1. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1

TABLE 12: CORRELATION BETWEEN MEDIAN HOURLY WAGE ACROSS DAYS – DAY3

Median wage *t*-1	0.420***	0.490***	0.474***
	(0.0971)	(0.106)	(0.110)
Median wage *t*-2		-0.0994	-0.0552
		(0.105)	(0.120)
Median wage *t*-3			-0.0755
			(0.107)
Observations	90	89	88
Adjusted R-squared	0.175	0.206	0.203
Constant	54.14***	57.15***	61.59***
	(9.135)	(10.72)	(12.50)
Number of drivers	47	47	47

Table 12: Correlation between median hourly wage across days DAY3 – The table contains autocorrelation between days by regressing the median wage at day t on the median wage at day *t*-1. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1

FIGURE 9: LOG HOURS – LOG WAGE RELATIONSHIP – SUB1

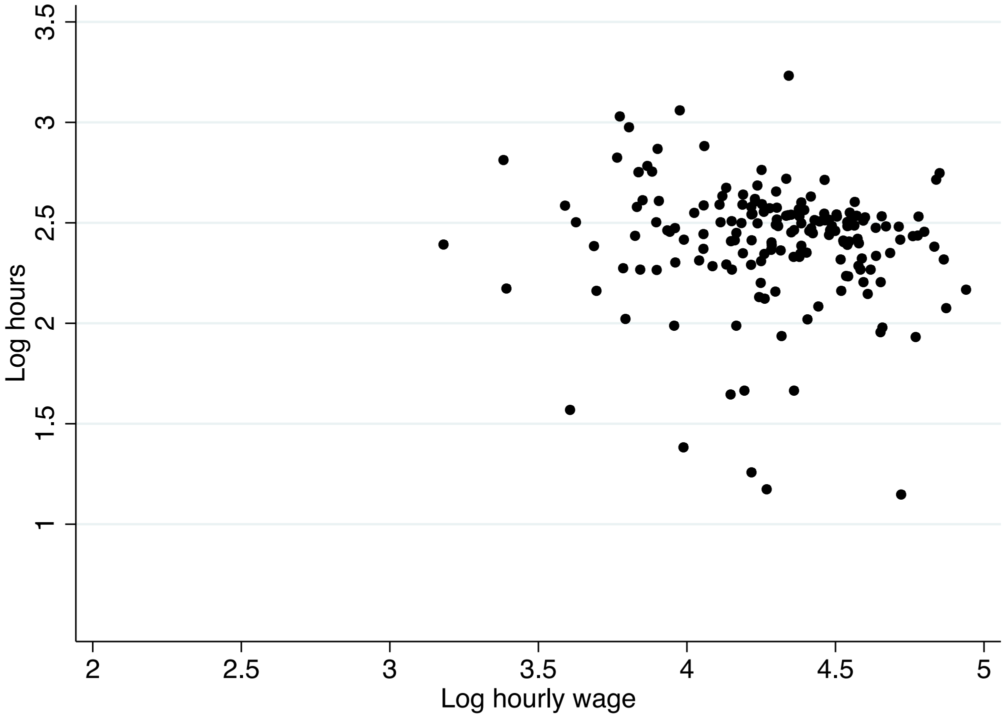

Figure 9: Hours-wage relationship SUB1 – Scatterplot describing log hours – log wage relationship.

TABLE 13: OLS LOG HOURS WORKED EQUATION – SUB1

	1	2	3	4
Log hourly wage	-0.119	-0.148*	-0.152**	-0.194**
	(0.0752)	(0.0840)	(0.0766)	(0.0861)
Rain			-0.0882*	-0.0959**
			(0.0460)	(0.0470)
Fixed effect	No	Yes	No	Yes
Observations	173	173	173	173
Adjusted R-squared	0.020	0.020	0.045	0.046
Constant	2.913***	3.043***	3.109***	3.302***
	(0.324)	(0.361)	(0.337)	(0.379)
Number of drivers	18	18	18	18

Table 13: OLS log hours worked equation SUB1 – The table contains regression results estimating hours worked equation including control variables and estimates with fixed effects. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1

TABLE 14: OLS LOG HOURS WORKED EQUATION – SUB2

	1	2	3	4
Log hourly wage	-0.0865	-0.0324	-0.0865	-0.0359
	(0.0675)	(0.0715)	(0.0675)	(0.0714)
Rain			-0.0592	-0.0560
			(0.0476)	(0.0474)
Fixed effect	No	Yes	No	Yes
Observations	178	178	178	178
Adjusted R-squared	0.01	0.01	0.01	0.01
Constant	2.851***	2.613***	2.885***	2.661***
	(0.299)	(0.316)	(0.302)	(0.318)
Number of drivers	15	15	15	15

Table 14: OLS log hours worked equation SUB2 – The table contains regression results estimating hours worked equation including control variables and estimates with fixed effects. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1

TABLE 15: IV LOG HOURS WORKED EQUATION – SUB1

	1	2	3	4
Log hourly wage	-0.158*	-0.186*	-0.224**	-0.272***
	(0.0894)	(0.101)	(0.0920)	(0.105)
Rain			-0.0974**	-0.107**
			(0.0467)	(0.0479)
Fixed effect	No	Yes	No	Yes
Observations	173	173	173	173
Constant	3.080***	3.205***	3.423***	3.640***
	(0.385)	(0.432)	(0.404)	(0.461)
Number of drivers	18	18	18	18
		FIRST STAGE		
	1	2		
Reported wage	1.285***	1.209***		
	(0.0643)	(0.0652)		
Adjusted R-squared	0.7199	0.7187		
P-value on F-test instrument of wage	0.00	0.00		
Fixed Effects	No	Yes		

Dependent variable is the log of average hourly wage.
Regressions also include weather characteristics as explanatory variables.

Table 15: IV log hours worked equation SUB1 – The table contains IV regression result estimating hours worked equation including control variables and estimates with fixed effects. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1