The Cost of Labor Supply Biases

Jack Welcome Fisher*

November 2, 2022

Preliminary: Comments welcome.

Abstract

This paper investigates an important dimension of the typical flexibility versus security trade-off that is used to frame self-employment. Namely, behavioral frictions that hinder workers from exploiting flexibility. I study the welfare cost of behavioral biases in intensive margin labor supply decisions for a group of self-employed workers who are free to pick their hours. In response to salient wage variation, workers' behavior implies a large and positive daily Frisch elasticity of 0.80 (s.e. 0.10). But in response to more common wage fluctuations their labor supply function is downward sloping for a range of wages, which is incompatible with even the most unrestrictive models of labor supply. In the spirit of Chetty-Looney-Kroft (2009), I use the salient Frisch elasticity to characterize preferences, and contrast outcomes under observed and optimal labor supply. A new sufficient statistics formula translates these deviations into daily welfare losses that are found to be economically significant; point estimates range from two to six percent of daily income. Annually, this can imply welfare losses of over £1,000 for those affected.

*j.w.fisher@lse.ac.uk. I would like to thank Giulia Giupponi, Supreet Kaur, Alan Manning, Guy Michaels, Steve Pischke, Daniel Reck, Johannes Spinnewijn, and John Van Reenen for many helpful comments and suggestions.

1 Introduction

Self-employment remains an important form of work in most economies. Across OECD countries, the median self-employment rate is 15 percent and the rise of the gig economy has precipitated an increase in the number of people who engage in self-employment as a source of extra income (Collins et al., 2019).[1] In low income countries, self-employment makes up a much larger proportion of total employment; for example, 76 percent of India's working population are self-employed.[2]

A key feature of self-employment is being one's own boss. As their own boss, the self-employed must organize their work but they may not necessarily do this efficiently. In this paper, I study how they manage a key input: their labor. Specifically, I examine how intensive margin labor supply decisions are made relative to an optimal benchmark and, most importantly, the welfare implications. I do this for a set of self-employed workers who pick their own hours and face regular variation in their return to work. Therefore, my approach and findings are most pertinent to the self-employed that operate in a similar environment. However, the results point more broadly to an important aspect of the typical flexibility versus security trade-off that is used to frame self-employment. Namely, behavioral frictions that prevent workers from fully exploiting flexibility.

Intensive margin labor supply responses to wage changes are the focus of a vast literature in economics (Blundell and MaCurdy, 1999). The results from this literature are often used to calibrate parameters in models that help to inform normative topics, such as the efficiency of income taxes (Keane, 2011). This rests on a revealed preference logic. Individuals are willing to work ε percent more when the wage increases by 1 percent, so the ordinal relationship between utility and work is governed by parameters that are a function of ε. But, in some settings, labor supply responses are inconsistent with even the most unrestrictive of neoclassical models. New York cabdrivers, who have received much attention from a section of the labor supply literature, have been estimated to exhibit negative daily Frisch elasticities (Camerer et al., 1997; Doran, 2014; Schmidt, 2018), positive daily Frisch elasticities (Farber, 2005, 2008, 2015), and to be influenced by hourly income effects (Thakral and Tô, 2017; Morgul and Ozbay, 2015), as well as to demonstrate behavior consistent with hours targeting (Crawford and Meng, 2011).[3]

These contradictory findings raise two key questions: Are they a consequence of behavioral biases, or deviations from normative assumptions embedded in neoclassical models? And if they are the product of behavioral biases, how do we distinguish between preferences and behavior so as to infer the cost of biases? Similar questions are pervasive in behavioral economics (Allcott and Taubinsky, 2015; Bernheim and Rangel, 2009; Bernheim and Taubinsky, 2018; Goldin and Reck, 2017; Mullainathan et al., 2011), but they remain unanswered in the labor supply context despite significant positive analysis.

This paper proposes and evinces a simple story about labor supply behavior in order to shed light on these questions. It argues that individuals vary daily labor supply substantially and in line with preferences, when faced with salient wage variation. However, if wage variation is not readily apparent, then they rely on suboptimal labor supply rules. Therefore, labor supply responses to salient wage variation can be used to make inferences about preferences over consumption and leisure. But general behavior will deviate from optimal behavior—it will be biased—since not all wage variation is salient. The extent to which general

[1] Source: the OECD (https://data.oecd.org/emp/self-employment-rate.htm).

[2] Source: the ILO (https://data.worldbank.org/indicator/SL.EMP.SELF.ZS).

[3] Negative short-run Frisch elasticities have also been found in observational studies of other settings (Chang and Gross, 2014; Chou, 2000; Agarwal et al., 2013; Nguyen and Leung, 2009), and experimental approaches have found results that are equally inconsistent with a neoclassical theory of daily labor supply decisions (Dupas et al., 2015; Fehr and Goette, 2007; Andersen et al., 2014).

behavior is biased depends on the proportion of wage variation that is salient and the labor supply rules that individuals use for the remaining wage variation. But, from the researcher's perspective, it is only necessary to observe average labor supply behavior, which is the confluence of these factors, and to contrast it with optimal labor supply in order to determine welfare losses.

I make three advances to establish this narrative using data on a set of self-employed taxi drivers in London. Firstly, I distinguish between different types of wage variation to separately infer preferences and behavior in the spirit of Chetty-Looney-Kroft (2009). London Tube strikes provide a source of salient wage variation, which suggest a large Frisch elasticity (0.80, s.e. 0.10). Conversely, plausibly exogenous wage variation stemming from the quasi-random allocation of jobs to drivers implies a much smaller elasticity (0.12, s.e. $<$0.01).

Secondly, I confirm that labor supply responses to the latter source of variation are suboptimal. A control function estimator reveals that a portion of drivers' labor supply function is downward sloping in a way that is reminiscent of income targeting. That is, hours increase with extreme wages and fall for wages around the mean. But downward sloping labor supply functions are not necessarily suboptimal because they can be the consequence of neoclassical income effects. Therefore, I exploit a permanent fare reform to estimate a Marshallian elasticity (-0.14, s.e. 0.04) that is too small in magnitude to explain the downward sloping daily labor supply function, which leads to the conclusion that labor supply is biased.

To further support the conclusion of biased decision making, I present survey evidence on how drivers determine their labor supply. Drivers readily subscribe to income and hours targeting but do not recognize the implications this has for the relationship between the wage rate and their hours. This suggests a cognitive dissonance between their behavior and preferences.

Lastly, I derive a behavioral welfare expression (BWE) for losses due to suboptimal behavior in continuous decision making settings, which imposes limited concerns over consumption and hours worked. This reveals the average marginal bias (Allcott and Taubinsky, 2015) of suboptimal labor supply in terms of the wedge in the intratemporal optimality condition between the going wage and the marginal rate of substitution for consumption and leisur (MRS).

To estimate the BWE, I use the Tube strikes Frisch elasticity and the Marshallian elasticity to parameterize preferences. I derive a condition under which this is appropriate since these statistics are identified by comparing behavior in general, which is biased, with optimal behavior on, for example, Tube strike days. Intuitively, the condition requires that biased behavior satisfies the intratemporal optimality condition in expectation. The presumption that behavior on Tube strike days is efficient seems a reasonable benchmark given that the implied magnitude of the Frisch elasticity is in line with other studies where biases are not a major concern. I calibrate the relative weight of consumption and work in the utility function to match either the observed level of hours during Tube strikes, or to keep average income across shifts constant under both biased and optimal behavior. Preferences are then contrasted with behavior given by the labor supply function, which is estimated with a control function estimator. Finally, welfare losses depend on the level of the wage, so an expected welfare loss is calculated by integrating the BWE with respect to a kernel estimated density of wages.

Expected daily welfare losses range from £2.09 to £5.29. The baseline estimate, which assumes all drivers are biased, points to a loss of £2.32 per day. However, when only a fraction of drivers are responsible for the aggregate deviation from optimal behavior, the estimated welfare loss rises rapidly for those affected.

I use survey evidence and lower bounds on biased labor supply elasticities to calibrate that one third of drivers supply labor optimally, which leads to an estimated loss of £5.29 amongst biased drivers—equal to six percent of average shift income. When these daily losses are accumulated over the course of a standard driver-year they point towards losses of up to £1,000 for biased drivers. This is comparable to annual welfare losses found in other settings, for example health insurance choices (Handel and Kolstad, 2015), which are considered large. Therefore, this paper suggest an economically significant impact of intensive margin labor supply biases on the welfare of self-employed workers.

I discuss the ramifications of these results for self-employed workers, organizations that contract with these workers, and for government policy. For workers, I note that hours targets dominate income targets because the latter introduces a costly covariance between the wedge in individuals intratemporal condition and the hours they work. Contracting companies could aid workers' optimization by making wages more salient and providing advice about reference dependent labor supply. However, firms may wish to exploit, rather than ameliorate, worker biases. This could necessitate a role for government; for example, laws could require firms to be transparent about expected earnings from working another hour.

Literature review. This research is related to several literatures. Most related is a strand of the behavioral economics literature that assesses the normative implications of policy interventions, such as commodity taxes and nudges (Thaler and Sunstein, 2009), via the estimation of reduced form statistics. This approach is explained in Mullainathan et al. (2011) and exemplified by Allcott and Taubinsky (2015), Berkouwer and Dean (2019), Chetty et al. (2009), and Spinnewijn (2015), among others.

The BWE derived below is estimable with a few sufficient statistics and identification of biased behavior, but it differs by considering a behavior change over a continuous choice variable, rather than a price induced behavior change for a binary variable.[4] The results in this paper also suggest that welfare losses increase on aggregate when a (locally) smaller proportion of drivers are biased, which provides another example of how heterogeneity in biases can exacerbate welfare losses (Taubinsky and Rees-Jones, 2018).

The empirical aspects of this paper lean on the many studies of income targeting amongst New York cabdrivers, for example, Camerer et al. (1997), Crawford and Meng (2011) , Farber (2005, 2008, 2015), and (Thakral and Tô, 2017), as well as the broader labor supply literature (Blundell and MaCurdy, 1999) and other methodological work (Wooldridge, 2015). I move beyond the income targeting literature in order to provide the first estimates of welfare losses due to behaviorally biased labor supply.

Lastly, this paper is also connected to the literature on alternative work arrangements. Two important papers in this literature are Mas and Pallais (2017) and Chen et al. (2017),[5] which focus on the value of flexible work. These papers use stated and revealed preference approaches, respectively, to find different results. Individuals are not willing to pay much for flexibility but their behavior indicates that they value it greatly.

The paper proceeds as follows. Section 2 discusses the institutional details of the empirical setting, and the data available. This is followed by an analysis of drivers' behavioral daily labor supply in section 3. In section 4, I argue and evince that these behavioral tendencies are not normative and generate welfare losses.

[4]Though I also derive an analogous BWE that considers price changes in appendix D.

[5]See also Chen et al. (2020) who use a similar approach to Chen et al. (2017) but allow for fixed costs of working, and correlation between wages and the cost of working.

In line with this argument, section 5 develops a theoretical apparatus to calculate this welfare loss and estimates the cost of labor supply biases. Section 7 concludes.

2 Institutional Details and Data

This section discusses institutional details and the data available. I also define the analysis sample and present some summary statistics to provide an introduction to the data.

Institutional Details. This paper uses data from a private hire taxi firm in London. The firm leases cars to its self-employed drivers,[6] who are allocated jobs to complete via an application on their mobile phone. These jobs are demanded by customers either on another application or over the phone. The car can be used for leisure purposes but may not be used for other professional reasons, such as serving other ride hailing businesses, so there is little risk of conflating intensive margin labor supply responses with switching between other work (Caldwell and Oehlsen, 2018).

Jobs are allocated by a central computer system; the system ranks a number of the closest cars according to how suitable they are for the job. The suitability of a car is determined by a number of factors which include the size of the car and whether the car is currently occupied. The top ranked car is then allocated to the job. While drivers can *de jure* turn down jobs, this is rare because they are disadvantaged in future job allocations if they do so. Moreover, the final destination of a job is not visible to drivers and this would likely be a key determinant of whether a driver would like to accept a job, or not. Nonetheless, drivers are able to determine when they finish work; at any point in time they can tell the firm that they are "Going Home" through the application. After this notification, the driver will receive no further jobs until they next log on or, if they are in the middle of a job, they will receive no further jobs after the completion of the current job.

Drivers are paid by the job and receive between 50 and 80 percent of the amount the customer pays. The amount received for a job is primarily determined by the distance of a job;[7] there is no separate compensation for the duration of a job. The amount also varies with the type of job,[8] the number of passengers, the number of stops, the time of day, and the location of the job. For some jobs, a value-added tax of 20 percent of the total transaction value is payable; drivers must pay their share of this from the fare they receive, which I account for in the analysis below. Drivers' earnings are paid out on a weekly basis by the firm.

Data. I observe job-level data from January 2012 to December 2019, which is electronically recorded from drivers' phone applications. Thus, each observation is a job and has variables for the driver ID, the total transaction value, the driver's fare, the start and finish time of the job, the start and finish location, and the distance of the job, among others. The raw data contains over 60 million jobs, which I cleaned analogously to Haggag and Paci (2014). This process removes rides if they contain anomalous variables, or are missing

[6]The leases are normally for a 12 week period and, over the span of this data, the average cost of a lease is around £200.00 per week. The firm provides incentives for drivers, which are similar to Uber's Driver Rewards, that lead to discounts on leases. Unfortunately, these have not been well documented and I have not been able to exploit this variation. The cost of a lease includes various maintenance costs and insurance for the driver, though drivers are responsible for fuel costs. As is common in the literature, incomes and wages in this paper are gross of these costs since they are not observed.

[7]A linear regression of the driver's fare on a 5th order polynomial of distance interacted with a dummy for the fare schedule yields an R-squared of 0.85.

[8]The type of job is affected by the type of car requested and supplied (e.g. Ford Galaxy or Toyota Prius) and the customer (e.g. corporate client or individual customer).

key variables, or are cancelled.[9] After cleaning, around 50 million jobs remain. From the clean job-level data, I constructed shift-level data by allocating jobs to the same shift, if they are completed by the same driver and are within six hours of one another (Farber, 2015; Thakral and Tô, 2017). If there is a break of more than one hour between two jobs, I deduct the excess duration of the break from the length of the shift in order to get a more accurate picture of actual labor supply. The shift wage is then calculated as the total income earned over a shift divided by the shift length. Finally, I cleaned these shifts in an analogous way to jobs, which leaves just under seven million shifts.

Analysis Sample. This paper focuses on intensive margin decisions made by drivers, so I use a sub-sample of drivers for whom I observe sufficient wage variation in their shifts. Specifically, I ensure that I observe a driver during at least one Tube strike and for 50 shifts either side of a fare reform that I use to estimate a Marshallian labor supply elasticity. This leaves me with around 3.5 million shifts driven by 2,600 drivers; I refer to this as the *balanced* sample. All analysis is conducted on this sample, unless otherwise stated. I also construct a *robust* sample, which imposes the same restrictions as the balanced sample but requires drivers to be observed during a minimum of four Tube strikes in order to check whether my selection along the extensive margin has important consequences for my results.[10] While extensive margins are certainly of interest more generally, this empirical setting is not best suited to investigate such questions because drivers in the sample essentially work full time.

Summary Statistics. I report summary statistics at the job and shift level. Table 1 shows job summary statistics from the balanced sample. The mean ride duration is just over half an hour, and 50 percent of all rides in my sample take between 21 and 42 minutes. In general, the rides are longer than those in the New York taxi context. The average job distance is 10.69 kilometers, and a driver receives £14.89 on average for a job.

Table 2 reports shift summary statistics. On average, a shift is comprised of just over six jobs, which translates to a working shift length of six and a half hours. If breaks exceeding an hour are included, this rises to eight hours. Shift income averages £91.62. The hourly wage, which is constructed as shift income divided by shift length, averages £14.32 with a standard deviation of £3.80. Table 3 shows sample sizes, and that shift variable means do not vary much between the full, balanced, and robust samples.

An examination of shift variable distributions is also informative of drivers' behaviors. Figures 1 and 2 illustrate the distribution of shift start and end times, respectively. The figures suggest a driver's typical day begins early in the morning and ends in the mid-afternoon. It is also common for drivers to begin shifts after lunchtime in anticipation of the afternoon rush, and to end shifts around midnight. These different shift patterns lead to a distribution of shift lengths that are displayed in figure 3; approximately half of all shifts fall between five and ten hours in length.

[9] The cleaning of job- and shift-level data is explained further in appendix A.

[10] I do not have any driver characteristics to compare my sub-sample with the full sample of drivers, but in table 3 I report shift variables means to show that they are very similar.

Table 1: Job Summary Statistics

Statistic	Mean	St. Dev.	Pctl(25)	Pctl(75)
Job time (minutes)	33.41	15.60	21.00	42.00
Job distance (kilometers)	10.69	8.85	4.53	13.69
Driver fare (£)	14.89	9.34	8.00	18.67

Notes: This table presents summary statistics of job-level variables from the balanced sample, which requires that drivers are observed during tube strikes and sufficiently often on either side of the fare reform.

Table 2: Shift Summary Statistics

Statistic	Mean	St. Dev.	Pctl(25)	Pctl(75)
Number of jobs	6.15	2.33	4	8
Shift length w/o breaks (hours)	6.56	2.41	4.64	8.27
Shift length w/ breaks (hours)	8.14	3.08	5.79	10.31
Shift income (£)	91.62	35.91	64.95	114.35
Shift wage (£/hour)	14.32	3.80	11.58	16.61

Notes: This table presents summary statistics of shift-level variables from the balanced sample, which requires that drivers are observed during tube strikes and sufficiently often on either side of the fare reform.

Table 3: Comparison of Samples' Size and Shift Means

Sample	Full	Balanced	Robust
Sample size	6,871,701	3,627,711	3,154,271
Number of jobs	6.12	6.15	6.18
Shift length w/o breaks (hours)	6.47	6.56	6.59
Shift length w/ breaks (hours)	8.04	8.14	8.16
Shift income (£)	89.3	91.62	91.24
Shift wage (£/hour)	14.14	14.32	14.21

Notes: This table compares means of shift-level variables from different samples: the full sample, the balanced sample, and the robust sample, as defined in section 2.

3 Empirical Motivation

In this section, I focus on intensive margin labor supply responses to daily wage changes via the estimation of two Frisch elasticities and a labor supply function. There are two primary reasons to focus on the daily level. First, wage rates *between* shifts are not correlated in an economically meaningful way,[11] so labor supply should be determined daily with reference to the wage relative to its usual level. Secondly, preempting the results below, it is hard to reconcile a downward sloping labor supply function with any other time horizon, as argued by Camerer et al. (1997).

I leverage two variables that induce wage variation which is temporary and exogenous: London Tube strikes and variation in the mean distance of jobs within a shift.[12] While these sources of variation induce theoretically equivalent variation, Tube strikes are seldom events whereas the mean distance of jobs varies regularly.

With this exogenous variation, I use different empirical frameworks to document contradictory labor supply responses to wage fluctuations, all in one setting. Firstly, I use Tube strikes in order to estimate a large and positive daily Frisch elasticity (0.80, s.e. 0.10). I refer to this as the *True* Frisch elasticity because, later in the paper, this statistic will be used to determine preferences. Secondly, I find a much smaller Frisch elasticity (0.12, s.e. <0.01) when I exploit variation in mean job distances within a shift. I refer to this latter estimate as the *Behavioral* Frisch elasticity because, as I will argue later, it reflects behavioral biases. Thirdly, I use a control function model to fully trace out the shape of the labor supply function, which reveals a negatively sloped portion of the labor supply curve. Then, I discuss and reject issues that could explain these different results and, in doing so, present estimates of strong hourly income effects from a probability stopping model (Thakral and Tô, 2017).

3.1 True Frisch Elasticity

London Tube strikes serve as a natural experiment to estimate a Frisch elasticity; they cause a significant change in the average wage, and are neither long enough nor severe enough to affect the marginal utility of income. Data on the dates and types of Tube strikes have been provided online by TfL thanks to a Freedom of Information request, and I verified these dates by checking coverage of media outlets at the time. There are two types of strikes: network wide and line specific. A network wide strike affects the capacity of all lines in the Tube network, while line specific strikes affect the running of only one or two lines. I restrict my analysis in this subsection to the period 1st January 2014 to 1st May 2015, which spans one fare schedule. This is for two interrelated reasons. First, this period contains the majority of network wide strikes, which causes the most systematic variation in the wage rate. Secondly, results are robust to time controls over this period and are also replicated for the full time horizon.

Given the data on Tube strikes, I investigate whether they are informative of wage rates at the driver-shift level. To do so, I regress the wage rate on a separate dummy for each strike, driver fixed effects, and controls for: start time of the shift, Ramadan,[13] bank holidays, days of the week, months and a flexible time trend. These are the main set of controls in the analysis. The results are shown in table 4. Five out of eight of the Tube strikes cause a statistically significant increase in the wage rate, and these are all network strikes. The largest effect was for the strike in April 2014, which caused a rise of five percent in the mean wage. These effects contribute to a first stage F-statistic of 29.99 although, for the results, the first and second stages are

[11] A £1 increase in today's shift wage increases the next shift's wage by £0.02 (s.e. <0.01) on average.

[12] As a result of the temporary nature of this variation, the marginal utility of income is likely fixed, which is required to estimate a Frisch elasticity.

[13] A large proportion of drivers are Muslim so controlling for Ramadan is logical though, in practice, it does not significantly alter results.

Table 4: True Frisch Estimation First Stage

	Dependent variable:
	log(Shift wage)
Network strike 04/02-06/02	0.007^{*} (0.004)
Network strike 28/04-30/04	0.055^{***} (0.004)
Network strike 09/05-10/05	0.010^{*} (0.005)
Network strike 13/06-14/06	-0.002 (0.005)
Network strike 01/07-09/07	0.033^{***} (0.003)
Network strike 09/07-15/07	0.014^{***} (0.003)
Line strike 22/08	0.001 (0.006)
Line strike 01/12-02/12	-0.005 (0.006)
First stage F-statistic	29.99
Observations	654,045
R^2	0.046
	*p<0.1; **p<0.05; ***p<0.01

Notes: This table presents coefficients on different London Tube strikes from the regression of driver wages on this variables, as well as driver fixed effects, a factor variable for start time of the shift, Ramadan, bank holidays, days of the week, months, and a 6th order polynomial time trend. The regression uses data from 01/01/2014 to 01/05/2015. Standard errors are clustered at the driver level.

jointly estimated below. The main conclusion of this investigation is that Tube strikes significantly affect drivers wages.

Therefore, the validity of my instrumental variable analysis rests on the exogeneity of Tube strikes, namely, that Tube strikes do not affect labor supply other than through their effect on the wage rate. Commonly cited reasons for Tube strikes include: insufficient pay, poor working conditions, and unfair dismissal of staff. There is no clear reason why the emergence of such concerns should be related to drivers' labor supply. However, the determinants of the specific dates of strikes may be different from their fundamental cause. It is often argued that the precise dates of strikes are set in order to cause maximum disruption.[14] This could imply that Tube strikes occur on dates when drivers are least able to expand their labor supply to meet the additional demand caused by a Tube strike. However, *ex post* this effect seems unlikely given the sizeable Frisch elasticity I estimate. Another concern is that Tube strikes cause major traffic. Obviously, traffic has implications for drivers' wages but a problem would only arise if drivers prefer more or less traffic aside from its impact on wages. I have no evidence of a systematic preference, but this concern is a valid caveat.

Before estimating a daily Frisch elasticity, I note that Tube strikes generally last two days. This is not problematic if drivers still make labor supply decisions one day at a time, which is likely given the labor supply function I estimate and the small sums of income at the daily level. However, if drivers do optimize over the course of a Tube strike, this could bias my daily Frisch elasticity estimate downwards. This is clear

[14] See, for example, https://www.telegraph.co.uk/news/2017/01/08/government-accuses-union-bosses-co-ordinating-transport-strikes/.

from a simple example: rather than driving a long shift during a Tube strike, a driver may decide to drive two normal shifts, which would cause me to estimate a zero response of hours to wages—despite a clear increase in labor supply. Again, this does not seem to be a significant problem given that my Frisch elasticity estimate is in the upper range of previously estimated elasticities.

Intuitively, identification rests on temporal variation within drivers; I compare the hours worked on a day with a Tube strike versus the hours worked on a day without a Tube strike, and their associated wages. In order to estimate the True Frisch elasticity, I implement a full information estimation of the structural equations (1) and (2) with driver fixed effects,[15] which is asymptotically equivalent to generalized 2SLS. The equations are as follows,

$$h_{i,s} = \alpha_i + \beta \cdot w_{i,s} + \Gamma \cdot X_{i,s} + u_{i,s}, \tag{1}$$

$$w_{i,s} = \delta_i + \Theta \cdot T_{i,s} + \Lambda \cdot X_{i,s} + v_{i,s}, \tag{2}$$

where subscripts i denotes a driver and s denotes the shifts, h is hours worked, w is the wage rate, T is a vector of length equal to the number of strikes with each element a dummy to indicate whether the shift took place during the respective strike, and X is a vector of controls that have been previously mentioned and are noted in the results table.

The estimates of β from equation (1) are shown in table 5, where standard errors are clustered at the driver level. The first column shows the OLS estimate, which is significantly negative because of endogeneity that is brought on for two reasons. Firstly, due to division bias because of the construction of the wage rate (Borjas, 1980) and, secondly, because demand and supply are conflated. The main estimate in the second column uses all strikes in 2014 and shows a significantly positive response of hours worked in a shift to the wage rate: a Frisch elasticity of 0.80 (s.e. 0.10). This estimate changes only marginally when restricting to network strikes or using the robust sample of drivers, which is shown in columns (3) and (4), respectively. When I use the full time horizon, which spans 2012 to 2019, I get a smaller point estimate as shown in column (5) but the 95 percent confidence intervals overlap considerably and the estimates are not statistically significantly different. The same is true when I vary the flexibility of the time trend, see columns (6) and (7), except this time the point estimates increase.

In summary, Tube strikes cause an increase in the wage rate, which significantly raises the length of drivers' shifts.

3.2 Behavioral Frisch Elasticity

Shifts in the composition of consumer demand act as a further instrument with which to estimate a second Frisch elasticity. Precisely, I use the mean distance of jobs in a driver's shift as an instrument for the wage rate. As discussed in section 2, a driver's fare for a job is primarily determined by the distance of the job; longer distance jobs have higher driver fares. Column (1) in table 6 shows that this effect maps to wages for shifts, so that shifts with longer distance jobs have higher wages. Since the distance of jobs is determined by the pick-up and drop-off that customers demand, different mean job distances across shifts reflect differences in the composition of demand from customers.

[15] Therefore, this regression identifies a positively weighted treatment effect for "changers", i.e. those drivers who experienced a wage change.

Table 5: True Frisch Elasticity Results

	Dependent variable:						
	log(Shift length)						
	OLS	All strikes	Network only	Robust sample	All schedules	4th order poly.	8th order poly.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(Shift wage)	-0.136^{***}	0.799^{***}	0.808^{***}	0.814^{***}	0.625^{***}	0.954^{***}	0.912^{***}
	(0.004)	(0.103)	(0.103)	(0.105)	(0.086)	(0.132)	(0.107)
Shift type dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ramadan dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank holiday dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DOTW dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Schedule dummies	No	No	No	No	Yes	No	No
Time trend poly. order	6th	6th	6th	6th	6th	4th	8th
Observations	654,519	654,519	654,519	613,941	3,466,056	654,519	654,519
R^2	0.114	0.013	0.013	0.012	0.013	0.008	0.009

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Notes: This table presents coefficients on wages from a series of regressions of shift length on this variable, as well as driver fixed effects, a factor variables for start time of the shift, Ramadan, bank holidays, days of the week, months, fare schedule, and a time trend of varying flexibility. An indicator for the fare schedule is also included in column 5. The wage variables is instrumented for with London Tube strikes. The regressions uses data from various periods and drivers. Standard errors are clustered at the driver level.

Table 6: Behavioral Frisch Estimation First Stage

	Dependent variable:		
	log(Shift wage)		
	Mean job dist.	LOM wage	Both
	(1)	(2)	(3)
Mean job distance	0.028*** (0.0001)		0.028*** (0.0001)
Leave-out mean wage		0.059*** (0.0004)	0.047*** (0.0003)
First stage F-statistic	$1.659 \times e^6$	$1.638 \times e^4$	$8.304 \times e^5$
Observations	3,466,056	3,466,055	3,466,055
R^2	0.552	0.290	0.571

*p<0.1; **p<0.05; ***p<0.01

Notes: This table presents coefficients on alternative instruments to the wage from a series of first stage regressions that control for driver fixed effects, a factor variable for start time of the shift, Ramadan, bank holidays, days of the week, months, fare schedule, and a 6th order polynomial time trend. The regression uses data from the full time horizon and balanced sample. Standard errors are clustered at the driver level.

This alternative instrument is valid if the mean distance of jobs in a shift does not affect the supply of hours aside from its effect on wages. I highlight two threats to this condition. First, longer jobs may systematically cause drivers more or less disutility; for example, longer jobs may be more fatiguing for drivers. There is no clear direction to this effect. Indeed, this instrument identifies a non-monotonic labor supply function in subsection 3.3, which is at odds with a preference for shorter, or longer, jobs driving the results. Second, if driver supply differentially effects the surplus for customers of different distance rides, the distance of rides may be correlated with labor supply separately from the wage. Broadly, this could work in two ways. If longer distance jobs initially have a higher value to customers then, when driver supply is low and waiting times are higher, only longer distance jobs will remain worthwhile. Or if the value of different distance jobs to customers is differentially affected by labor supply, and only long distance jobs remain beneficial. These concerns are alleviated by the fact that the firm prices according to distance—and so discriminates precisely on this variable.

I estimate the Behavioral Frisch elasticity analogously to Tube strikes, but I replace the vector of Tube strike dummies T in equation (2) with the scalar mean job distance d. Therefore, equation (2) is replaced with,

$$w_{i,s} = \delta_i + \mu \cdot d_{i,s} + \Lambda \cdot X_{i,s} + v_{i,s}, \tag{3}$$

The resulting estimate is shown in column (2) of table 7 and, again, standard errors are clustered at the driver level. The estimate is less than a sixth of the True Frisch elasticity implied by the Tube strikes, which is shown in column (1) for comparison. The standard errors in column (2) are small because mean job distance explains vastly more of the variation in wages. The Behavioral Frisch elasticity estimate is very statistically

Table 7: Behavioral Frisch Elasticity Results

	Dependent variable:			
	log(Shift length)			
	Strikes IV	Mean job dist. IV	LOM wage IV	Both IV
	(1)	(2)	(3)	(4)
log(Shift wage)	0.801^{***}	0.117^{***}	0.114^{***}	0.117^{***}
	(0.102)	(0.005)	(0.013)	(0.005)
Observations	654,045	3,466,056	3,466,055	3,466,055
R^2	0.013	0.081	0.082	0.081
				*p<0.1; **p<0.05; ***p<0.01

Notes: This table presents coefficients from regressions of shift length on the wage. The wage is instrumented for with alternative, and the regression includes controls for driver fixed effects, a factor variable for start time of the shift, Ramadan, bank holidays, days of the week, months, fare schedule, and a 6th order polynomial time trend. The regression uses data from the full time horizon, except in the strikes column, and balanced sample. Standard errors are clustered at the driver level.

significantly different from the True Frisch elasticity estimate. Moreover, the point estimate of 0.12 is at the lower bound of previous estimates of the *Hicksian* labor supply elasticity, which are typically in the range 0.1 to 0.3 (Keane, 2011). Columns (3) and (4) display estimates of the Frisch elasticity using the leave-out mean (LOM) wage and a combination of the LOM wage and mean job distance as instruments, respectively. The results barely change.

3.3 Control Function Model

In order to explore the difference between the True and Behavioral Frisch, I implement a control function model to trace out the full shape of the labor supply function using the mean job distance instrument. This requires two further assumptions, which are stronger than those necessary for instrumental variable estimation. Firstly, an independence assumption implies the mean shift distance contains no additional information about the shift length after conditioning on the wage. Secondly, a functional form assumption relates to how one controls for distance's effect on the wage. I state these assumptions more precisely as I outline the four steps that underlie my estimation procedure.

First, I residualize hours h, wages w, and mean job distances d with respect to all my controls; I denote these residualized variables with a dot $\dot{\bullet}$. Residualizing ensures that these variables are uncorrelated with the controls, but it does not guarantee independence. Therefore, this step is not without loss of generality because the identifying assumption in control function models revolves around independence. However, this simplification makes the analysis more convenient and, importantly, allows me to specify a more flexible control function.

Second, I specify a first stage where the residualized wage $\dot{w}$ is stated in terms of the residualized mean distance $\dot{d}$. There is a trade-off in how flexibly this relationship is formulated. On the one hand, making the relationship more flexible will make the independence assumption more plausible but, against this, it makes specifying a more flexible control function less feasible. With this balance in mind, I specify a fifth order polynomial for the first stage relationship,

$$\dot{w}_{i,s} = \zeta + \sum_{j=1}^{5} \eta_j \cdot \dot{d}_{i,s}^{j} + e_{i,s}. \tag{4}$$

Third, I specify a functional form for the object of interest $\mathbb{E}\left[\dot{h}|\dot{w}, \dot{d}\right]$. For this specification the aforementioned trade-off does not exist—the more flexible the better. I specify a tenth order polynomial for the relationship,

$$\dot{h}_{i,s} = \iota + \sum_{j=1}^{10} \kappa_j \cdot \dot{w}_{i,s}^{j} + \varepsilon_{i,s}. \tag{5}$$

Results are robust to this specification; the estimated labor supply function resembles a cubic function and so is not constrained in practical terms.

In the fourth and final step, I state and use the two assumptions that enable me to sketch out the labor supply function. Firstly, my approach requires an independence assumption: the joint distribution of the residuals from regressions (4) and (5) is independent of the residualized distance variable $(e, \varepsilon) \perp \dot{d}$. This allows me to write,

$$\mathbb{E}\left[\dot{h}\middle|\dot{w}, \dot{d}\right] = \iota + \sum_{j=1}^{10} \kappa_j \cdot \dot{w}^{j} + \mathbb{E}\left[\varepsilon\middle|e, \dot{d}\right] \tag{6}$$

$$= \iota + \sum_{j=1}^{10} \kappa_j \cdot \dot{w}^{j} + \mathbb{E}\left[\varepsilon\middle|e\right]. \tag{7}$$

The second assumption is the specification of the conditional expectation in equation (7). Often this is assumed to be linear, but given the size of my data I can be more flexible. As a baseline, I specify the control function as a quadratic, however, I check the robustness of my findings up to a fourth order polynomial. That is, I write,

$$\mathbb{E}\left[\varepsilon\middle|e\right] = \nu \cdot e + \xi \cdot e^2. \tag{8}$$

In order to implement this empirically, knowledge of $e_{i,s}$ and $e_{i,s}^2$ is required. I first construct the estimates $\hat{e}_{i,s}$ with the regression specified in equation (4). Then, I square both sides of equation (4) to yield a specification which can estimate $\hat{e}_{i,s}^2$. Note that this latter regression requires knowledge of $\hat{e}_{i,s}$. I proceed in similar steps when constructing the third and fourth order polynomial control functions. The number of parameters that are estimated in these latter regressions rises rapidly with the order of the polynomial, hence, the trade-off mentioned in the specification of (4).

The independence assumption required to implement a control function model is stronger than the exogeneity assumption required in an instrumental variables regression. This condition will not be perfectly met in practice, but I offer three defenses of the approach: (i) the output is consistent with results from the instrumental variable analysis, which suggests that (ii) although the assumption may be violated mildly, the results are still informative; and (iii) behavioral theories of labor supply, such as income targeting, predict a non-monotonic relationship between hours and wages, and this is the correct way to identify non-monotonicities.[16]

Given the steps above, I use the following regression to identify drivers' labor supply function,

$$\dot{h}_{i,s} = \iota + \sum_{j=1}^{10} \kappa_j \cdot \dot{w}_{i,s}^{j} + \nu \cdot \hat{e}_{i,s} + \xi \cdot \hat{e}_{i,s}^{2} + r_{i,s}. \tag{9}$$

The estimated labor supply function is illustrated by the blue line in figure 4. The grey bars show the distribution of residualized wage rates $\dot{w}$ used in the analysis and the dashed red line marks the mean wage rate. The function is striking; it is steeply rising for many wage rates—in line with the findings from the Tube strike analysis—but is *negatively* sloped for wage rates just above the mean wage. This result is robust to the specification of the control function, as shown in figure 5 which illustrates labor supply functions for different polynomial orders for the control function.

The labor supply function reveals that the Behavioral Frisch elasticity masks significant non-monotonicity in drivers responses to changing wage rates. Indeed, while the latter is qualitatively compatible with a neoclassical model of labor supply, even if not quantitatively, the downward sloping portion of the labor supply function is incompatible with optimal decision making, which I explain further in section 4.

3.4 Discussion

The hours-wages nexus is a form of aggregation which lends itself to familiar concepts in labor economics, however, it may over simplify the economic environment that drivers face. I discuss the impact of non-constant wages, schedule rigidities, and expectations in order to ameliorate these concerns.

Non-constant wage. In reality, drivers do not face a constant wage. Figure 6 illustrates the median hourly wage for drivers throughout the day for weekdays and weekends, separately. It is evident that wages have a peak in the morning, which is more pronounced during weekdays, and then rise initially steadily and then rapidly from midday to midnight, before declining until the start of the morning peak. This is problematic because the wage a driver has experienced through their shift may be different from the wage they would have received had they continued their shift, so that regressions of hours on wages do not capture the actual trade-off which drivers face. I present three pieces of evidence which suggest this is not a significant problem.

First, fares are a key driver of wages and exhibit significant auto-correlation. I regress current fares on fixed effects for the shift and the previous fare. Given the dynamic panel structure, I estimate this equation in first differences and instrument for the previous fare with its own preceding fare. A one standard deviation in the current fare level leads to an increase of the next fare of £2.76 (s.e. 0.13) in addition to level effects between shifts. This suggests there is strong persistence in the wage rate caused by auto-correlation in fares.

Second, alternative approaches that do not rely on constant wages produce results that are consistent with the findings here. For example, Farber (2015) and Thakral and Tô (2017) use probability stopping models in order to test if labor supply behavior deviates from the neoclassical benchmark. The focus of this approach is to test whether accumulated income in a shift is predictive of a driver ending their shift. In the

[16] Farber (2015) estimates a Frisch elasticity for wages close to the mean because of suspected non-monotonicities. However, such an approach is akin to a "forbidden regression". Stronger assumptions, as used here, are necessary.

neoclassical model, the amounts of income under consideration are not sufficient to affect the marginal utility of income, therefore, there should be no relationship between accumulated income and the probability of ending a shift.[17] I follow the approach of Thakral and Tô (2017), which non-parametrically controls for the disutility of hours worked, to assess the importance of income in determining the end of a shift. The approach regresses a dummy variable for ending the shift after a job on variables that summarize a drivers experience throughout the shift thus far, such as accumulated income and other job-level controls. A separate regression is run for different durations in a shift so that each regression is run on different data. These data bins are defined by 30 minute windows through a shift.[18] This method is akin to a local linear regression and flexibly controls for the disutility of work. The regressions are specified as,

$$\mathrm{q}_{i,j} = \pi_i + \xi \cdot y_{i,j} + \Upsilon \cdot Z + \rho_{i,j},$$

where the new subscript j denotes the job, the variable q is a dummy that takes value one if a driver ends their shift after the job, y is the logarithm of accumulated income, and Z is a rich vector of controls. The coefficients on log accumulated income ξ from each regression are shown by the black points in figure 7. All are significantly positive at durations through a shift, as in Thakral and Tô (2017).[19] Mean stopping probabilities are plotted with grey bars, which allows a naïve calculation of the elasticity of the probability of stopping with respect to income[20] in order to get a sense of the magnitude of these effects. At the average shift length duration, the elasticity is of the order 0.25, which indicates a significant behavioral response; this is consistent with a downward sloping labor supply function.

Third, a non-constant wage cannot easily explain the shape of the labor supply function that I uncover with the control function model. For example, drivers could perceive the wage to be mean reverting. In such a case, if a driver experienced a wage just above the mean, they would expect their wage to drop in future. As a result they may end their shift earlier on an above average wage because they anticipate lower wages in the future. This is incompatible with the fact that hours increase with wages either side of the downward sloping portion of the labor supply function.

Schedule rigidities. Schedule rigidities can cause non-convexities in drivers' decision making environments, which confound the interpretation of labor supply elasticities. The existence of schedule rigidities should show up as a low response of hours to wages. If the rigidities are fixed, then the daily Frisch elasticity will be estimated as zero. However, stochastic rigidities in combination with a non-constant wage can lead to the spurious estimation of non-zero elasticities. The logic is highlighted by Thakral and Tô (2017); if the wage rate is falling over time and drivers face a stochastic schedule rigidity, then a regression of hours on wages will yield a negative coefficient because longer shifts—that are only longer because of a shock to the schedule rigidity—necessarily entail a lower wage. This does not appear to be a major concern in my setting

[17] If wages are positively auto-correlated—as they are in the data—then accumulated income should lead to a lower probability of ending a shift, *i.e.*, a negative coefficient on accumulated income.

[18] For example, the first regression uses job level data where the jobs occur in the first 2.5 to 3 hours of the shift. The second regression uses job level data where the jobs occur in the first 3 to 3.5 hours of the shift, and so on.

[19] I also replicate the timing effects of these authors; income later in the shift has a greater effect on the probability of stopping, as illustrated in figure 8, and income at one point in time becomes less and less influential. However, unlike Thakral and Tô (2017) I find that the influence of earlier income can be negative on the likelihood of stopping. This finding can be reconciled with the authors' adjusting reference point, if the wage is auto-correlated throughout the day, as I argue, and drivers update their targets in a Bayesian fashion.

[20] This is naïve because it does not account for the covariance between mean stopping probabilities and behavioral responses to income.

for two reasons. Firstly, my results suggests that drivers adapt their hours very positively to variation in the wage rate due to Tube strikes and to low or very high wages. Secondly, most drivers end their shift between 10:00am and 11:00pm, which is a period of time where the wage rate is rising and continues to rise until midnight. There is still a significant mass of drivers who finish their shifts around midnight, which is around the time the wage falls substantially. I replicate the analysis without these shifts and the results are unchanged; for example, the Behavioral Frisch is not statistically significantly different.

Expectations. Another potential reason for the difference between the two Frisch elasticities is that one source of wage variation is expected, while the other is not. In particular, variation in the wage rate that comes from fluctuations in mean job distances may not be expected and, as a result, drivers do not have time to substitute labor intertemporally. This phenomenon would imply that hours do not respond significantly to variation in the wage rate. However, the labor supply function shows the shift length does respond strongly to wages, albeit in different directions depending on the level of the wage.

The inability of these factors to explain the disparate estimates of labor supply responses demands a more fundamental explanation, which I seek to give in section 4.

4 Behavioral Interpretation

I posit that there is a fundamental difference between the True Frisch and Behavioral Frisch: the True Frisch is the output of decision making when behavioral biases are attenuated, and the Behavioral Frisch and control function model identifies labor supply which is representative of general daily behavior—and subject to behavioral biases.

The Tube strikes and the mean job distance instruments cause wage variation that induces identical labor supply responses in typical models where drivers only optimize over hours worked and income, but in reality these sources of variation are very different in nature. Tube strikes are seldom events which receive substantial media coverage and, as a result, lead to salient wage variation with potential framing effects. These characteristics of Tube strikes attenuate driver biases, which allows them to behave closer to optimal.[21] Conversely, variation in mean job distance is very common and engages the standard heuristics used by drivers. This hypothesis simultaneously explains the non-neoclassical behavior evident in the labor supply function, as well as the probability stopping model, and the differences between the labor supply responses uncovered by the two instruments.

The behavioral nature of the labor supply function follows from its negative slope for a range of wages. The possibility that this is the result of any neoclassical income effects is ruled out by the estimation of a *less* negative Marshallian labor supply elasticity.[22] In a neoclassical model, this elasticity bounds all labor supply responses from below since it incorporates the income effect of price changes. Therefore, I can reject a neoclassical model of labor supply because the most negative elasticity from the labor supply function

[21] Salience is the quality of being particularly noticeable, therefore, its role in debiasing drivers can be thought of as a form of information provision. Allcott and Taubinsky (2015) provide a clear and concise discussion of how information provision can eliminate some biases, including biased beliefs, exogenous inattention, costly information acquisition, and costly thinking models. Moreover, the potential framing effects, that is the sense in which days with Tube strikes feel somewhat "special", may counter other biases, such as self-control issues.

[22] This estimation leverages a permanent fare reform that raised driver wages by 10 percent, see table A1. A detailed description of the estimation that yields the results in table A2, which are close to Ashenfelter et al. (2010), is contained in appendix B.

(-0.36) significantly exceeds the benchmark Marshallian estimate from table A2 (-0.14, s.e. 0.04). Further, the mean job distance instrument and other controls explain over half of all wage variation, which supports the interpretation of the labor supply function as indicative of general behavior. In contrast, the True Frisch elasticity falls into the range of previously estimated Frisch elasticities (Reichling and Whalen, 2012), where behavioral biases are not perceived to be severe, and easily exceeds the Marhsallian elasticity.

Survey evidence from a different set of drivers, who also pick their hours and are paid by the job, supports the postulation of a divergence between driver behavior and preferences too. As part of a general survey,[23] we asked these drivers how they determine their shift length. The options we offer are shown in table 8 and we randomize the order in which they appear to respondents. The two rules in bold and the frequency with which they are selected are of note; both rules imply the same relationship between the wage and shift length, however, working to earn a certain amount each day is vastly more popular than working less if pay is higher. This is not surprising because the latter is patently suboptimal, but under the guise of an income target it can appear quite reasonable. Further, when asked explicitly about how they would respond to temporary versus permanent wage increases, all drivers' answers were consistent with the ordering of Frisch and Marshallian elasticities implied by the neoclassical model. A similar cognitive dissonance amongst the drivers I analyze could sustain the divergence between behavior and preferences that is revealed in the conflicting Frisch elasticities and labor supply function.

If Tube strikes reveal decisions designed to maximize utility and the labor supply function implies general behavior, then I can use revealed preference logic in order to uncover the parameters that govern the trade-off between hours worked and income and, in turn, estimate the implications of suboptimal behavior for welfare. This approach does not require a firm stance on the behavioral biases that are afflicting drivers; I only need to identify behavior and preferences. However, the source of biases remains important for three reasons. Firstly, the source of biases determines the extent to which the salience of Tube strikes removes drivers' biases. Secondly, it can affect the interpretation of the welfare cost that I will estimate. For example, if behavior is not optimal because of optimization costs, the welfare loss I estimate could provide a lower bound on the cost of optimization. And thirdly, the source of biases could have consequences for the optimal policy response; if optimization costs are responsible for suboptimal behavior, then forcing individuals to cognize would likely not improve welfare.

There are numerous potential biases that could affect drivers, including: limited attention to the wage rate (Gabaix and Laibson, 2006; DellaVigna and Pollet, 2009; Bordalo et al., 2013), costly information acquisition since the going wage is not readily available (Gabaix et al., 2006; Sallee, 2014), costly thinking stemming from the cognitive burden of optimizing labor supply (Gabaix, 2014; Caplin and Dean, 2015; Sims, 2003), imperfect self-control (Gruber and Köszegi, 2001; Gruber and Kőszegi, 2004; Bernheim and Rangel, 2004), and biased beliefs (Spinnewijn, 2015). In the language of Allcott and Taubinsky (2015), I require Tube strikes to be a "pure nudge". That is, Tube strikes resolve drivers' biases without any confounding effects on labor supply. It is clear that salience will counter some biases more than others; information acquisition should be much less costly than otherwise since Tube strikes are very prominent to drivers. On the other hand, self-control biases may persist because the intertemporal properties of the costs and benefits to driving are unchanged.

Section 6, where I use the framework of Bernheim and Rangel (2009), formalizes how I leverage my empirical results to learn about preferences and behavior. But beforehand, in the next section, motivated by the

[23] More details of this survey are provided in appendix C.

Table 8: Surveyed Labor Supply Rules

Labor Supply Rules	Frequency
Work for a certain number of hours	126
Work to earn a certain amount each day	**123**
Work as long as possible	96
Work longer if pay is higher and *vice versa*	88
Work shorter if pay is higher and *vice versa*	**43**

Notes: This table presents the frequency of responses to alternative labor supply rules from a sample of 476 drivers. The two rules in bold imply the same qualitative relationship between the wage and hours worked.

idea of a divergence between behavior and preferences, I derive an estimable expression that quantifies the welfare cost of drivers' deviations from the optimal benchmark.

5 Behavioral Welfare Theory

Motivated by the empirical evidence presented in section 3 and the behavioral interpretation in section 4, I construct a simple static model to conduct a normative analysis of behaviorally biased labor supply. The model yields an object that captures an individual's change in utility, when labor supply moves from biased to optimal.[24] This object can be approximated by an expression which contains only a small number of sufficient statistics and is estimable given sufficient data. It reveals that the welfare cost of biased labor supply stems from the wedge in the intratemporal optimality condition, *i.e.* the difference between the wage and the marginal rate of substitution (MRS) between work and consumption. The model is static but the key dynamic effects, namely whether labor supply behavior translates to different consumption levels, can easily be captured in the empirical implementation.

I will use the language of drivers, consumption, and hours to match the empirical context of this paper but the potential applications of this expression are much more broad. Indeed, this formula is closely related to a continuous generalization of expressions for changes in consumer surplus stemming from a price change of a binary good (Allcott and Taubinsky, 2015).[25]

5.1 Behavioral Environment

Drivers derive utility from consumption c and disutility from hours worked h according to a utility function $U(c, h)$. They face a budget constraint defined by $c \leq w \cdot h + I$, where w is the wage rate and I is an exogenous, additional source of income. For convenience, I omit the latter variable from notation below since it is not important for what follows. Labor supply is suboptimal because drivers suffer from biases, which leads to

[24]Though the theory is adaptable to compare welfare under different types of behavior since it does not rely on any application of the envelope theorem.

[25]This generalization is presented in the appendix D, alongside the proof of the BWE.

two decision rules for consumption and hours, respectively,

$$\{\tilde{c}(w), \tilde{h}(w)\} \notin \underset{\{c,h\}\in\mathbb{B}(w)}{\arg\max}\; U(c,h),$$

where $\mathbb{B}(\cdot)$ is the choice set defined by the budget constraint inequality and non-negativity constraints on c and h. If drivers did behave optimally, then they would follow two optimal rules for consumption and hours, respectively,

$$\{c^*(w), h^*(w)\} \in \underset{\{c,h\}\in\mathbb{B}(w)}{\arg\max}\; U(c,h).$$

I am interested in a money-metric measure of the change in utility due to a change in labor supply from biased to optimal, namely,

$$\Delta(w) = \frac{U\big(c^*(w), h^*(w)\big) - U\big(\tilde{c}(w), \tilde{h}(w)\big)}{U_c\big(\tilde{c}(w), \tilde{h}(w)\big)}. \tag{10}$$

This quantity varies with the wage rate, which varies between days. Therefore, I treat the wage rate as a random variable so that the object of interest is the expected change in utility from a move to optimal behavior for one shift,

$$\mathbb{E}_w\left[\Delta(w)\right], \tag{11}$$

where $\mathbb{E}_w[\cdot]$ is the expectations operator that integrates with respect to the wage random variable. Higher order moments, such as the variance, can also be calculated and may be of interest; if a driver considers an investment to reduce the biases they suffer, they may also care about the variance of the return of that investment.

5.2 Sufficient Statistics Formula

While equation (10) is the precise quantity of interest, it provides little insight or practical use. In the theorem below, I derive an approximation of this expression that is estimable with a small number of sufficient statistics, which can be used to back out the necessary parameters. Moreover, the sufficient statistics are familiar labor supply elasticities that can be taken "off the shelf" if needed.

Theorem 1 (Behavioral Welfare Expression - BWE) *If the utility function is additively separable in consumption and hours, $\Delta(w)$ can be approximated by a second order Taylor series approximation and difference quotients in order to yield,*

$$\Delta(w) \approx \Delta c - \frac{1}{2}\cdot\eta(w)\cdot\frac{\Delta c^2}{\tilde{c}} - MRS\cdot\Delta h + \frac{1}{2}\cdot\Delta MRS\cdot\Delta h, \tag{12}$$

where,

$$MRS = \frac{v'(\tilde{h}(w))}{u'(\tilde{c}(w))}, \qquad \Delta h = h^*(w) - \tilde{h}(w),$$

$$\Delta MRS = -\frac{MRS}{\tilde{h}(w)/\gamma(w)} \cdot \Delta h, \quad \gamma(w) = \frac{\tilde{h}(w) \cdot v''(\tilde{h}(w))}{v'\left(\tilde{h}(w)\right)},$$

$$\eta(w) = -\frac{\tilde{c}(w) \cdot u''\left(\tilde{c}(w)\right)}{u'(\tilde{c}(w))}, \qquad \Delta c = c^*(w) - \tilde{c}(w).$$

See appendix D for proof.[26]

When consumption is written as earned income, the approximation is made up of three intuitive components,

$$\Delta(w) \approx \underbrace{(w - \mathrm{MRS}) \cdot \Delta h}_{\text{Wedge}} + \underbrace{\frac{1}{2} \cdot \Delta(w - \mathrm{MRS}) \cdot \Delta h}_{\text{Change in wedge}} - \underbrace{\frac{1}{2} \cdot \frac{\eta(w)}{\tilde{c}} \cdot (w \cdot \Delta h)^2}_{\text{Income effect}},$$

which highlight the source of the welfare loss: a wedge in the FOC adjusted for diminishing marginal utility. This is easily explained with an example. Take a hypothetical day where the wage rate was high but the driver worked too little to equate the wage rate and their MRS between labor and consumption. Then, the *Wedge* term captures foregone income corrected for the cost of earning this income. The *Change in wedge* term corrects the Wedge term for the fact that the cost of earning income increases with hours already worked, and the *Income effect* corrects the Wedge term again for diminishing marginal utility. Conversely, if the wage was low and the driver worked too much, the formula captures foregone utility that stems from working for too low a wage with the analogous corrections.

In order to think about how this plays out in expectation, it is useful to consider a simple labor supply heuristic, such as perfect income targeting. Under this heuristic, the driver always works until they earn a set income target so that the elasticity of hours with respect to wages is -1. If we ignore the second order terms in equation (12) and take expectations, this yields,

$$\mathbb{E}_w \left[(w - \mathrm{MRS}) \cdot \Delta h)\right] = \mathbb{E}_w \left[(w - \mathrm{MRS})\right] \cdot \mathbb{E}_w \left[\Delta h\right] + \mathbb{C}_w \left[(w - \mathrm{MRS}), \Delta h)\right],$$

where $\mathbb{C}_w[\cdot]$ is the covariance operator integrating with respect to the wage random variable. For simplicity, assume that the income target is set such that $\mathbb{E}_w \left[(w - \mathrm{MRS})\right] \approx 0$, then the welfare consequences of this labor supply bias depend only on $\mathbb{C}_w \left[(w - \mathrm{MRS}), \Delta h)\right]$, which is positive. Why? When the wage is high, drivers reach their income target quickly and work too little relative to the optimum—Δh is positive and the wedge between the wage and MRS is large. While when the wage is low, the driver who follows an income targeting labor supply bias works too much, so that Δh is negative, and the MRS exceeds the wage so that the wedge is negative. This highlights the inefficiency of income targeting more generally; it causes a negative covariance between hours worked and the wedge in the intratemporal optimality condition.

The accuracy of the approximation embodied in the BWE depends on the size of higher order derivatives of the utility function—if they are small, the approximation is good. In practical terms, the second order

[26] I leave the curvature objects $\eta(\bullet)$ and $\gamma(\bullet)$ as explicit functions of the wage, while suppressing this notation for other terms, in order to highlight a further approximation I use in the empirical implementation. Further, this notation also emphasises that I assume consumption is constant in both the biased or optimal paradigm.

approximation corrects for changes in the cost of labor, as measured by the MRS, in a linear way. I find that the MRS is close to linear in hours and so third order and higher terms are not of concern. In situations where the cost of labor (or benefit of a good) is more convex this may not be the case.

The BWE succinctly captures drivers' welfare losses due to suboptimal behavior, and is amenable to estimation. I lay out and implement a roadmap to estimation in section 6.

6 Welfare Analysis

The empirical evidence from section 3 and the theory from section 5 can be brought together in order to conduct a welfare analysis of drivers' labor supply decisions. Here, I discuss an approximation and an assumption to facilitate this, before outlining and implementing a roadmap to the estimation of welfare of losses due to labor supply biases. I conclude the section by presenting and discussing the results.

6.1 An Approximation and an Assumption

The BWE contains several objects that are related to driver preferences. In this subsection, I explain the steps necessary to identify these with familiar labor supply elasticities.

The approximation. To reduce the burden of estimation I impose that $\eta(w)$ and $\gamma(w)$ are constant rather than functions of the wage rate. I note that the BWE is linear in these terms, and that the MRS is also linear in these terms up to a first order approximation around some level of hours. Therefore, this imposition yields a further approximation of $\Delta(w)$. It also has clear implications for the shape of the utility function; approximating $\Delta(w)$ with $\eta(w)$ and $\gamma(w)$ as constants is equivalent to assuming utility is constant relative risk aversion (CRRA)—*conditional on using the BWE*—because,

$$\{\eta(w), \gamma(w)\} = \{\eta, \gamma\} \iff U(c,h) = \frac{c^{1-\eta}}{1-\eta} - \theta \cdot \frac{h^{1+\gamma}}{1+\gamma}.$$

The CRRA equivalence offers two convenient corollaries. First, the shape of the MRS is known without any further loss of generality,

$$\text{MRS} = \theta \cdot c^{\eta} \cdot h^{\gamma}, \tag{13}$$

and so optimal labor supply is implied by $w = \text{MRS}$. Second, there is a simple mapping from the True Frisch elasticity and Marhsallian elasticity estimates to the curvature parameters,

$$\varepsilon^{\text{Frisch}} = \frac{1}{\gamma}, \tag{14}$$

$$\varepsilon^{\text{Marsh.}} = \frac{1-\eta}{\gamma+\eta}, \tag{15}$$

where the latter equality assumes that drivers have no source of capital income.

The assumption. A Frisch elasticity that is estimated with data from always optimizing drivers maps one-to-one with the curvature of the disutility of labor γ according to equation (14). However, the True Frisch elasticity implicitly compares behavior when optimizing under Tube strikes with biased behavior otherwise, and the associated wage differences. To see this clearly, consider that the Tube strikes instrument is a single binary instrument and that this instrument, hours, and wages have been residualized with the respect to all controls. Then I could estimate my True Frisch elasticity with a Wald estimator,

$$\hat{\beta}^{\text{Wald}} = \frac{\bar{h}_1 - \bar{h}_0}{\bar{w}_1 - \bar{w}_0},$$

where the bar notation $\bar{\bullet}$ indicates the empirical average, and the subscripts $\{0, 1\}$ denote whether the average refers to outcomes during or not during Tube strikes, respectively. If $h_{i,t}$ is always defined by $\log(w_{i,t}) = \log(\theta) + \gamma \cdot \log(h_{i,t})$ it is simple to show that $\hat{\beta}^{\text{Wald}} = 1/\gamma$, where I have omitted the influence of consumption for simplicity. However, when labor supply is biased outside of Tube strikes shift length does not necessarily satisfy the first order condition and so it is not clear that anything about preferences is revealed. But, if biases satisfy,

$$\mathbb{E}\left[\log(w_{i,t})\right] = \log(\theta) + \gamma \cdot \mathbb{E}\left[\log(h_{i,t})\right]. \tag{16}$$

then γ is still identified. Therefore, I require biased labor supply to obey the equality in equation (16): the logarithm of drivers' first order conditions holds in expectation. Note that this is a restriction on the average level of hours, but does not constrain the shape of the biased labor supply function. Beyond its convenience, this condition is attractive for three reasons. Firstly, it grants drivers a degree of sophistication and so does not mechanically overstate the welfare cost of behavioral biases. Secondly, the condition would hold if drivers aim to ensure their intratemporal condition holds in expectation despite their biases, but neglect Jensen's inequality. Lastly, the condition would hold without this neglect if the MRS is linear in hours, which is roughly the case with my results. Given (16), the curvature of consumption utility can also be backed out of equation (15) with the Marshallian elasticity, and no other condition on the nature of biases is necessary.

6.2 Estimation

The estimation of the BWE in expectation requires a number of ingredients; table 9 lists these components alongside how they are estimated and a graphical, mathematical, or numerical representation.

First, I estimate the distribution of wages with a kernel density estimator which is run on the wages observed in my sample. This gives a clear interpretation to my results: the expected welfare loss due to suboptimal labor supply, if one were draw a random shift from the sample.[27] Labor supply under biases is specified as the function implied by the control function model, which is estimated with the mean job distance instrument. Optimal labor supply is characterized by the intratemporal optimality condition, which equates the wage and the MRS. The MRS is specified as in equation (13), which requires the estimation of γ and $\theta \cdot \tilde{c}^{\eta}$, where $\tilde{c}$ is the constant consumption level under baised labor supply, and η is also required separately for the income effect. The curvature parameters η and γ are backed out from the True Frisch elasticity and Marshallian elasticity estimates using equations (14) and (15); these estimates are presented in table 10, where standard errors are calculated with the delta method.

[27] Instead, if one were to take the distribution of wages a particular driver receives, then the interpretation would be the expected welfare loss on any given shift for that driver.

Table 9: Ingredients for BWE in Expectation

Function/Parameter	Estimation	Reference
$f_w(\bullet)$	Kernel density estimator	Figure 10
$\tilde{h}(w)$	Control function model	Figure 4
$h^*(w)$	Intratemporal optimality condition	Figure 9
$\eta(w) = \eta$	True Frisch + Marshallian + eq. (15)	Table 10, row one
$\gamma(w) = \gamma$	True Frisch + eq. (14)	Table 10, row two
$\theta \cdot \tilde{c}^{\eta}$	Hour levels during Tube strikes	$\mathbb{E}_{w\vert\text{strikes}}\left[\frac{w}{h^{\gamma}}\right]$
	Calibrated to income level	$\mathbb{E}_w\left[w \cdot \tilde{h}(w)\right] = \mathbb{E}_w\left[w \cdot h^*(w)\right]$
$\tilde{c}$	Average shift income	$\mathbb{E}_w\left[w \cdot \tilde{h}(w)\right]$
Δc	Change in average shift income	$\mathbb{E}_w\left[w \cdot \left(\tilde{h}(w) - h^*(w)\right)\right]$
ν	% income + hours targeters in survey	≈0.66

Notes: This table presents the different objects required to calculate the expectation of the BWE, how these objects are estimated, and where they are illustrated in the paper.

Table 10: Curvature Parameter Estimates

Parameter	Estimate	Std. error
η	1.37	0.02
γ	1.25	0.03

Notes: This table presents estimates for the parameters η and γ. The estimates are produced using equations 15 and 14, and the coefficients from column (1) of table A2 and column (2) of table 5, respectively. Standard errors are constructed using the delta method.

I consider two approaches to parameterize $\theta \cdot \tilde{c}^{\eta}$, which relates the optimal level of hours supplied. Firstly, I set it to match the observed level of hours during Tube strikes when drivers are hypothesized to be behaving optimally. Given this hypothesis, there are a number of ways to derive the level parameter, which make different implicit assumptions about the nature of random (not behavioral) errors in driver behavior. This is discussed in appendix E, and I use the specification in row 6 of table 9. Secondly, I calibrate the level of optimal hours to ensure the daily level of income is unchanged when behavior moves from bias to optimal. Both of these approaches assume constant consumption, either when behaving with biases or optimally since the exercise considers daily fluctuation in wages, which drivers are able to insure against. Naturally, this has ramifications for the estimation of $\tilde{c}$. I assume that constant consumption under biased behavior equals average daily income that is generated by $\tilde{h}(w)$ and the distribution of wages $f_w(\bullet)$. This parameter only mediates the strength of the income effect in the BWE and so it has no effect on the results when $\theta \cdot \tilde{c}^{\eta}$ is calibrated to maintain constant consumption. The change in consumption between the optimal and behavioral regime is set as the difference in expected income generated by the two labor supply schedules.

It is plausible that only a proportion of drivers ν are responsible for the deviation from optimal behavior. That is,

$$\tilde{h}(w) = \nu \cdot \hat{h}(w) + (1 - \nu) \cdot h^*(w),$$

where $\hat{h}(w)$ is the labor supply of the purely biased drivers, such that $\tilde{h}(w)$ is now viewed as the confluence of optimal and suboptimal labor supply by drivers. Given ν, it is easy to infer $\hat{h}(w)$, however, ν is unknown. The smaller ν is, the more severe biased behavior deviates from the optimal. I calibrate ν to equal the proportion of income and hours targeters in the survey presented in table 8. This calibration is also appealing because if biases were concentrated amongst an even smaller group of drivers, the most negative elasticity of $\hat{h}(w)$ would fall below -1, which is hard to reconcile with any behavioral theories. Note that two welfare loss estimates can be derived from this exercise: one for the biased individuals, and an average for all individuals, *i.e.* the former estimate scaled by ν.

With two alternative calibrations of $\theta \cdot \tilde{c}^{\eta}$ and homogeneous and heterogeneous behavior, I present four results for the expected daily welfare loss due to labor supply biases. I also present a further fifth result, where I strictly assume CRRA preferences in combination with homogeneous behavior and an optimal level of hours during Tube strikes, in order to check that the BWE yields plausible results.

6.3 Results

The results of the estimation are summarized in table 11. The cell in the first row and column considers that all drivers are equally biased, and derives the level parameter from the hours observed during Tube strikes. This reveals an expected daily welfare loss of £2.32. Notably, optimal behavior implies a significant loss in consumption utility of £9.01 (approximately 10 percent of average daily income), which is exceeded by savings in disutility from work. The overall welfare loss only falls slightly, to £2.09, when the level parameter is calibrated to keep expected daily income, and thus consumption, constant. Both these numbers are relatively close to £2.44, which is the loss implied when CRRA preferences are assumed without any approximation.

The expected daily welfare loss rises sharply when a fraction of drivers behave optimally. If two thirds of drivers are responsible for biased behavior then the welfare loss more than doubles. If the level parameter is set to match hours during Tube strikes, the expected loss equals £5.29 for biased drivers. This result must be scaled by $1/\nu$ in order to determine the unconditional, expected welfare loss, which leads to a loss of £3.49. Again, this loss is composed of a consumption loss and an overwhelming gain due to more efficient labor supply. Moreover, this welfare loss does not change markedly when the level parameter is calibrated to keep expected income constant. In this scenario, the loss for biased individuals equals £4.72 and is £3.11 across all drivers.

Discussion. In order to gain a sense of the economic importance of these losses, it is necessary to account for how regularly they are incurred. Driver-weeks most commonly contain five shifts and the mean number of shifts in a week is four. Annually, this regularity of work combined with holidays leads to an average of 200 shifts. Therefore, given preferences that are additively separable across time and negligible discounting over the course of a year, the results imply losses ranging from £418.00 to £1058.00 per annum. For context,

Table 11: Expected Daily Welfare Losses for Biased Drivers

<table>
<tr><td></td><td></td><td colspan="2">$\theta \cdot \tilde{c}^{\eta}$</td></tr>
<tr><td></td><td></td><td>Tube strike level</td><td>Expected income level</td></tr>
<tr><td rowspan="2">ν</td><td>Homog.</td><td>2.32
(p5, p95)
$\Delta u(c)$: -9.01 $\Delta u(h)$: +11.32</td><td>2.09
(p5, p95)</td></tr>
<tr><td>Heterog.</td><td>5.29
(p5, p95)
$\Delta u(c)$: -13.78 $\Delta u(h)$: +19.07</td><td>4.72
(p5, p95)</td></tr>
</table>

Notes: This table presents the expected welfare losses in pounds (£) for the 2×2 variations under considerations. When the level parameter is not calibrated to keep expected income equal under biased and optimal behavior, I report how the welfare loss is composed of changes in utility due to consumption and labor supply behavior. 5^{th} and 95^{th} percentiles from 500 bootstraps are presented in the parentheses. The heterogeneity scenario considers $\nu = 0.66$.

Handel and Kolstad (2015) find average losses due to suboptimal health insurance choices, which are made annually, of approximately £1,200.00.[28] Such losses are considered large and are of a comparable magnitude to the results presented here. Accordingly, the losses in this paper point towards significant welfare losses for drivers due to behavioral frictions in exploiting flexibility.

The results also have implications for the role of heterogeneity in the cost of biases and the nature of biases. Firstly, aggregate welfare losses increase even as biases are concentrated amongst a smaller proportion of drivers. This is because welfare losses for biased drivers increase quickly as they become more responsible for the aggregate deviation from optimal behavior, such that aggregate welfare losses increase locally as a smaller proportion of drivers are biased. More precisely, the elasticity of $\mathbb{E}_w\left[\Delta(w)\right]$ with respect to ν is greater than one. It is important to recognize that this non-unitary elasticity arises because drivers make different decisions, which imply different size internalities, for different wages, and they have to make these decisions repeatedly. This mechanism contributes another channel to those identified by Taubinsky and Rees-Jones (2018) as to how heterogeneity in biases can accentuate welfare losses.

Secondly, the results suggest that suboptimal labor supply is not likely to be driven by time-inconsistent decision making, but they are in line with a reference-dependent labor supply hypothesis. Column 1 of table 11 implies drivers spurn a saving in the expenditure of effort—an immediate benefit—at the price of lower income—a delayed cost—which is opposite to the expected behavior of an individual with self-control issues. Moreover, if a driver were to deviate to the optimal labor supply schedule for just one day, the cost of lower income would be negligible since consumption can be smoothed with (dis)saving, which reinforces the argument that self-control is not immediately at play. However, it is plausible that sophisticated drivers with self-control issues may use goal-setting, which is consistent with the results here. Indeed, figure 4 resembles a labor supply schedule for drivers with reference dependence,[29] which has been used to motivate goal-setting amongst time inconsistent individuals (Hsiaw, 2013). Recent work by Reck and Seibold (2020)

[28] Handel and Kolstad (2015) model information frictions and hassle costs as the source of suboptimal choice.
[29] See figure 1 in DellaVigna (2009).

explains how decreases in the reference point generally improve welfare because they reduce over consumption, or in this case over work, that takes place to reach a reference point. This is what I observe; drivers work too much at low wages and this is the main cause of welfare losses.

These lessons suggest some remedies for the welfare losses experienced by drivers. Making wages more salient may help purge drivers of their biases in a way similar to Tube strikes. Further, drivers should be trained to avoid using income targets since this causes a negative correlation between the wedge in their intratemporal condition and their labor supply. Hours targets are less harmful, if targets are necessary as a way to avoid self-control issues. Lastly, if over working for low wages due to reference-dependent labor supply is a broader feature of self-employment, this could form part of the argument in favor of a minimum wage for the self-employed in some contexts.

7 Conclusion

The approach in this paper builds on previous work that has established non-neoclassical labor supply behavior in many settings, especially amongst New York taxi drivers, in order to ask whether such observations indicate a deviation from optimal behavior and are important for welfare; the answer is yes to both. I characterize preferences and behavior for a group of self-employed workers using salient and common wage variation, respectively, which resolves any normative ambiguity. Typical labor supply is non-monotonic in the wage rate and, for some wage rates, the elasticity of labor supply is more negative than an estimate of the Marshallian elasticity, which indicates labor supply is generally biased. Conversely, salient wage variation due to Tube strikes causes a large increase in hours worked that is consistent with Frisch elasticity estimates from the literature. I derive and estimate a theoretical expression to quantify the welfare losses due to behavioral biases, where responses to salient wage variation characterize preferences. This exercise reveals a significant welfare loss for these individuals over the course of a working year. The size of the welfare loss rises steeply as a smaller proportion of drivers are assumed to be biased, and the estimated labor supply function is suggestive of reference-dependence.

Given these findings, it is important to pin down more precisely what biases are afflicting drivers so that it is possible to facilitate choice making which is in line with individuals desires. I see this as a valuable area of future research, which requires experimentation to understand workers' incentives and surveys to understand their motivations. However, even without this knowledge, there are actionable points now. This paper suggests salience is important in affecting optimization, so organizations that contract with self-employed workers could help make predicted wages salient. Moreover, labor supply "training" could inform self-employed workers of the acute inefficiencies due to heuristics such as income targeting. However, it is not clear whether it is in the interest of these organizations to debias drivers. In this regard, this research could provide part of the rationale for a minimum wage for the self-employed with the aim to prevent individuals working for too long at low wages.

Further, this paper has abstracted from extensive margin labor supply decisions because drivers in this setting tend to work a full working week. But in many settings, such as Uber drivers, extensive margin labor supply decisions are much more important and have been used as a key ingredient to determine the value of flexible work. Therefore, better understanding biases at the extensive margin level, and their interaction with intensive margin biases, is an important area of future research.

References

Agarwal, Sumit, Mio Diao, Jessica Pan, and Tien Foo Sing, "Labor Supply Decisions of Singaporean Cab Drivers," 2013.

Allcott, Hunt and Dmitry Taubinsky, "Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market," *American Economic Review*, 2015, *105* (8), 2501–38.

Andersen, Steffen, Alec Brandon, Uri Gneezy, and John A List, "Toward an Understanding of Reference-Dependent Labor Supply: Theory and Evidence from a Field Experiment," Working Paper 20695, National Bureau of Economic Research November 2014.

Ashenfelter, Orley, Kirk Doran, and Bruce Schaller, "A Shred of Credible Evidence on the Long-run Elasticity of Labour Supply," *Economica*, October 2010, *77* (308), 637–650.

Berkouwer, Susanna B and Joshua T Dean, "Credit and attention in the adoption of profitable energy efficient technologies in Kenya," 2019.

Bernheim, B Douglas and Antonio Rangel, "Addiction and cue-triggered decision processes," *American economic review*, 2004, *94* (5), 1558–1590.

__ and __ , "Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics," *The Quarterly Journal of Economics*, 2009, *124* (1), 51–104.

__ and Dmitry Taubinsky, "Behavioral public economics," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 1, Elsevier, 2018, pp. 381–516.

Blundell, Richard and Thomas MaCurdy, "Labor supply: A review of alternative approaches," in "Handbook of labor economics," Vol. 3, Elsevier, 1999, pp. 1559–1695.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Salience and consumer choice," *Journal of Political Economy*, 2013, *121* (5), 803–843.

Borjas, George, "The Relationship between Wages and Weekly Hours of Work: The Role of Division Bias," *Journal of Human Resources*, 1980, *15* (3), 409–423.

Caldwell, Sydnee and Emily Oehlsen, "Monopsony and the gender wage gap: Experimental evidence from the gig economy," *Massachusetts Institute of Technology Working Paper*, 2018.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler, "Labor Supply of New York City Cabdrivers: One Day at a Time*," *The Quarterly Journal of Economics*, 05 1997, *112* (2), 407–441.

Caplin, Andrew and Mark Dean, "Revealed preference, rational inattention, and costly information acquisition," *American Economic Review*, 2015, *105* (7), 2183–2203.

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

Chen, Kuan-Ming, Claire Ding, John A List, and Magne Mogstad, "Reservation Wages and Workers' Valuation of Job Flexibility: Evidence from a Natural Field Experiment," Technical Report, National Bureau of Economic Research 2020.

Chen, M. Keith, Judith A. Chevalier, Peter E. Rossi, and Emily Oehlsen, "The Value of Flexible Work: Evidence from Uber Drivers," NBER Working Papers 23296, National Bureau of Economic Research, Inc March 2017.

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and taxation: Theory and evidence," *American economic review*, 2009, *99* (4), 1145–77.

Chou, Y.K., "Testing Alternative Models of Labor Supply. Evidence from Taxi-Drivers in Singapore," Technical Report 2000.

Collins, Brett, Andrew Garin, Emile Jackson, Dmitri Koustas, and Mark Payne, "Is Gig Work Replacing Traditional Employment? Evidence from Two Decades of Tax Returns," 2019.

Crawford, Vincent P. and Juanjuan Meng, "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, August 2011, *101* (5), 1912–32.

DellaVigna, Stefano, "Psychology and economics: Evidence from the field," *Journal of Economic literature*, 2009, *47* (2), 315–72.

__ and Joshua M Pollet, "Investor inattention and Friday earnings announcements," *The Journal of Finance*, 2009, *64* (2), 709–749.

Doran, Kirk, "Are long-term wage elasticities of labor supply more negative than short-term ones?," *Economics Letters*, 2014, *122* (2), 208–210.

Dupas, Pascaline, Jonathan Robinson, and Santiago Saavedra, "The Daily Grind : Cash Needs , Labor Supply and Self-Control," 2015.

Farber, Henry S., "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers," *Journal of Political Economy*, February 2005, *113* (1), 46–82.

__ , "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers," *American Economic Review*, June 2008, *98* (3), 1069–82.

__ , "Why you Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers*," *The Quarterly Journal of Economics*, 07 2015, *130* (4), 1975–2026.

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

Gabaix, Xavier, "A sparsity-based model of bounded rationality," *The Quarterly Journal of Economics*, 2014, *129* (4), 1661–1710.

__ and David Laibson, "Shrouded attributes, consumer myopia, and information suppression in competitive markets," *The Quarterly Journal of Economics*, 2006, *121* (2), 505–540.

__ , __ , Guillermo Moloche, and Stephen Weinberg, "Costly information acquisition: Experimental analysis of a boundedly rational model," *American Economic Review*, 2006, *96* (4), 1043–1068.

Goldin, Jacob and Daniel Reck, "Optimal defaults with normative ambiguity," *Available at SSRN 2893302*, 2017.

Gruber, Jonathan and Botond Köszegi, "Is addiction "rational"? Theory and evidence," *The Quarterly Journal of Economics*, 2001, *116* (4), 1261–1303.

__ and Botond Kőszegi, "Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes," *Journal of Public Economics*, 2004, *88* (9-10), 1959–1987.

Haggag, Kareem and Giovanni Paci, "Default tips," *American Economic Journal: Applied Economics*, 2014, *6* (3), 1–19.

Handel, Benjamin R and Jonathan T Kolstad, "Health insurance for" humans": Information frictions, plan choice, and consumer welfare," *American Economic Review*, 2015, *105* (8), 2449–2500.

Hsiaw, Alice, "Goal-setting and self-control," *Journal of Economic Theory*, 2013, *148* (2), 601–626.

Keane, Michael P, "Labor supply and taxes: A survey," *Journal of Economic Literature*, 2011, *49* (4), 961–1075.

Mas, Alexandre and Amanda Pallais, "Valuing alternative work arrangements," *American Economic Review*, 2017, *107* (12), 3722–59.

Morgul, Ender Faruk and Kaan Ozbay, "Revisiting Labor Supply of New York City Taxi Drivers: Empirical Evidence from Large-Scale Taxi Data," Technical Report 2015.

Mullainathan, Sendhil, Joshua Schwartzstein, and William J Congdon, "A reduced-form approach to behavioral public finance," 2011.

Nguyen, Quang and PinngSun Leung, "Revenue Targeting in Fisheries: The Case of Hawaii Longline Fishery," 2009.

Reck, Daniel and Arthur Seibold, "The Welfare Economics of Reference Dependence," *Working Paper*, 2020.

Reichling, Felix and Charles Whalen, "Review of estimates of the Frisch elasticity of labor supply," 2012.

Sallee, James M, "Rational inattention and energy efficiency," *The Journal of Law and Economics*, 2014, *57* (3), 781–820.

Schmidt, Marc-Antoine, "The Daily Labor Supply Response to Worker-Specific Earnings Shocks," Technical Report, Working Paper 2018.

Sims, Christopher A, "Implications of rational inattention," *Journal of monetary Economics*, 2003, *50* (3), 665–690.

Spinnewijn, Johannes, "Unemployed but optimistic: Optimal insurance design with biased beliefs," *Journal of the European Economic Association*, 2015, *13* (1), 130–167.

Taubinsky, Dmitry and Alex Rees-Jones, "Attention variation and welfare: theory and evidence from a tax salience experiment," *The Review of Economic Studies*, 2018, *85* (4), 2462–2496.

Thakral, Neil and Linh T. Tô, "Daily Labor Supply and Adaptive Reference Points," *Working Paper*, 2017.

Thaler, Richard H and Cass R Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*, Penguin, 2009.

Wooldridge, Jeffrey M, "Control function methods in applied econometrics," *Journal of Human Resources*, 2015, *50* (2), 420–445.

A Data Cleaning

New: completed time seems the most reliable measure, so job times are computed by deducting journey time or, if that is not possible, difference with minimum pick up arrival.

At the ride level, I drop observations if,

- The ride was cancelled,
- The start time of the ride is not observed,
- The fare or total transaction value is not observed,
- The ride distance is not observed,
- The duration of the ride is not observed.

Then I trim the data if rides fall above the 2.5 percentile according to the following variables,

- Speed,
- Driving wage,
- Journey distance,
- Ride duration,
- Fare and total transaction value,
- Waiting time,
- Driver extras, *e.g.*, additional fees for toll gates.

I also drop rides if,

- The average speed was below 5 kilometres per hour,
- The driving wage fell below £3.00,
- The journey time was less than one minute,
- The fare or total transaction was less than £1.00.

From these rides I construct shifts by allocating rides that are within six hours of one another to the same shift. At the shift level, I ensure that shifts,

- Contain at least three jobs,
- Do not have more than five jobs an hour,
- Are not shorter than two hours and not longer than 18 hours.

Over the course of this project, I have tried variations on these restrictions and none have significantly impacted the results.

B Marshallian Labor Supply Elasticity

I estimate a Marshallian labor supply elasticity using a permanent fare reform that affected drivers' wage rates. The permanent nature of this reform implies that the resulting elasticity incorporates income effects. In the absence of any cross-sectional variation in the application of the reform, identification rests on an exogenous change in the wage that is not confounded by other factors which affect labor supply.

The fare reform took place on June 6$^{\text{th}}$ 2016 and its effect on the wage rate is illustrated in figure 11. After residualizing the wage with respect to controls, as shown in figure 12, it is clear the wage rose by approximately £1.75, from £14.00 to £15.75. Table A1 estimates this as a percentage increase of around 10 percent.

Figure 13 demonstrates how shift hours evolve over the same time horizon. A drop in hours is apparent but, like the wage rate, trends over time complicate inference. Residualizing shift length on controls, as in figure 14, reveals a clear though small drop in shift length of around five minutes.

The results from the formal estimation of the Marshallian elasticity are shown in table A2. In this analysis, I regress shift length on the wage rate and controls, and instrument for the wage rate with the fare reform. Controls are analogous, where applicable, to the Tube strike analysis; they include dummies for Tube strikes, factor variables for when the shift was started, a dummy for Ramadan, dummies for bank holidays, and time controls. Given the significant role of trends over time, I test robustness to different specifications of the the time trend polynomial and the length of the window on which the estimation is run. The resulting estimates range from zero to slightly negative—as in Ashenfelter et al. (2010). My preferred estimate, which falls in the midrange of the estimates, is a Marshallian elasticity of -0.142 (s.e., 0.04). When a smaller window is considered, an elasticity much closer to zero is estimated, which is not surprising given the declining trajectory of shift length seen in the raw data. Adjusting the polynomial order of the time trend leads larger and smaller Marshallian elasticities, but none of these alternative estimates are statistically different.

C Survey Details

Table 8 presents results from a survey conducted on a separate group of self-employed workers who are predominantly based in London. These individuals work for a variety of ridesharing and food delivery platforms, and the survey was conducted via a firm which provides hire and reward vehicle insurance to these individuals. In order to incentivize completion of the survey, individuals were entered into a lottery for a £50.00 Amazon voucher upon completion. The survey contained 12 questions, including the question presented here, on general working patterns, wages, and costs. Participants were invited to select one of the responses displayed in table 8 to the question "How do you decide the number of hours to work in a day?". The ordering of the possible responses was randomized. For the temporary versus permanent wage increase question, drivers were asked "How would you change your daily hours in response to a temporary [permanent] wage increase of 10 percent?". Drivers had the option of reducing their hours, keeping their hours the same, or increasing them. In total, 476 individuals completed this question.

Table A1: True Marshallian Elasticity First Stage

	Dependent variable:						
	log(Shift wage)						
	Full window	3 mo. window	6 mo. window	4th order poly.	5th order poly.	7th order poly.	8th order poly.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Reform	0.114***	0.102***	0.101***	0.105***	0.115***	0.090***	0.082***
	(0.003)	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)
Strike dummies	Yes	No	Yes	Yes	Yes	Yes	Yes
Shift type dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ramadan dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank holiday dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DOTW dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trend poly. order	6th	1st	1st	4th	5th	7th	8th
Observations	994,486	140,958	266,279	994,486	994,486	994,486	994,486
R^2	0.114	0.156	0.172	0.113	0.114	0.114	0.115

*p<0.1; **p<0.05; ***p<0.01

Notes: This table presents first stage results from the influence of a fare reform on drivers wages. The regressions control for driver fixed effects, a factor variables for start time of the shift, Ramadan, bank holidays, days of the week, months, and a time trend of varying flexibility. The regressions either use data from different windows around the reform, or allow for different polynomial time trends. Standard errors are clustered at the driver level.

Table A2: True Marshallian Elasticity Results

	Dependent variable:						
	log(Shift length)						
	Full window	3 mo. window	6 mo. window	4th order poly.	5th order poly.	7th order poly.	8th order poly.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(Shift wage)	-0.142^{***}	0.013	0.076	-0.234^{***}	-0.073^{**}	-0.212^{***}	-0.240^{***}
	(0.040)	(0.063)	(0.061)	(0.038)	(0.036)	(0.053)	(0.059)
Strike dummies	Yes	No	Yes	Yes	Yes	Yes	Yes
Shift type dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ramadan dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank holiday dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DOTW dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trend poly. order	6th	1st	1st	4th	5th	7th	8th
Observations	994,486	140,958	266,279	994,486	994,486	994,486	994,486
R^2	0.092	0.054	0.044	0.097	0.084	0.097	0.098

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Notes: This table presents coefficients on wages, which are instrumented with a permanent fare reform, from a series of regressions of shift length on this variable. The regressions control for driver fixed effects, a factor variables for start time of the shift, Ramadan, bank holidays, days of the week, months, and a time trend of varying flexibility. The regressions either use data from different windows around the reform, or allow for different polynomial time trends. Standard errors are clustered at the driver level.

D Behavioral Welfare Expression Derivation

In this part of the appendix, I prove theorem 1 and derive an analogous expression for changes in consumer surplus caused by price changes.

D.1 Proof of Theorem 1

In order to derive $\Delta(w)$ in terms of sufficient statistics, I make use of the fact that utility is assumed to be additively separable in consumption and hours worked. To start, I will work with consumption utility. I consider a change in consumption induced by switching from a biased consumption rule to an optimal consumption rule at a given wage rate. Mathematically I use a second order Taylor series approximation to show,

$$\begin{aligned} u(c^*(w)) &\approx u\big(\tilde{c}(w)\big) + u'\big(\tilde{c}(w)\big) \cdot \big(c^*(w) - \tilde{c}(w)\big) + \frac{1}{2} \cdot u''\big(\tilde{c}(w)\big) \cdot (c^*(w) - \tilde{c}(w))^2 \\ \iff \frac{u(c^*(w)) - u(\tilde{c}(w))}{u'(\tilde{c}(w))} &\approx c^*(w) - \tilde{c}(w) + \frac{u''(\tilde{c}(w))}{u'(\tilde{c}(w))} \cdot \frac{(c^*(w) - \tilde{c}(w))^2}{2}. \end{aligned}$$

The same operations with hours disutility yield,

$$\frac{v(h^*(w)) - v(\tilde{h}(w))}{u'(\tilde{c}(w))} \approx \frac{v'(\tilde{h}(w)}{u'(\tilde{c}(w))} \cdot (h^*(w) - \tilde{h}(w)) + \frac{v''(\tilde{h}(w))}{u'(\tilde{c}(w))} \cdot \frac{(h^*(w) - \tilde{h}(w))^2}{2}.$$

Combining these terms gives,

$$\begin{aligned} \frac{U\big(c^*(w), h^*(w)\big) - U\big(\tilde{c}(w), \tilde{h}(w)\big)}{u'(\tilde{c}(w))} \approx & c^*(w) - \tilde{c}(w) + \frac{u''(\tilde{c}(w))}{u'(\tilde{c}(w))} \cdot \frac{(c^*(w) - \tilde{c}(w))^2}{2} + \dots \\ & \dots \frac{v'(\tilde{h}(w)}{u'(\tilde{c}(w))} \cdot (h^*(w) - \tilde{h}(w)) + \frac{v''(\tilde{h}(w))}{u'(\tilde{c}(w))} \cdot \frac{(h^*(w) - \tilde{h}(w))^2}{2} \\ \approx & c^*(w) - \tilde{c}(w) + \frac{1}{2} \cdot \frac{\tilde{c}(w) \cdot u''(\tilde{c}(w))}{u'(\tilde{c}(w))} \cdot \frac{(c^*(w) - \tilde{c}(w))^2}{\tilde{c}(w)} + \dots \\ & \dots \frac{v'(\tilde{h}(w)}{u'(\tilde{c}(w))} \cdot (h^*(w) - \tilde{h}(w)) + \frac{1}{2} \cdot \frac{v''(\tilde{h}(w))}{u'(\tilde{c}(w))} \cdot (h^*(w) - \tilde{h}(w))^2 \\ \approx & c^*(w) - \tilde{c}(w) + \frac{1}{2} \cdot \frac{\tilde{c}(w) \cdot u''(\tilde{c}(w))}{u'(\tilde{c}(w))} \cdot \frac{(c^*(w) - \tilde{c}(w))^2}{\tilde{c}(w)} + \dots \\ & \dots \frac{v'(\tilde{h}(w)}{u'(\tilde{c}(w))} \cdot (h^*(w) - \tilde{h}(w)) \dots \\ & \dots + \frac{1}{2} \cdot \frac{\tilde{h}(w) \cdot v''(\tilde{h}(w))}{v'(\tilde{c}(w))} \cdot \frac{v'(\tilde{h}(w))}{u'(\tilde{c}(w))} \cdot \frac{(h^*(w) - \tilde{h}(w))^2}{\tilde{h}(w)}, \end{aligned}$$

which is in the same form as equation (12) without the simplified notation.

D.2 Change in Consumer Surplus Due to Price Change

In this subsection of the appendix, I consider the change in consumer surplus due to a wage change from w' to w but keep the biased policy rules that define behavior constant. Take a second order Taylor approximation of the value function at wage w around the wage w',

$$\begin{aligned}V(w) =&U\big(\tilde{c}(w), \tilde{h}(w)\big) \\ \approx&V(w') + \frac{dV(w')}{dw}(w-w') + \frac{1}{2}\frac{d^2V(w')}{dw^2}(w-w')^2.\end{aligned}$$

It is not possible to apply the Envelope theorem here because of the policy functions do not maximize utility, so,

$$\begin{aligned}\frac{dV(w')}{dw} =&U_c\left[\tilde{h}(w') + \left(w' + \frac{U_h}{U_c}\right)\frac{d\tilde{h}(w')}{dw}\right], \\ \frac{d^2V(w')}{dw^2} =&U_c\left[\frac{d\tilde{h}(w')}{dw} + \frac{d\tilde{h}(w')}{dw} + w'\frac{d^2\tilde{h}(w')}{dw^2}\right] + U_{cc}\left[\tilde{h}(w') + w'\frac{d\tilde{h}(w')}{dw}\right]^2 \dots \\ &\dots + U_h\frac{d^2\tilde{h}(w')}{dw^2} + U_{hh}\left[\frac{d\tilde{h}(w')}{dw}\right]^2.\end{aligned}$$

where I have assumed additively separable utility in consumption and hours. Rearranging the Taylor series approximation and substituting in the above yields:

$$\begin{aligned}\frac{V(w)-V(w')}{U_c} =&h\cdot\Delta w + \frac{1}{2}\cdot\Delta h\cdot\Delta w - \frac{1}{2}\cdot\frac{\eta}{c}\cdot(\Delta hw)^2\dots \\ &\dots + (w-\text{MRS})\cdot\Delta h + \frac{1}{2}\cdot\Delta(w-\text{MRS})\cdot\Delta h,\end{aligned} \tag{17}$$

where $\eta = \frac{-cU_{cc}}{U_c}$, $\gamma = \frac{hU_{hh}}{U_h}$, $\text{MRS} = -\frac{U_h}{U_c}$, $h = \tilde{h}(w')$, $\Delta h = \tilde{h}(w) - \tilde{h}(w')$, $\Delta w = w - w'$, $c = \tilde{c}(w')$, $\Delta hw = \tilde{h}(w)w - \tilde{h}(w')w'$, and $\Delta(w-\text{MRS}) = \Delta w - \gamma\cdot\frac{\text{MRS}}{h}\cdot\Delta h$. Relative to the BWE, this expression incorporates the mechanical change due to the price change as well as changes in the internalities. Further if the consumption and hours policy functions were optimal rules, the wedge between the wage and the MRS would always be zero such that equation (17) would collapse to the first line.

E Parameterization of $\theta\cdot\tilde{c}^{\eta}$

If drivers behave optimally during Tube strikes there are a number of theoretically equivalent ways to identify the level parameter $\theta\cdot c^{\eta}$. However, given the inclusion of econometric errors, which may not be i.i.d., this equivalence breaks down. Therefore, it is important to consider how best to identify the level parameter $\theta\cdot c^{\eta}$. In the below, I consider only observations during Tube strikes and start with the following model,

$$w_{i,t} = (\theta\cdot c^{\eta})\cdot u_{i,t}\cdot h_{i,t}^{\gamma} + v_{i,t} \tag{18}$$

where there are two possible sources of deviation from the intratemporal optimality condition: an additive error $v_{i,t}$ and a multiplicative error $u_{i,t}$, where $\mathbb{E}[v_{i,t}] = 0$ and $\mathbb{E}[u_{i,t}] = 1$. If both $v_{i,t}$ and $u_{i,t}$ were i.i.d., then equation (18) would imply,

$$\mathbb{E}\left[\frac{w_{i,t}}{h_{i,t}^{\gamma}}\right] = \frac{\mathbb{E}[w_{i,t}]}{\mathbb{E}\left[h_{i,t}^{\gamma}\right]},$$

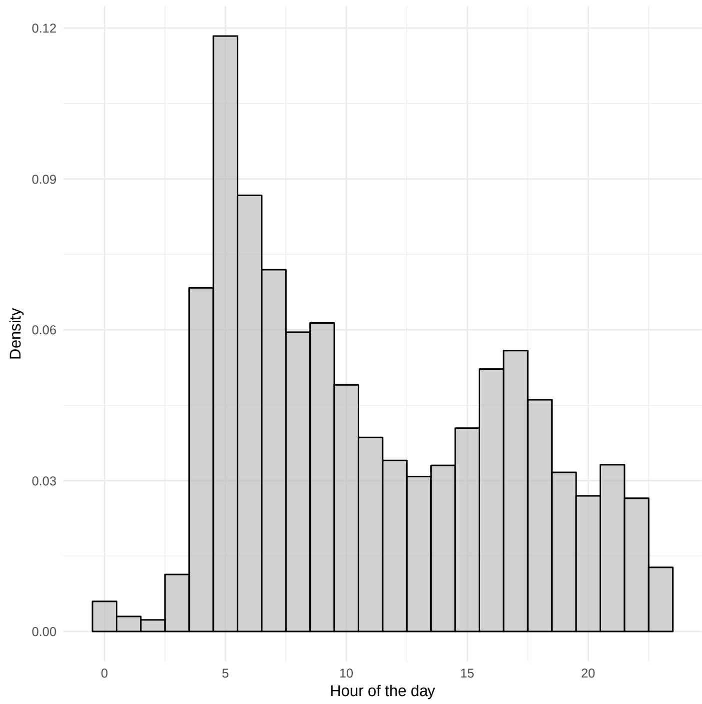
which is *not* the case empirically. Therefore, one or more of $v_{i,t}$ and $u_{i,t}$ is not i.i.d. If neither error term is i.i.d. then estimation is near impossible so, practically, the challenge is to determine which source of error is more severely not i.i.d. I argue that the multiplicative error $u_{i,t}$ term is most likely to be related to hours $h_{i,t}$ since $u_{i,t}$ intuitively captures idiosyncratic variation in consumption and the distribution of hours will be linked to the distribution of income, which is related to consumption c and, in turn, the level parameter $\theta \cdot c^{\eta}$. As a result, I use the following estimator,

$$\mathbb{E}\left[\frac{w_{i,t}}{h_{i,t}^{\gamma}}\right] = (\theta \cdot c^{\eta}) \cdot \underbrace{\mathbb{E}[u_{i,t}]}_{=1} + \underbrace{\mathbb{E}[v_{i,t} \cdot h_{i,t}^{-\gamma}]}_{=\mathbb{E}[v_{i,t}]\cdot\mathbb{E}[h_{i,t}^{-\gamma}]=0} .$$

I note that the estimate generated by this estimator is consistent with the assumption made on biases, ***i.e.***, that the logarithm of drivers' first order condition holds in expectation, though the assumption itself cannot be used to estimate the level parameter $\theta \cdot c^{\eta}$ because of Jensen's inequality.

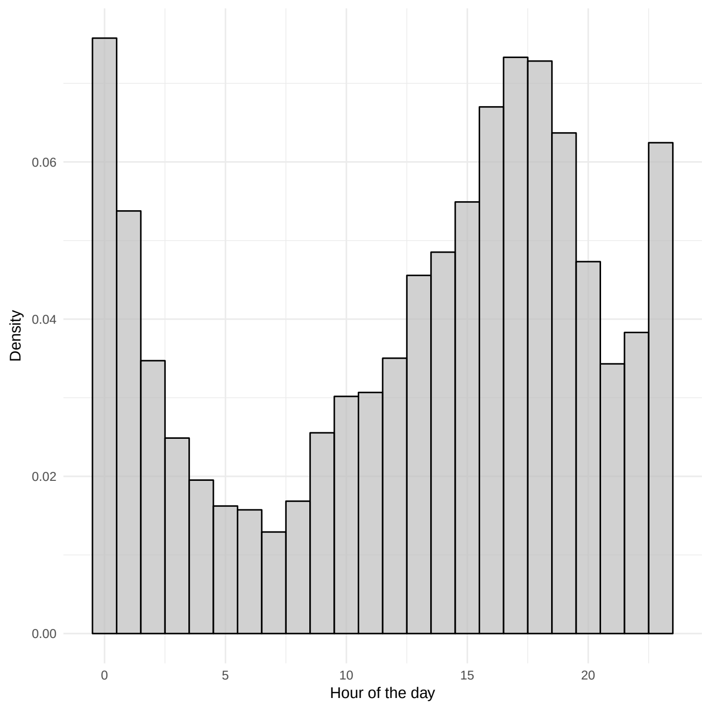
List of Figures

Figure 1: Shift Start Time Density



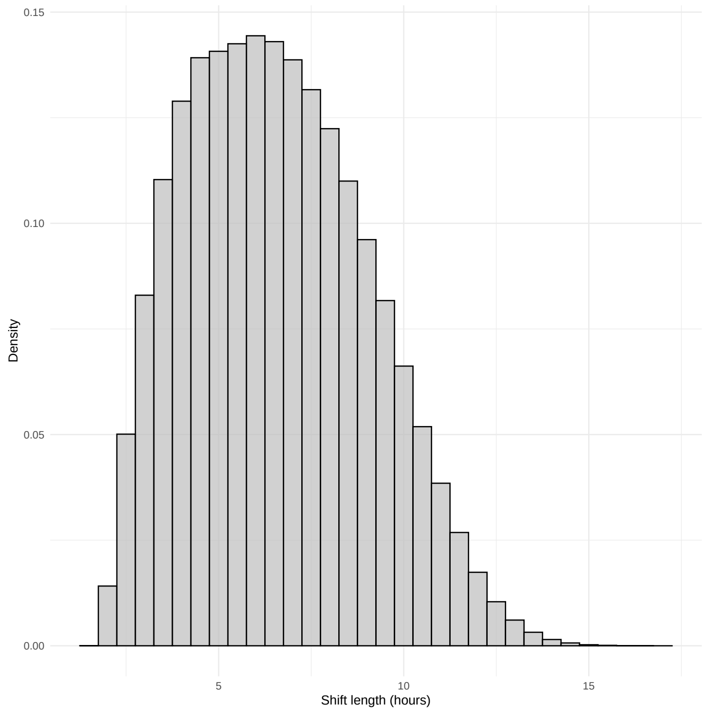
Notes: This figure presents the empirical density histogram of the start times of shifts for the balanced sample, where each bin is an hour of the day.

Figure 2: Shift End Time Density



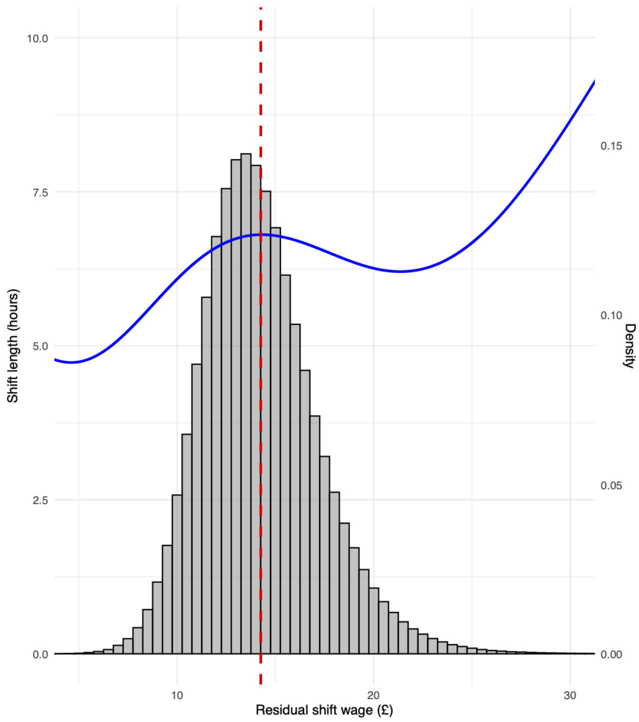
Notes: This figure presents the empirical density histogram of the end times of shifts for the balanced sample, where each bin is an hour of the day.

Figure 3: Shift Length Density



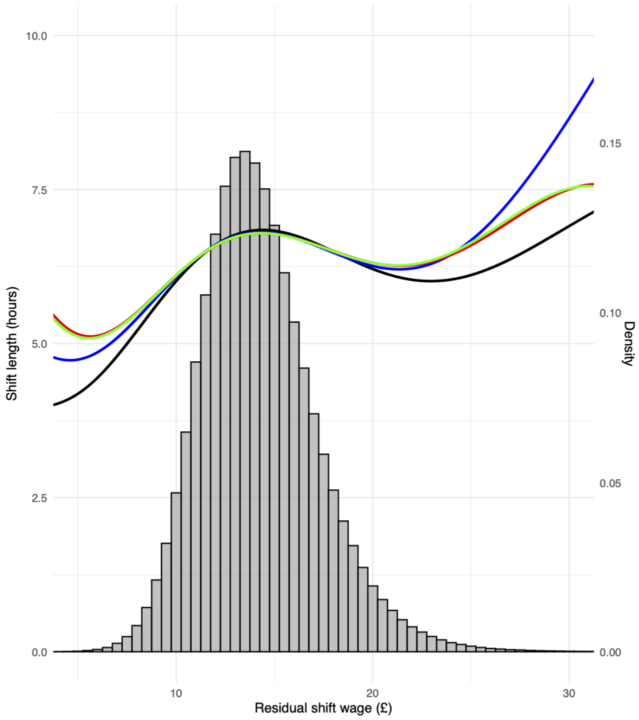
Notes: This figures presents the empirical density histogram of shift length for the balanced sample, where each bin is 30 minutes.

Figure 4: Labor Supply Function



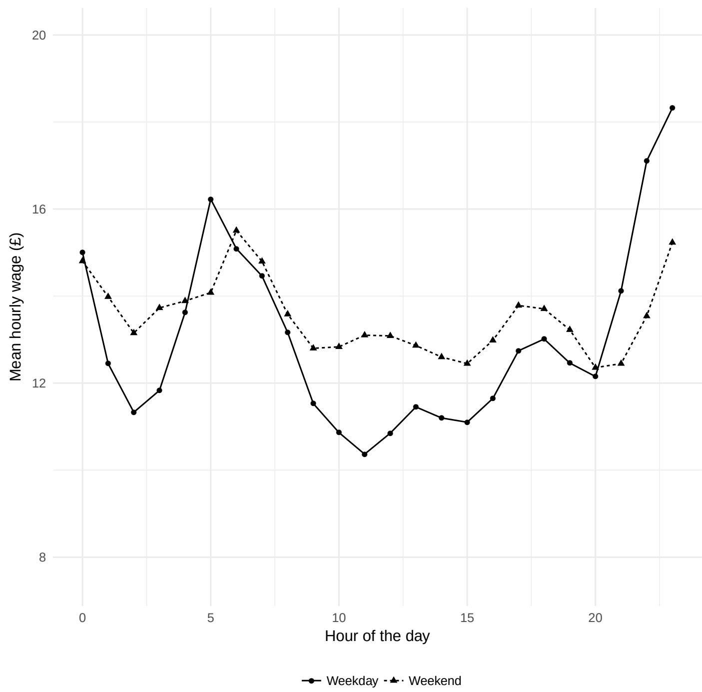
Notes: This figures presents the function identified by the control function model. The grey bays denote the empirical density histogram of the residualized wage rate, where each bin is an hour of the day, and the dashed red line marks the mean of that variable.

Figure 5: Robustness to Control Function Form



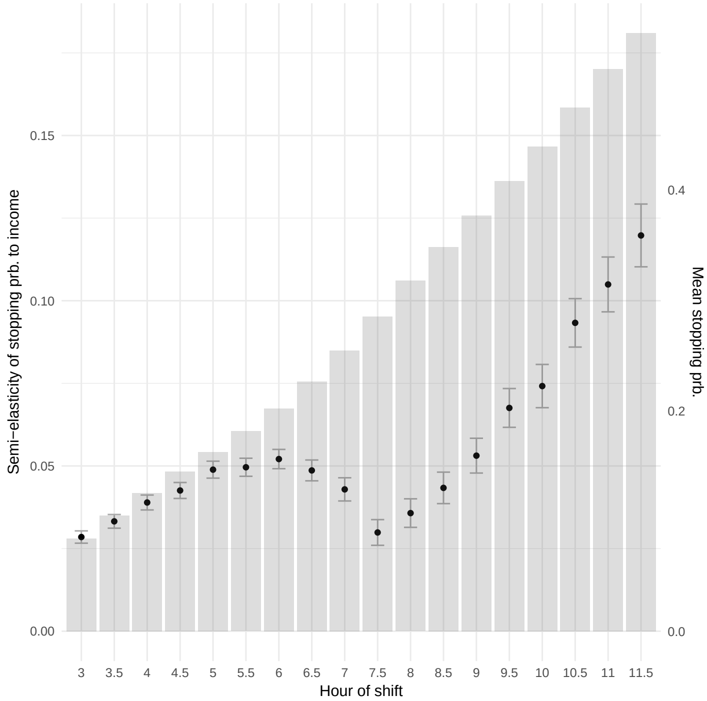
Notes: This figures presents the functions identified by the control function models which specify the control function as linear, quadratic, cubic, and quartic.

Figure 6: Wage by Hour of the Day



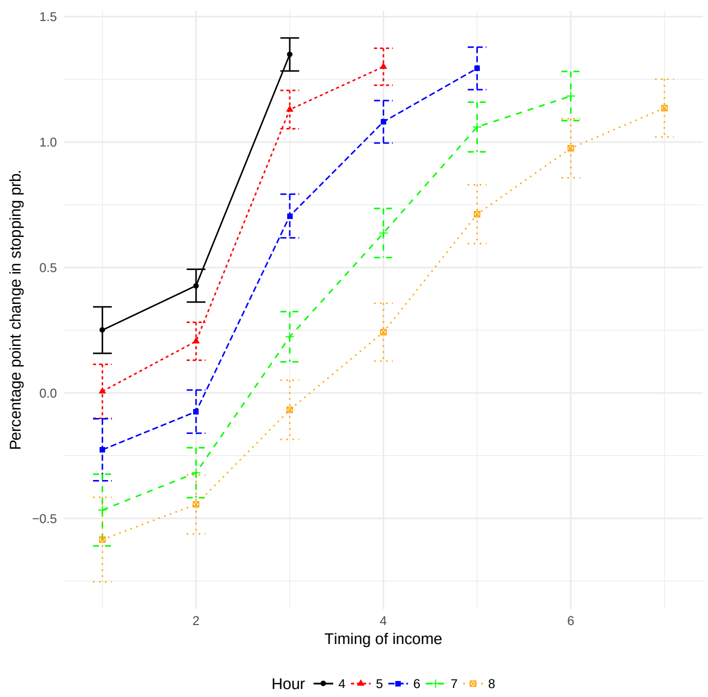
Notes: This figure presents the mean wage at every hour of the day, which is constructed by allocating earned fares to hours of the day proportional to the hours in which the job took place.

Figure 7: Stopping Model Income Effect Estimates



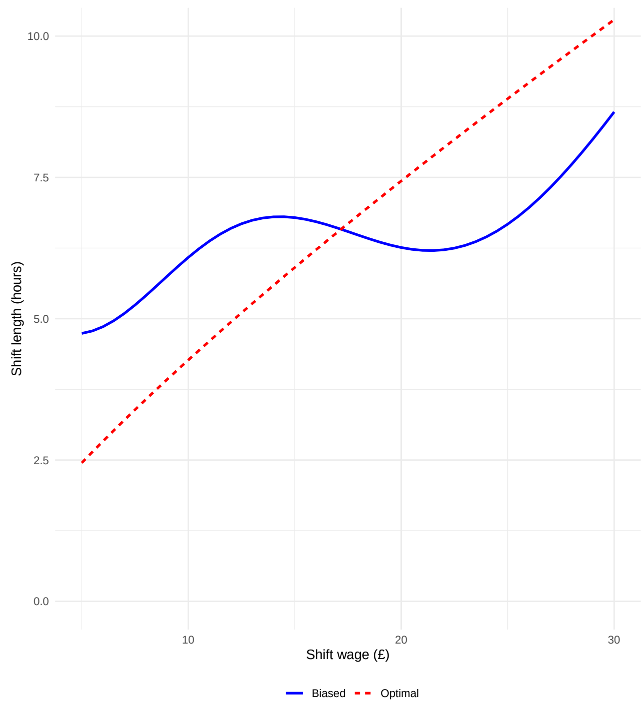
Notes: This figure presents the coefficients on accumulated income from the probability stopping model regressions as black dots with standard error bars, and the mean stopping probabilities as grey bars.

Figure 8: Stopping Model Timing Effect Estimates



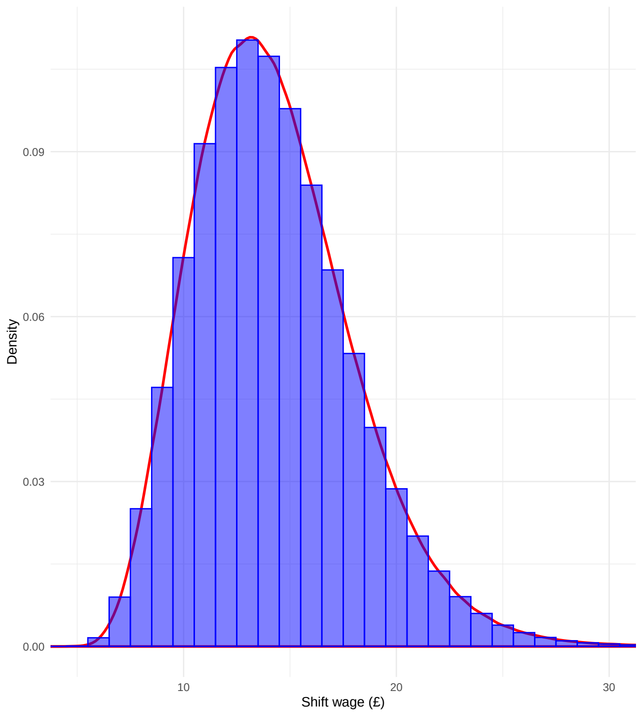
Notes: This figure presents the coefficients on income earned at different durations in shifts for shifts of different durations. For example, the blue line shows coefficients from a regression on shifts six hours through, and shows the influence of income earned in hour one, two, three, and so on, on the probability of stopping.

Figure 9: Biased Versus Optimal Labor Supply



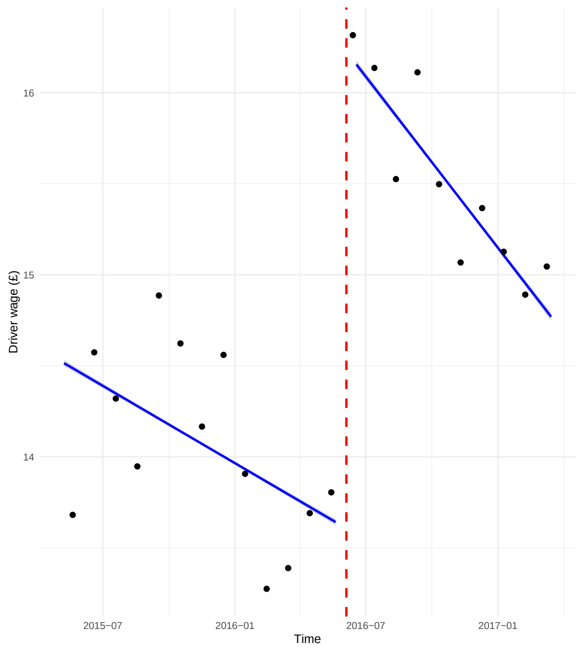
Notes: This figure plots the labor supply function implied by the quadratic control function model in contrast with the labor supply function implied by satisfying the intratemporal optimality condition, where the level parameter is specified by the level of hours during Tube strikes.

Figure 10: Shift Wage Density



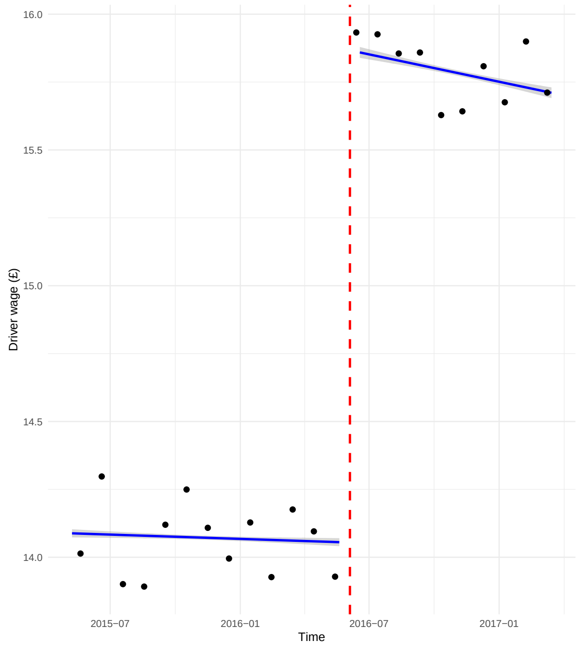
Notes: This figure plots the empirical density of wages in the sample with a fitted kernel density, which is used to integrate over welfare losses.

Figure 11: Raw Marshallian First Stage



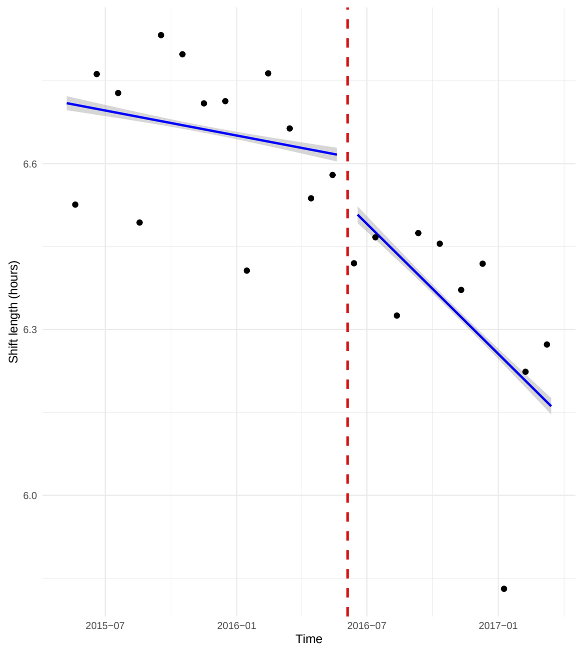
Notes: This figure plots mean monthly wages with a blue line representing a fitted linear trend either side of a fare reform, which is marked with a dashed red line.

Figure 12: Residual Marshallian First Stage



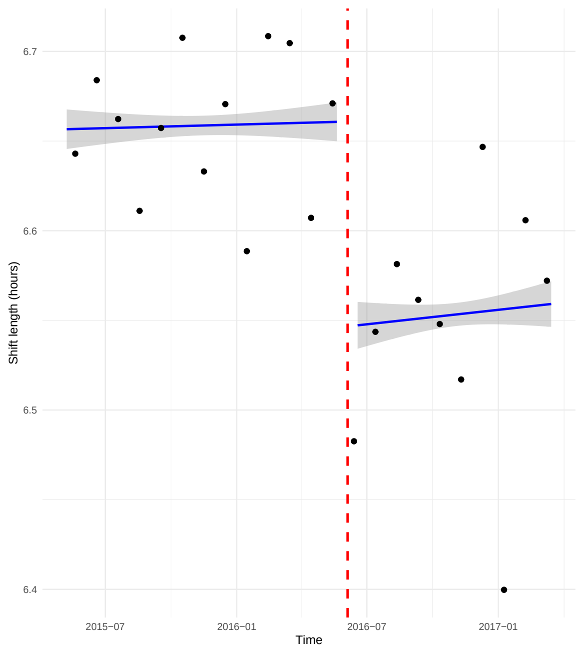
Notes: This figure plots mean residualized monthly wages with a blue line representing a fitted linear trend either side of a fare reform, which is marked with a dashed red line.

Figure 13: Raw Marshallian Second Stage



Notes: This figure plots mean monthly shift lengths with a blue line representing a fitted linear trend either side of a fare reform, which is marked with a dashed red line.

Figure 14: Residual Marshallian Second Stage



Notes: This figure plots mean residualized monthly shift lengths with a blue line representing a fitted linear trend either side of a fare reform, which is marked with a dashed red line.