Note: The results in this paper are made with proprietary data which cannot be released. Furthermore, as a condition of using the data, I cannot identify the source of the information, and I cannot identify appropriate summary statistics which may give away market information. This also means I cannot release standard errors which may give approximations of sample size. However, I do indicate the levels of significance where appropriate.

Income Targeting and the Ridesharing Market

Michael Sheldon

February 18, 2016

1 Abstract

In this paper, I replicate and expand upon the findings of Camerer et al. (1997) using a novel set of data centered on the supply side of a ride-sharing market. I analyze nine months of data from a peer-to-peer ride-sharing firm in which independent contractors freely choose their hours and receive a fixed commission for every trip they complete. First, I run a regression of (log) hours on (log) hourly earnings to compute a labor supply elasticity. I find that, in contrast to Camerer et al., elasticities are positive and highly significant. These results also hold when using the average income of all partners as an instrument for one partner's own hourly income. Then, I demonstrate that partners' elasticities increase with experience, suggesting that income-targeting behavior, if present, is only temporary. Finally, I demonstrate that small measurement errors will lead to spuriously negative income elasticities–potentially explaining the negative labor supply elasticities found by previous studies.

2 Introduction

In 1997, Camerer et al. published a widely influential, and controversial, paper on the labor supply of New York City taxi drivers. The authors found that, in contrast to the standard neoclassical theory, drivers demonstrated negative labor supply elasticities; that is, as earnings for a day increased, taxi drivers drove less total hours. Camerer et al. justified these findings conceptually by developing the theory of income-targeting, which supposes that taxi drivers (and other workers in similar environments) decide their labor supply based on daily, heuristic targets for total earnings. The income-targeting theory is rooted in the reference-dependent preferences literature, and has provided one of the most provocative examples of reference-dependent preferences outside of laboratory settings.

Camerer et al.'s findings, however, have not come without substantial contention. Considerable challenges question the original Camerer et al. findings on both conceptual and econometric grounds. First of all, as emphasized by Farber (2005), the income-targeting theory imposes strong assumptions of irrationality on taxi drivers; by allocating labor supply disproportionally to times of low earnings, taxi drivers are losing substantial amounts of both income and leisure time. While violations of rationality are not uncommon, they are rarely found in decisions with high stakes and repeated opportunities to learn. More importantly, econometric challenges question whether Camerer et al.'s findings of negative elasticities are valid. In their base regression, Camerer et al. encounter a well-known "division bias" which can lead to spuriously negative elasticities when measurement error is present. Though the authors acknowledge this and attempt to correct for it with an instrument, the validity of their instrument is also questionable. Though Camerer et al. remains hugely influential, these lingering questions cast enough doubt that further research is imperative to address them.

In this paper I attempt to replicate the findings of Camerer et al. by using a new, comprehensive dataset of a scale previously unrealized in this literature. I use the trip and supply activity data for partners of a peer-to-peer ride-sharing platform (henceforth "partners" and "the firm", respectively) to run an analysis which mirrors and then expands upon the original Camerer et al. setup. Using data on partners' earnings and hours logged for the city of Chicago, I regress (log) hours on (log) hourly earnings in order to calculate a labor supply elasticity. In contrast to Camerer et al., I find significantly and substantially positive elasticities. These results hold when instrumenting for income as in the Camerer et al. analysis. I then follow Farber (2014) in order to demonstrate how partner elasticities change over experience. I find that for partners overall, elasticities start positive and increase with experience; for partners who start with negative elasticities, I demonstrate that elasticities quickly increase and eventually become positive. Finally, I utilize the precision of my data to show the effect of "white noise" measurement error on the estimation of labor supply elasticities; by artificially adding error to my data, I demonstrate that calculated elasticities can become downward biased and appear negative.

2.1 Theories of Labor Supply

The neoclassical and behavior perspectives on labor supply offer not only vastly different frameworks for the worker's labor supply decision, but opposing predictions on the sign of labor supply elasticities. The neoclassical view of labor supply follows from the lifecycle model, in which individuals maximize their lifetime utilities in a multi-period maximization problem with substitution across periods[1]. Individuals hold some forecast of future wages and can decide their amount of labor and leisure in each period with complete flexibility. In this setting, a transitory wage shock (i.e, a single-period wage change such that income effects are negligible) predicts a change in labor supply in the same direction. Simply put, individuals respond to a temporary wage increase by

[1]For more information about the lifecycle hypothesis, see Modigliani (1966)

supplying more labor now in exchange for future leisure where the opportunity cost of leisure is lower. Labor supply elasticities along both the extensive and intensive margins are therefore predicted to be strictly positive.

Behavioral perspectives of labor supply stem from the concept of loss aversion within prospect theory (Kahneman and Tversky, 1979; 1991) and its refinement through the reference-dependent preference models of Kőszegi and Rabin (2006; 2007; 2009). Prospect theory postulates a kink in an individual's utility curve around a contextual, predetermined reference point[2], where losses relative to this point are weighted more heavily than gains. Camerer et al. (1997) argue that taxi drivers have a reference point for income and that this reference point greatly influences a taxi driver's daily labor supply decision. Furthermore, Camerer et al. suppose that taxi drivers consider their labor supply only on the daily level, rather than apart of a lifetime problem, as a result of "narrow bracketing" (Thaler, 1985). Thus, workers are highly influenced by daily earnings relative to their daily target. Since losses are weighted more heavily than gains around the target, taxi drivers are more highly motivated to work when they are below their income target than above it. So on days when income rates are lower, taxi drivers will work more; on days when income rates are higher, taxi drivers will work less. Therefore, the income-targeting hypothesis predicts a negative relationship between hours worked and rates of income, and thus a negative (intensive margin) labor supply elasticity.

To explicitly address the determinants of a partner's labor supply, I present and discuss the supply model penned by Farber (2005). As Farber notes, the partner's true optimizing model of labor supply is the solution to a dynamic programming problem that is much beyond the scope of this and other studies

[2]The theory surrounding the development of a reference point, its longevity and heterogeneity between individuals and contexts is altogether unclear. In context of labor supply Crawford and Meng (2013), following the work of Kőszegi and Rabin (2006), claim individuals develop reference points for both hours and income based on each worker's rational expectations

in this literature. Nevertheless, this model provides a useful reference for the differing assumptions and implications between the neoclassical and behavioral theories.

Consider a partner's intertemporal labor supply problem defined over some periods T, where the partner derives utility from consumption x_t and leisure l_t in each t period. As with Farber, I assume utility that is additively separable within and between days.

The partner's utility for period t is then

$$U_t = c(x_t) + r(l_t)$$

Where $c(\cdot)$ is utility from consumption and $r(\cdot)$ is utility from leisure; both functions are strictly concave in their arguments.

The partner's lifetime utility is then

$$U = \sum_{t=0}^{T}(1-\rho)^{-t}[c(x_t) + r(l_t)]$$

where ρ is the partner's rate of time preference. The lifetime budget constraint is given by:

$$Y_0 + \sum_{t=0}^{T}(1+r)^{-t}y_t(1-l_t) = \sum_{t=0}^{T}(1-r)^{-t}x_t$$

Where y_t is daily earnings, $(1-l_t)$ is work hours, and r is the market discount rate.

The partner wishes to maximize lifetime utility subject to his budget constraint. The optimization problem takes the following Lagrangian form:

$$V = \sum_{t=0}^{T}(1-\rho)^{-t}[c(x_t) + r(l_t)] + \lambda[Y_0 + \sum_{t=0}^{T}(1+r)^{-t}y_t(1-l_t) - x_t]$$

Where λ has the interpretation of the marginal utility of lifetime wealth.

Solving for the partner's first order conditions, we get the following set of equations:

$$\frac{\partial V}{\partial x_t} = c'(x_t) - \lambda\theta^t = 0$$
$$\frac{\partial V}{\partial l_t} = r'(x_t) - \lambda\theta^t y_t(1 - l_t) = 0$$
$$\frac{\partial V}{\partial \lambda} = Y_0 + \sum_{t=0}^{T}(1 + r)^{-t} y_t'(1 - l_t) - x_t = 0$$

Where, as in Farber (2005), $\theta = \frac{(1-\rho)}{(1-r)}$ measures the partner's patience relative to the market. As θ becomes large, the partner is impatient relative to the market and vice versa. Rearranging the first pair of first order conditions, a partner's optimal choice is characterized by the following relationship between his marginal wage and his marginal rate of substitution between consumption and leisure:

$$y_t'(1 - l_t) = \frac{r'(l_t)}{c'(x_t)}$$

While the model does not explicitly present a model of partner labor supply, the implicit model makes clear that optimal labor supply depends crucially on λ and θ. As the time horizon becomes longer, short-term fluctuations in income matter less and therefore income effects are less important in the t period labor decision. However, as the time horizon becomes shorter, labor supply decisions depend more on these income effects. θ governs this relationship; the more patient the partner is relative to the market, the less important income effects become. If the partner is sufficiently patient, transitory variation in the marginal wage evokes a positive supply response; the result is the neoclassical prediction of a positive supply elasticity.

However, the behavioral income-targeting hypothesis introduces a variety of assumptions that distort this relationship. First of all, the income-targeting hypothesis imposes the "narrow bracketing" restriction, implying a high degree of partner impatience as partners consider their labor supply decision only on the daily level–or, in Camerer et al.'s terms, "one day at a time". In terms of the model, this is the case of an extremely high θ. Furthermore, the behavioral theory combines this condition with imposing both: a kink in the marginal utility of consumption at the "target income", altering the value of additional consumption from very high to very low, and a kink in the opposite direction with respect to the marginal utility of leisure. Under these conditions, negative elasticities will exist near the target income. To see this, say a partner has reached their target level of income, x_t^* (modeled here by reaching a target level of consumption goods). As the partner transitions from x_t to x_t^*, the marginal utility of consumption drops substantially. Even if the marginal utility of leisure remains constant, the marginal rate of income would need to increase discontinuously to motivate a partner to supply additional labor[3]. Therefore, if positive income variation pushes a partner to his target income, this partner is likely to stop supplying labor; thus, higher rates of income will be negatively related with labor supply.

2.2 Findings in the Literature

Though for decades the lifecycle model had been the predominant theory for conceptualizing labor supply, early researchers attempting to verify its predictions in the field met little success. For one, the model assumes a flexibility in hours such that a worker can respond to transitory wage incentives; in reality,

[3]As can be inferred from this discussion, the effect of the income target is a local phenomenon–at earnings far below the target and far above the target, we would expect the partner to behave in the neoclassical manner. Though this is an important feature of an income target, it is not discussed further in this paper for two main reasons: first of all, Camerer et al. (1997) do not explicitly address this in their empirical design and secondly, if the driver is so often below/above their income target such that the income target does not substantially alter their labor supply, it is not very important in the first place

labor supply hours come with considerable rigidity. Secondly, shifts in wages are rarely transitory and therefore often accompanied with substantial income effects. Lastly, researchers have traditionally relied on aggregate workforce data which often suffers from limited comprehensiveness and great potential for measurement errors.

While these issues have likely contributed to the long series of insignificant or weakly positive estimates of labor supply elasticity[4], literature emerging around the new century offered a better approach. Centered on markets where workers could freely choose hours, many of these studies find significant and substantially positive labor supply elasticities along both the intensive and extensive margins. Carrington (1996) studies behavior of workers on the Trans-Alaskan pipeline, and finds that both the number of workers (extensive margin) and hours worked per worker (intense margin) are quite elastic and positively related to large, temporary, and anticipated demand shocks to labor. Oettinger (1999) studies the labor supply behavior of vendors in a baseball stadium, who are free to choose which games and for how long in each game to work for. Oettinger again finds substantially and significantly positive elasticities along the extensive and intensive margins. Furthermore, Oettinger stresses and demonstrates the importance of wage variation being endogenously determined by the labor supply of vendors–when wage variation is treated solely as a product of demand shocks, as in most studies presented here (including the current study), labor supply elasticities can be severely downward biased. Stafford (2013) studies the labor supply of Florida lobster fishermen, and again finds that wage shocks are positively related to labor supply.

The most substantial findings of negative income elasticities, key in the development and support of the behavioral model, come from studies of taxi drivers.

[4]For examples, see Killingsworth and Heckman 1986; Pencavel 1986; Blundell and MaCurdy 1999

Camerer et al. (1997) introduced their famous income-targeting hypothesis based on a dataset of trip sheets obtained from New York City cab drivers. Camerer et al. claim that the taxi market provides an ideal setting to study labor supply elasticities since daily wages are influenced by a variety of transitory shocks with low variance within days and high variance between days. From an analysis on the available data concerning driver incomes and time spent working, Camerer et al. found substantial and significantly negative elasticities. In 2000, Chou replicated this study with a similar approach in Singapore; again finding evidence of significantly negative elasticities. The most substantive support for Camerer et al.'s original finding comes from Crawford and Meng (2013), who test an alternative model and specification with targets for both income and hours. They find strong effects of the targets on local labor supply decisions in support of Camerer et al. and reconcile findings by Farber (2008), who suggests income targets (even if present) are not important in supply decisions. Outside of the taxi industry, support for Camerer et al. has been lacking. However, Fehr and Goette (2007) do find mixed evidence in tepid support of the behavioral model; they find when bicycle messengers experience a month-long piece-rate increase, the messengers exhibit positive elasticities overall but work less on the daily level (compensated by working more days in total).

The original Camerer et al. paper, which laid the groundwork for the behavioral model of labor supply, remains highly influential. However, many researchers have questioned these findings on both theoretical and econometric grounds; the most prominent being Farber (2005; 2008). Farber questions the methodology of Camerer et al. who run their analysis by regressing hours on the reciprocal of hours; thus creating a division bias which will push the coefficient downward in the presence of any measurement error in the hours variable. Even though Camerer et al. acknowledge this issue and attempt to remedy it with an instrument of average hourly wage, Farber notes that due to small sample size and the possibility of calendar date effects that the instrument may be ineffective. Furthermore Farber (2014), in a yet-unpublished report using five years of electronic NYC taxi data, finds substantially positive wage elasticities once controlling for this bias. This upcoming analysis provides a serious challenge to the original findings in Camerer et al., with Farber suggesting that Camerer et al.'s results were not genuine findings of income-targeting behavior but rather misapplied econometrics.

3 Data and Approach

For this analysis, I use trip and hourly activity data for drivers partnered with a peer-to-peer ride-sharing firm within the city of Chicago over a period of nine months. This data includes detailed information for each partner, including but not limited to: total number of completed trips, gross fares from each trip, time spent on the application (in which the partner is available to requests from clients), and the relevant timestamps for all of these activities. All data is collected automatically and stored on company servers, with little to no risk for direct measurement error in any of the variables. For each partner the date of their first completed trip is known. Although precise numbers cannot be disclosed, the dataset I analyze contains data for over one-thousand unique partners and hundreds of thousands of unique trips. A significant advantage of this dataset compared to other similar studies is that the sheer sample size and accuracy of information allows for a high degree of power and confidence in the results on both an aggregate and individual basis.

3.1 Definitions and Data Cleaning

3.1.1 Market Overview

The focus of this analysis is the supply side of a peer-to-peer (P2P) transportation market in which a technology firm acts as a conduit between clients seeking a ride and independent contractors (partners) who elect to complete the ride. The firm provides both the partners and clients with a free mobile phone application which the firm maintains. The client selects which type of ride service she wants, and a partner who provides that service will receive a dispatch notification which he can either accept or reject. If the partner rejects the dispatch, it will be forwarded to another available partner. If the partner accepts the dispatch request, he will be routed to pick up the client. Once the trip is completed the partner receives a fixed commission of the total fare which is a product of trip distance, trip time, and a flexible pricing schedule which is known to both the partner and client before the trip begins.

The firm provides different tiers of products and services and, in some cases, a partner can serve on multiple tiers. For example, some partners may be licensed chauffeurs with luxury vehicles, and clients can request (for higher prices) this specific tier of service. For this analysis we focus only on partners who provide solely on the non-licensed peer-to-peer service; partners who provide across multiple service tiers are not considered since their labor supply cannot be reliably and accurately measured[5]. All payments are electronic and handled by the firm; the partner receives regular payments based on their gross earnings and commission. Partners, unlike taxis, are not permitted to accept fares from street-hails or accept any fares outside of the system while on the application.

[5] For example, some partners will work primarily for private companies outside of the firm's network. Many partners use the firm's product only between scheduled trips from outside companies

Tipping is also discouraged by the firm and is likely a rare occurrence.

The firm records data based on a partner's application activity, the partner's number of trips, and the gross amount of earnings brought in by the partner. This data is complete and accurately timestamped within the precision of a minute. I use the application activity of a partner as a direct measure of labor supplied; this includes time spent on the application searching for fares, driving to clients, and time with clients. Partners can only make trips with the application on, but otherwise have no clear incentive to leave the application on if they have no intention of completing fares[6].

3.1.2 Partner Shifts and Hourly Earnings

An important consideration and unique challenge of this dataset is the conceptualization of a partner's "day" or "shift". In contrast to papers studying the labor supply of taxi drivers (i.e. Camerer et al. 1997, Chou 2000, Farber 2005; 2008), the firm's partners are truly free to choose their hours in that they are not systematically limited by daily leasing periods and that their status as independent contractors does not come with any obligatory hour constraints. Barring individual constraints outside the industry, partners can choose exactly which days to work and how many hours to work on any given outing. While this flexibility provides a clear benefit to the analysis in that partners can have greater ability to adjust supply to changes in earnings, the flexibility poses a challenge in defining the exact day-level time horizon (particularly important for income targeting).

Multiple approaches could be considered to define a partner's day. First, one could define a partner's short-term horizon by the actual calendar day. However, this approach would split in two any shift which crosses over midnight–which,

[6]In fact, partners are slightly discouraged from this behavior since they are expected by the firm to accept a reasonable amount of dispatches, and will be temporarily forced off the system during periods of inactivity.

particularly on weekends, are times when many partners continue to drive. Such an action would bias elasticities toward zero if enough partners started their shift before and ended past midnight[7], as well as add unnecessary noise into the system. A modified but similar approach is used by Farber (2014) in his analysis of New York City taxi drivers; however, instead of separating a day at midnight he defines a day transition at 4AM. The reasoning is that 4AM is consistently the time of day with the lowest amount of drivers on the road so this effect from splitting shifts is, in his view, negligible. However, since this approach still runs into the same bias issues as above (although plausibly at a lower magnitude), it would be desirable to find a more intuitive and logical approach which minimizes the possible effect of introduced errors.

As a remedy to this issue, I define a shift such as the cluster of all trip and application activity which occur within a period of four hours in any direction. That is, every gap greater than four hours between application or trip activity marks the beginning of a new shift in the data. This definition is loosely based on a clustering analysis of shifts detailed in Appendix A. Conceptually, the use of a four-hour gap allows partners to have mid-shift breaks without counting such breaks as distinct shift; that is, treating returning to drive after a break as a decision on the intensive margin rather than the extensive margin. This is consistent with the clustering analysis as it appears that many partners are accustomed to taking midday breaks in periods when demand is low, but then continue to work later in the day when higher demand returns. Although the cutoff of four hours is somewhat arbitrary, altering the gap by a range of one or two hours does not produce substantial changes to the results.

[7]Say a partner systematically starts his shift at 10PM and ends at 2AM the next day if the shift had a particularly high earnings or 1 AM if the shift had low earnings. The first "shift" created by the split would reflect hours targeting, and the second "shift" would reflect intertemporal substitution. The hours targeting shift, however, is artificial and biases the elasticity towards zero. The same effect would occur if the partner was an income target earner

Hourly earnings, the central variable to the OLS analysis, are defined per partner, per shift as the quotient of gross fares over hours spent on application (shift length). This assigns an approximated "hourly income rate" for a shift. In the 2SLS section, I use the average hourly earnings of all other partners active during the same hours as an instrument for a partner's hourly earnings. One can view this instrument as a representation of a "city earnings rate" for all partners active in the city at the given time. This should be a valid instrument as I do not expect measurement errors in hours to be correlated across partners (thus the instrument should be orthogonal to a partner's own measurement error), and the average earnings for all other partners should be highly correlated with any given partner's earnings.

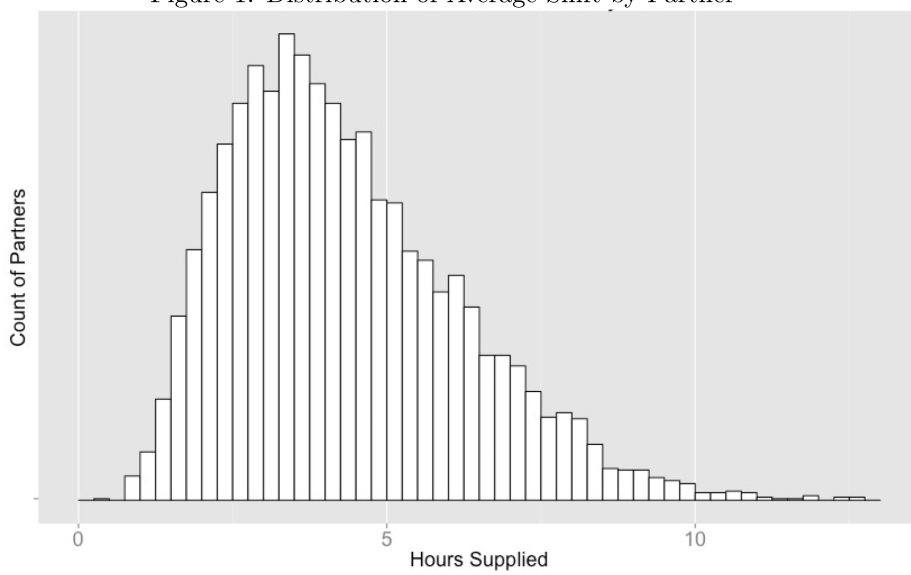
3.1.3 Data Cleaning

Simple measures were undertaken to ensure that only data relevant to the analysis were included in the final dataset. Since the analysis of labor supply elasticity along the intensive margin involves looking at partner behavior over the course of different shifts, minimum requirements for both the number of trips and shifts were imposed. To be included in the analysis, partners are required to have at least ten completed trips and at least five total shifts as defined above. Shifts, to be counted as such, must have at least one completed trip. This is to exclude partners who log into the application without the direct intention of making a trip; such as a partner looking to gather information on current demand and pricing but who decides not to commit to driving.

3.1.4 Shift Distribution

One of the biggest and most interesting differences between this P2P market and other comparable transportation markets is the flexibility in which labor

Figure 1: Distribution of Average Shift by Partner



can be supplied; partners can choose supply hours on their own accord with no particular institutional barriers. In a market such as this, understanding the distribution of partner hours is potentially interesting in many respects. First of all, it would give insight into how comparable this market is with other transportation markets in terms of first-order supply decisions. The aggregate variation could also provide a sense of heterogeneity between partners as well as an overview of how "flexible" hours truly are. Lastly, the distribution may demonstrate "lumpiness" around certain hours thresholds, which could suggest that partners supply hours based on an hours-targeting behavior[8]. Figure 1 provides a histogram of partner hours per shift. The mean shift is 4 hours 18 minutes, the median is 4 hours, and the standard deviation is 1 hour and 52 minutes.

The shift lengths are particularly interesting because of their short duration

[8]For example, one may expect clusters around 6 or 8 hour shifts; a sign that partners use hours-based commitment devices in determining shift lengths

compared to other similar markets[9]. Combined with the relatively high variance, this suggests that partners have a great deal of flexibility in determining their hours on any given shift. Furthermore, there is a lack of noticeable clusters in the histogram around certain hour or half-hour marks–indicating that hours-targeting heuristics may not be systematically employed.

3.2 Approach

This analysis follows the main approach of measuring labor supply elasticities as described in Camerer et al. (1997) and Farber (2014). I begin by providing a deeper understanding of the time-series properties which are essential for understanding partner decisions and behavior. This includes autocorrelation of hourly earnings within days and variation of earnings between days. Following the approach of Camerer et al., I complete an OLS regression analysis of a partner's supply hours explained by his own earnings and other relevant controls. To deal with potential complications from the "division bias" described by Farber (2005) and also for consistency, I then proceed with a two-staged least squares regression by instrumenting for own hourly earnings with the average hourly earnings for all other partners active during the same hours. Then I go into broader arguments concerning evidence of learning among partners as well as highlighting challenges of measuring labor supply elasticities in the field with artificial injections of measurement error into my own data.

[9]These descriptive statistics do not change with comparable measures of a shift. For example, defining a shift as an entire day does not change the results dramatically.

4 Empirical Analysis

4.1 Autocorrelation of Hourly Earnings

As noted by Farber (2005), understanding the time-series properties of a partner's earnings opportunities is critical in determining optimal partner behavior. Autocorrelations in wages tell us what a partner can infer about future earnings from current earnings; these inferences, in turn, dictate a partner's response to income shocks. If autocorrelations are negative, then negative wage elasticities may actually result from classically rational behavior and therefore we could not distinguish between the lifecycle and behavioral models of labor supply. Only if autocorrelations are significantly positive can we make a claim that intertemporal substitution is optimal and therefore be able to test different predictions made by the lifecycle and behavioral models.

To test this, I use the time series for the median hourly income for each hour in the dataset. I run the autocorrelation over a lag of eight hours. These autocorrelations are provided in Figure 2 and Table 1 below. All of the provided autocorrelations are positive and significant to the 1% level. In contrast to Farber (2005) and in line with Camerer et al. (1997), this suggests that when shifts start as high earning days, they are fairly likely to remain that way.

Instead of doing this on the aggregate level, it would be ideal to run autocorrelations on a partner-level basis. However, on an individual level, the earnings per hour variable over a shift is rather noisy; I can only capture the fare in the hour the trip ends in, so there will be predictable hour to hour noise in the individual data. To deal with this, I take shifts that span six or more hours and regress the average earnings of the first half of the shift on the second half. The resulting autocorrelation is .44, which is also significant to the 1% level. These findings, in combination with the aggregate level data, confirm that within-shift

Table 1: Autocorrelations by Hour

Hours:	1	2	3	4	5	6	7	8
Autocorrelation:	.722	.471	.283	.202	.169	.122	.096	.114

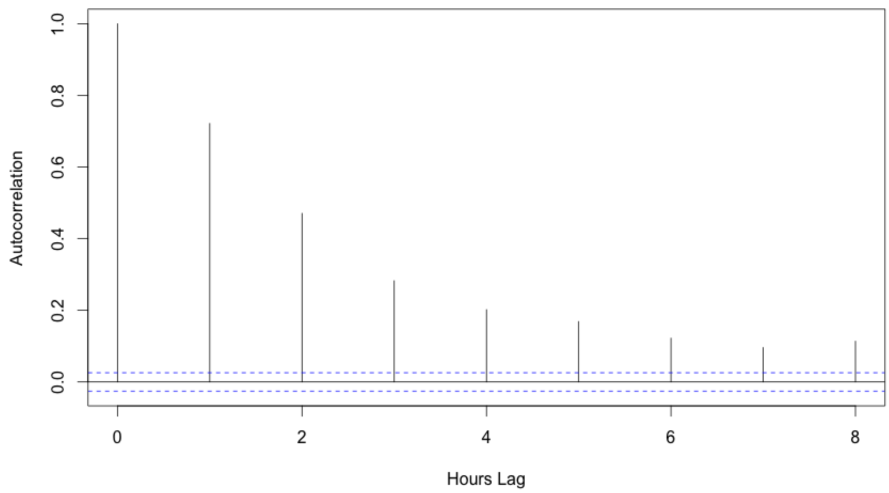
earnings are relatively stable.

Another important time-series trend, especially for the behavioral model, is the relationship of earnings between days. If earnings are significantly related between days, the "one day at a time" bracketing of earnings may not be as convincing. In testing this autocorrelation, I find that the second day earnings are significant to the 1% level with a coefficient of .376 when controlling for day-of-the-week and month fixed effects. No other days are significantly related. However, there are valid reasons to doubt this has a substantial impact on the analysis. First of all, this relationship captures the earnings relationship before and after midnight; which naturally are highly related on certain weekdays. Secondly, beyond this midnight transition, many partners do not work consecutive days. This suggests a two-day bracket, if present, could only apply to a small percentage of the total partner base. Overall, this analysis favors the description in Camerer et al. (1997) which describes highly correlated earnings within days and insignificantly related earnings between days.

4.2 OLS Regression

In the spirit of the original analysis, I run a basic regression to determine the relationship between (log) hours and (log) hourly earnings. I regress hours of application activity on the previously defined hourly earnings variable along with other relevant controls. I also allow partner fixed effects to account for potential systematic differences between partners (particularly important in this market,

Figure 2: Autocorrelation by Hour



where a large amount of partner heterogeneity is expected). Hourly records of precipitation, temperature, wind speed, and humidity are included in the weather category[10] to determine potential shifts in the labor supply from direct costs weather may impose on partners or from changes in opportunity costs. The analysis results are provided below in Table 2.

The regression results show that the income elasticities are both positive and highly significant, in contrast to the findings in Camerer et al. (1997). Controlling for partner fixed-effects alters the results substantially; this is likely due to the high degree of heterogeneity between different partners. Day of the week and month effects are also highly significant. For the weather variables, only precipitation and temperature significantly alter partner behavior. These results also hold when using median regressions to control for outliers. The elasticities suggest that, on average, partners are not income target earners but

[10]Hourly weather data is available for almost every hour in the dataset from the Wunderground Historical Weather Database

Table 2: OLS of (log) Hours

Variables:	(1)	(2)	(3)
Log Own Income	0.05***	0.15***	0.14***
Constant	1.194***	1.244***	1.341***
Weather Controls			X
Fixed Effects:			
Partner		X	X
Month and Day of Week			X
R^2	0.001	0.36	0.38

*** Indicates significance to the 1% level. Robust standard errors are used, but withheld due to previously stated privacy concerns. R-sqaured values take intercepts as explanatory.

rather act in line with the lifecycle model. The positive elasticities found here are particularly remarkable since the OLS regression, as is, still has potential for a downward bias.

4.3 2SLS Regression

As mentioned previously, Camerer et al. note that “white noise” measurement error can produce a downward bias in the estimated elasticities. Their response is to employ a two-stage regression by instrumenting for the driver’s own earnings for the average earnings of other drivers in the same day. I repeat this procedure here and use the average hourly earnings of every other partner active during the same hours as an instrument. The results are given below in Table 3.

Compared to the OLS regression results in Table 2, the elasticities here are even greater in magnitude and still highly significant. They are also similar to elasticities found by Farber (2014) in his analysis of NYC cab data, as well to elasticities found in other markets (e.g., Ottenginer, 1999; Jonanson & Wållgren, 2013; Stafford, 2013). All other variables have effects similar to those described in the OLS regression results. Once again, including partner fixed effects has a dramatic effect on the analysis in terms of coefficient magnitude and significance; the inclusion of partner fixed effects has a substantial effect on the elasticity in the positive direction.

The results again run in direct contrast to Camerer et al., who still find negative elasticities in their 2SLS. The differences between results likely owe much to the validity of the instrument; though Camerer et al. and I use identically constructed instruments, their low sample size provides few overlapping drivers to calculate average earnings. Thus, their instrument is not as strongly correlated with a driver's own earnings as one would expect, and the weak instrument could fail to purge the bias. Furthermore, my analysis highlights the importance of heterogeneity between partners–controlling for partner fixed effects has a strong influence on calculated elasticities. Camerer et al. have little ability to include fixed effects due to limited data, but when included it diminishes some of the significance and coefficient magnitudes within their results. Overall, the results from both the OLS and 2SLS cast doubt that the original findings in Camerer et al. provide evidence of systematic income targeting within their market. At very least, the results here demonstrate that income-targeting behavior is not the norm within this P2P market, with partners on average demonstrating behavior consistent with the lifecycle model.

4.4 The Effect of Experience on Income Elasticity

As with Camerer et al. and Farber (2014), I look to see whether experience effects a partner's labor supply elasticity. The analysis here and in the previous studies is a subset of the well established learning-by-doing literature. However, unlike the majority of the literature which focuses on increases in production due

Table 3: 2SLS of (log) Hours

Variables:	(1)	(2)	(3)
Log Average Income	0.13***	0.25***	0.22***
Constant	1.194***	1.244***	1.341***
Weather Controls			X
Fixed Effects:			
Partner		X	X
Month and Day of Week			X
R^2	0.001	0.36	0.38

*** Indicates significance to the 1% level. Robust standard errors are used, but withheld due to previously stated privacy concerns. R-sqaured values take intercepts as explanatory.

to increases in worker experience[11], the focus here is not in changes of output levels but in direct changes in behavior. In this sense, the analysis is surprisingly novel. Levitt, List, and Syverson (2013) note that while many studies have support the notion of learning-by-doing, few studies have discussed the mechanism of *how* worker behavior changes, as a result of learning, in order to increase production. This analysis provides an interesting insight into the behavioral changes workers undergo by gaining work experience in a new environment.

While my analysis thus far demonstrates that, on average, partners behave in accordance with the neoclassical predictions, it may be that income-targeting behavior is present among inexperienced partners. To gauge the effect of experience, I run a 2SLS of hours active on average hourly earnings (still as an instrument for a partner's own hourly earnings) with interactions between the instrument and experience. Experience is measured by "bins" of the chronological lifetime shifts in increments of ten; that is, the first ten shifts a partner drives

[11]The classic examples include Wright's (1936) and Alchian's (1950) studies on airplane production during WWII, as well as Rapping's (1965) analysis on the production of *Liberty Ships*

comprise the first bin, shifts eleven through twenty comprise the next bin, and so forth until the last bin which includes all shifts following the seventieth. These bins of experience are incorporated into the 2SLS through dummy variables, and the interaction term with the instrument measures the marginal effect on the income elasticity. For this analysis I only use the subset of partners whose first trip occurred after the beginning of my data horizon (i.e., partners whose first trip occurred in the span of dataset) and whose lifetime shifts are greater than 80 (in order to control for potential selection effects, since target-earners may be more likely to leave the market over time).

I run two separate regressions: one using the overall cohort of partners meeting the previously described requirements, and another cohort meeting the same conditions and additionally whose elasticities for the first 25 shifts were strictly negative. The results are presented in Figure 3 below as well as Table 4. For the overall cohort, there is a general upward (albeit noisy) trend in income elasticities with respect to experience. This suggests that partners learn over time to substitute hours with a greater responsiveness to changes in income, even as most partners enter with positive elasticities. Once again, standard errors cannot be disclosed, but differences between the first and last bins are significant to the 1% level.

The cohort of drivers who begin as income targeters is particularly interesting as the drivers reveal a dramatic learning curve with respect to elasticities over experience. Though these individuals enter the market with quite substantial and significantly negative labor supply elasticities, these elasticities move sharply in the positive direction with shift experience such that by the 31st shift these partners, on average, are no longer exhibiting income-targeting behavior. Notice that this cannot be explained by selection since all partners in the sample have at least 80 shifts. This evidence strongly suggests that the role

Figure 3: Elasticities over Shift Experience

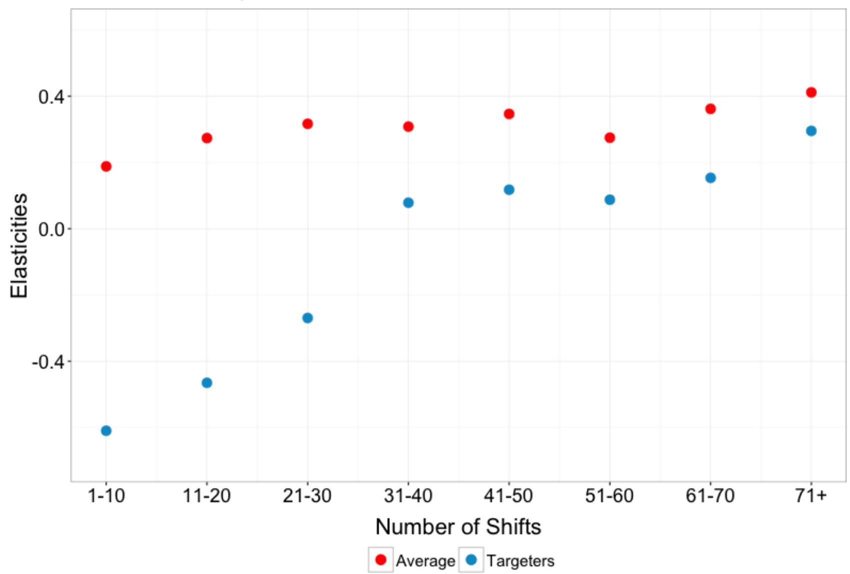


Table 4: Elasticities over Shift Experience

Shifts	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71+
All	0.188	.273	.317	.308	.346	.275	.362	.411
Targeting-Cohort:	-0.610	-0.465	-0.270	0.078	0.118	0.087	0.153	0.295

of income-targeting might be as a heuristic for new partners unfamiliar with how to engage a market with no requirements for hours. However, as partners gain experience in this market, it appears they respond less to this suboptimal heuristic and more to the earnings incentives which allow them to enjoy both higher incomes and additional leisure hours.

4.5 Estimating the Impact of Measurement Error

To try and explain the negative elasticities arrived at by Camerer et al. (1997), Farber (2005), and Chou (2000) in their OLS regression estimates, I attempt to gauge the effect of measurement error[12]. Since the firm has highly accurate data on partners' application activity (the only times they can be dispatched by the system), I can plausibly assume minimal measurement error in the recording of time spent "working". Given this assumption, we can then test the effect of measurement error by introducing artificial white noise into the system. Though this is a bias acknowledged by Camerer et al., they do not state the extent to which this bias can alter their results. I rerun the OLS regression in Section 4.2 here, except I insert noise with an average magnitude of 15 minutes into the "application activity" measurement. The results are given below in Table 5.

This analysis shows that only a relatively small amount of noise in minutes worked can greatly change the results of the OLS[13]. Since both Camerer et al. (1997) and Farber (2005) have to rely on trip sheet data, and since the beginning and end of shifts is marked by the first and last trip (thus neglecting any time spent searching for the first fare or after the last one), measurement error of a large magnitude is likely to arise and put a downward bias in the estimates. Furthermore, even though Camerer et al. implement an instrument to control for this, aforementioned validity concerns question whether this instrument effectively purges the relevant bias. Overall, the results presented here and from the previous regression appear to contradict the findings of Camerer et al.; only when measurement error is present and uncontrolled for do I find any indication of negative labor supply elasticities. Thus, a plausible and certainly supported

[12] I elaborate on the nature of measurement error's effect on the model in Appendix B

[13] It is important to note that 15 minutes of measurement error in my data would have a stronger effect than the same amount in Camerer et al.'s or Farber's data since they have systematically longer shifts. A likely equivalent amount of error in their data would be around 25-30 minutes worth.

Table 5: OLS of Log Hours with Noise in Hours

Variables:	(1)	(2)	(3)
Log Own Income	-0.08***	-0.004***	-0.04***
Constant	1.194***	1.244***	1.341***
Weather Controls			X
Fixed Effects:			
Partner		X	X
Month and Day of Week			X
R^2	0.001	0.36	0.38

critique of Camerer et al. is that measurement error may be driving the findings of negative labor supply elasticities in their analysis.

5 Discussion

My analysis fails to replicate the original results of Camerer et al. (1997) which supported the income-targeting hypothesis; instead, my analysis provides results consistent with neoclassical theory. Both the OLS and 2SLS estimates produce significant and substantially positive income elasticity estimates. These results are bolstered by an unusually comprehensive and large dataset which avoids many of the common issues found throughout the field; some of which can bias income elasticity estimates downward. Here I briefly outline what, in my opinion, is the role for income targeting in this market, potential market differences which may explain why my results differ from Camerer et al.'s, and issues with my analysis which could be corrected in further studies.

5.1 The Role of Income Targeting in the P2P market

Although I find that income elasticities are on average positive throughout my analysis, this does not mean that income-targeting behavior is absent from influencing partner decisions. In fact, the income-targeting hypothesis has a certain intuitive appeal which should not be ignored. The P2P market in this analysis is extraordinarily unusual in that it is one of the few major modern markets where individuals have complete discretion over hours supplied. Individuals new to this market might seek a soft commitment device; a way of mentally obligating themselves into a certain amount of output to prevent against procrastination, underperformance, or any other related (and well documented) behavioral traits. I believe this is the best way to interpret the learning data from Section 4.4. A substantial, although not most, fraction of partners do in fact come into the market with income targeting behavior. However, since intertemporal substitution provides better returns for both income and leisure, income targeting behavior is rather quickly learned away in favor of more optimal decision making.

As Farber (2005) explains, it is difficult to imagine that the behavioral model would persist since it means that partners are missing easy opportunities to earn large rewards; partners who income target could in fact earn substantially more income and enjoy more leisure if they simply took advantage of high earning hours. Partners who income-target are systematically leaving "money on the table" and so it is conceptually difficult to believe that this behavior would persist over multiple, repeated opportunities. Many suboptimal behavioral traits occur in contexts where iterations are few and consequences are not immediate; such as with Thaler and Benartzi's (2004) analysis with savings and retirement plans. But given that partners have the opportunity to learn on each successive shift with immediate effects on rewards, the psychological justification for the persistence of income-targeting behavior is weaker. Therefore, intuition backed by the empirical results of Section 4.4 suggests that income targeting may be a useful tool for inexperienced partners, but one that is quickly learned away in favor of more rewarding decisions.

5.2 Other Potential Sources of Differences

Since my approach is very similar to Camerer et al.'s and since the markets we study are of a comparable nature, I find it unlikely that structural differences drive the discrepancies in our findings. Thus far, explanations for our differences have been driven by extensive differences in our data. However, one could suggest that the P2P service (which, among other things, requires a partner to bring their own vehicle onto the system) attracts a fundamentally different kind of population. In order to dramatically alter the results, this P2P partner population would somehow have to be more attuned to the concept of intertemporal substitution and less likely to fall for the income-targeting heuristic. While a lack of demographic information prevents me from exploring this in comparison to the taxi industry, I find it difficult to create any compelling arguments in this direction. Also, given that Camerer et al. object to the idea that their results are driven by traits particular to taxi drivers, it seems unlikely that partners in the market I study are systematically "biased" to optimize relative to the general population.

One plausible argument, as indicated by the distribution of shift lengths, is that the P2P population may be concentrated on part-time earners and that these part-time earners, as opposed to full-time earners, are more efficient in responding to income incentives. Perhaps these individuals have more entrepreneurial spirit than the average worker, and their participation in this

P2P market (likely on top of another form of work) indicates that they are particularly savvy in choosing hours. I find this explanation weak at best. In the absence of any demographic information on partners, it is difficult to say that any part-time contractor would know more about the market than a full-time cab driver; especially given the evidence on learning.

Nevertheless, I cannot rule out population differences with my dataset alone. The most compelling evidence to this extent comes from Farber (2014), who analyzes five years worth of NYC taxi data in an identical analysis. Farber finds positive elasticities in his 2SLS, as well as a learning trend similar to the one observed in my analysis. In combination with my analysis, this strongly suggests that the differences in results between this replication and the original Camerer et al. paper are derived not from structural differences in our markets but rather systematic differences in our data.

5.3 Issues within the Analysis and Future Steps

I am aware of at least a few issues within the data which are likely to produce biases in the results, albeit in the downward direction. Ottenginer (1999) demonstrates that failing to take into account the ability for income shocks to pull individuals into the market along the extensive margin will produce a downward bias in the elasticity estimate. Since I only observe the intense margin decisions of partners, I neglect to take this effect into account. However, this bias is only present if partners can accurately predict transitory income shocks before committing to driving. Although devices do exist to this extent, I have no information that shocks are systemically predictable.

Similarly, there is an endogeneity bias in that income is determined by both demand and supply shocks (Stafford, 2013). Failing to separate this out again results in a downward bias of the elasticity estimate. Although it would be ideal, I simply lacked an instrument which could separate this difference. Further analysis could attempt to find appropriate instruments in this manner, although few analyses have yet been able to accomplish this (an important example being Ottenginer 1999).

6 Conclusion

Camerer et al.'s work on labor supply elasticities remains among the most influential in the field of behavioral economics. Not only has it spurred a large amount of following literature, but it to this day provided one of the few and major examples of reference-dependent preferences “in the wild”. However, my replication on a similar market with significantly higher quality data fails to reproduce the findings in Camerer et al. In contrast, I find evidence of substantially positive and highly significant income elasticities. This indicates that the behavioral model of income targeting fails to explain the labor supply of individuals, and the results instead favor the predictions of the lifecycle model. I find that the most likely place for income-targeting in the market is as a short-term heuristic for new partners, though it is quickly learned away with experience. Furthermore the evidence suggests that the original Camerer et al. results, along with other similar studies producing negative wage elasticities, may be driven by a mixture of measurement error and potential misspecification.

References

Blundell, R., & MaCurdy, T. (1999). Handbook of labor economics. In O. Ashenfelter & D. Card (Eds.), (Vol. 3A, chap. Labor supply: a review of alternative approaches).

Borjas, G. J. (1980). The relationship between wages and weekly hours of work: The role of division bias. *The Journal of Human Resources*, *15*(3), 409-423. Retrieved from `http://www.jstor.org/stable/145291`

Camerer, C., Babcock, L., Lowenstein, G., & Thaler, R. (1997). Labor supply of New York City cabdrivers: one day at a time. *The Quarterly Journal of Economics*, *112*(2), 407-441.

Carrington, W. (1996). The Alaskan labor market during the pipeline era. *Journal of Political Economy*, *104*(1), 186.

Chou, Y. (2000). Testing alternative models of labor supply: evidence from taxi drivers in Singapore. *Research Paper no. 768, Department of Econonoics, University of Melbourne*.

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

Farber, H. S. (2005). Is tomorrow another day? the labor supply of New York City cabdrivers. *Journal of Political Economy*, *113*(1).

Farber, H. S. (2008). Reference-dependent preferences and labor supply: The case of New York City taxi drivers. *American Economic Review*, *113*(1), 46-82.

Farber, H. S. (2014). Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. In *Albert rees lecture: Society of labor economists meeting.*

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

Jonanson, D., & Wållgren, J. (2013). *Labor supply of Stockholm cab drivers: revisiting the evidence.* (Retrieved August 21, 2014 from http://arc.hhs.se/download.aspx?mediumid=1957)

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*(2), 263.

Killingsworth, M., & Heckmann, J. (1986). Handbook of labor economics,. In O. Ashenfelter & R. Layard (Eds.), (Vol. 1, chap. Female labor supply: a survey). Amsterdam: North Holland.

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

Kőszegi, B., & Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review*, *97*(4), 1047-1073.

Kőszegi, B., & Rabin, M. (2009). Reference-dependent consumption plans. *American Economic Review*, *99*(3), 909-936.

Leisch, F. (2006). A toolbox for k-centroids cluster analysis. *Computational Statistics and Data Analysis*, *51*(2), 526-544.

Levitt, S. D., List, J. A., & Syverson, C. (2013, 08). Toward an understanding of learning by doing: Evidence from an automobile assembly plant. *Journal of Political Economy*, *121*(4), 643-681.

Modigliani, F. (1966, Summer). The life cycle hypothesis of saving, the demand for wealth and the supply of capital. *Social Research*.

Oettinger, G. (1999). An empirical analysis of the daily labor supply of stadium vendors. *Journal of Political Economy*, *107*(2), 360.

Pencavel, J. (1986). Handbook of labor economics,. In O. Ashenfelter & R. Layard (Eds.), (Vol. 1, chap. Labor Supply of Men: A Survey). Amsterdam: North Holland.

Stafford, T. (2013). *What do fishermen tell us that taxi drivers don't? an empirical investigation of labor supply.* Retrieved from `http://ssrn.com/abstract=2262677`

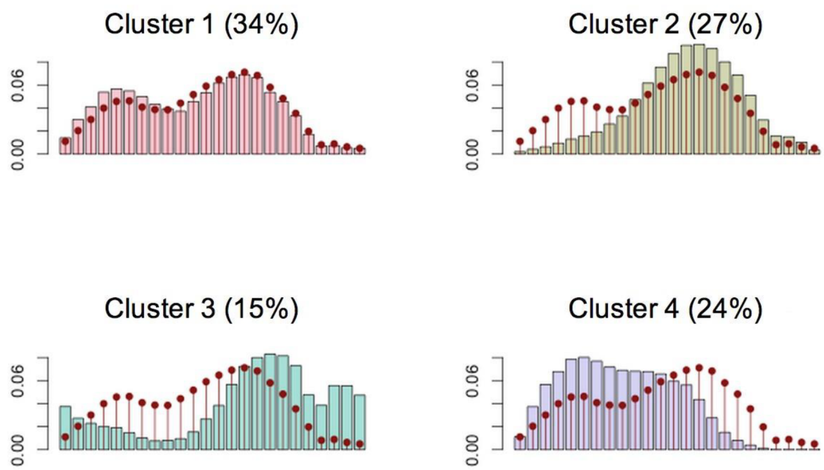
Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, *4*(3), 199-214.

Thaler, R., & Benartzi, S. (2004). Save more tomorrow^{TM}: Using behavioral economics to increase employee saving. *Journal of Political Economy*, *112*(S1), 164-187.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039-1061.

Wunderground. (2014, August). *Historical weather.* Retrieved from `http://www.wunderground.com/history/`

Figure 4: K-Median Clustering of Hour Activity



A Clustering Analysis of Shifts

Knowing the most common "types" of partner shifts is conceptually important for visualizing how partners allocate their time contingent on them driving that particular day. To determine the distribution and most common "types" of partner shifts, I do a k-median analysis on the distribution of a partner's hours. For each partner I count the number of times they were active on a given hour, and create a histogram where the y-axis gives the proportion of times they were active on that hour. I then cluster these distributions into four different groups, as shown below in Figure 4.

The clusters show clear and intuitive patterns in the distribution of partner hours. In each graph, the bars correspond to the cluster of interest while the dots give the overall "mean" distribution. Clusters are determined by a k-medians approach[14]. Based on purely conceptual knowledge, we can assign each cluster a subjective time-slot: Cluster 1, which applies to the greatest amount of partners,

[14] For more information on the clustering algorithm, see Leisch (2006)

and shows high activity around the two rush hour periods. Cluster 2 corresponds to partners who center their hours on the evening rush hour, and cluster 4 on the morning rush hour. Cluster 3 corresponds to partners who drive a night shift. From this clustering, we can see that the majority of shifts follow a contiguous pattern; even if breaks are taken, there is no evidence that they span more than a few hours. However, in deciding the optimal cutoffs between shifts, one needs to balance the normal "gaps" in the distribution without counting regular breaks; such as those which occur in Cluster 1. A four-hour span, in my interpretation, does this successfully; it seems to cover close gaps which reflect breaks without including conceptually distinct shifts. However, it should be noted that this analysis is almost entirely subjective; there is no explicit reason why four-hour cutoffs are any more useful than three or five hour ones. Although I cannot come up with a sufficient claim against this point, I believe basic intuition and comparability to other approaches indicates that the decision made here will not systematically skew the results of the analysis. Furthermore, as I have stated in my analysis, reasonable deviations from the 4-hour cutoff do not alter the results substantially.

B Measurement Error in OLS Regression

While measurement error is a persistent problem across economic studies, the effects of measurement error are typically limited to attenuation of the estimators. However, in the model analyzed here, the effects of measurement error in the hours variable are more perverse[15]. Consider the model

$$ln(Hours_{it}) = \beta_0 + \beta_1 ln(\frac{Earnings_{it}}{Hours_{it}}) \quad (1)$$

[15] The following discussion owes much to Borjas (1980)

Where i denotes the individual partner and t the partner's t^{th} shift in the data.

Consider the case where there is white noise measurement error in observed hours, such that we observe $Hours^*_{it} = Hours_{it} * \epsilon_{it}$, where $Hours^*_{it}$ is the observed hours worked on the shift and ϵ_{it} is a multiplicative error term independent of both income and hours.

Running this regression in the data, we get

$$\hat{\beta_1} \to \frac{cov(ln(Hours^*_{it}), ln(\frac{Earnings_{it}}{Hours^*_{it}}))}{var(\frac{Earnings_{it}}{Hours^*_{it}})}$$

Simple algebra reveals...

$$\hat{\beta_1} \to \beta_1(\frac{\sigma_I^2}{\sigma_I^2 + \sigma_\epsilon^2}) - (\frac{\sigma_\epsilon^2}{\sigma_I^2 + \sigma_\epsilon^2})$$

Where σ_ϵ^2 is the variance of the log hours error, σ_I^2 is the variance of income where $Income_{it} = \frac{Earnings_{it}}{Hours_{it}}$.

The decomposition of $\hat{\beta_1}$ is particularly informative here: the first term is the classical attenuation effect, which biases $\hat{\beta_1}$ toward zero as the error variance dominates the income variance. The second term is the term of particular interest for this study, referred to as the "division bias" in the literature. As the error variance dominates the income variance, this term goes to -1, putting a downward bias on the estimator. So the measurement error produces a twofold effect; not only does it attenuate $\hat{\beta_1}$ but it also produces a negative bias, and both effects increase as σ_ϵ^2 increases with respect to σ_I^2.

The decomposition here may help to explain the results in Table 2 when compared to Camerer et al. (1997). Since the data used in this analysis is more precise, we would expect that the measurement error variance is smaller within this data than within Camerer's. Additionally, it is reasonable to believe P2P partners face greater variance in earnings than cab drivers; dynamic pricing in the P2P market allows for greater fluctuations in fares a given partner may face. Therefore, greater income variance relative to measurement error variance would reduce the overall bias of elasticity estimator, even with a simple OLS design.