

LABOR SUPPLY OF NEW YORK CITY CABDRIVERS: ONE DAY AT A TIME*

COLIN CAMERER
LINDA BABCOCK
GEORGE LOEWENSTEIN
RICHARD THALER

Life-cycle models of labor supply predict a positive relationship between hours supplied and transitory changes in wages. We tested this prediction using three samples of wages and hours of New York City cabdrivers, whose wages are correlated within days but uncorrelated between days. Estimated wage elasticities are significantly negative in two out of three samples. Elasticities of inexperienced drivers average approximately -1 and are less than zero in all three samples (and significantly less than for experienced drivers in two of three samples). Our interpretation of these findings is that cabdrivers (at least inexperienced ones): (i) make labor supply decisions “one day at a time” instead of intertemporally substituting labor and leisure across multiple days, and (ii) set a loose daily income target and quit working once they reach that target.

I. INTRODUCTION

Dynamic models of labor supply predict that work hours should respond positively to transitory positive wage changes, as workers intertemporally substitute labor and leisure, working more when wages are high and consuming more leisure when its price—the forgone wage—is low (e.g., Lucas and Rapping [1969]). While this prediction is straightforward, it has proved difficult to verify. Estimated elasticities of intertemporal substitution have generally been low and insignificant, or even negative, whether they are based on aggregate [Mankiw, Rotemberg, and Summers 1985], cohort [Browning, Deaton, and Irish 1985], or panel [Altonji 1986] data (see also Laisney, Pohlmeier, and

*Many thanks to Bruce Schaller (NYC Taxi and Limousine Commission) for data and helpful discussions; James Choi, Kim Morgan, and Dov Rosenberg for research assistance; Charles Brown, Jeffrey Dominitz, John Engberg, John Ham, Seth Sanders, and Lowell Taylor for helpful discussions; Andrei Shleifer for the Weber quote; two referees, and editor Lawrence Katz for extraordinarily thorough comments, and colleagues at many workshops: The Caltech brown bag lunch seminar, NBER Behavioral Labor Economics meeting, and Behavioral Economics Summer Camp (sponsored by the Russell Sage Foundation), Carnegie Mellon University's Heinz School of Public Policy and Management, the University of California (Irvine and Berkeley) Departments of Economics, the MIT/Harvard Behavioral Economics seminar, the University of Chicago Labor Workshop, the Judgment/Decision Making Society, and the Econometric Society meetings. Address correspondence to the first author at Division of Social Sciences 228-77, California Institute of Technology, Pasadena CA 91125, camerer@hss.caltech.edu.

Staat [1992]; Pencavel [1986]; and cf. Mulligan [1995]). However, these results are difficult to interpret because actual wage changes are rarely transitory, so the hypothesis of intertemporal substitution must be tested jointly along with auxiliary assumptions about persistence of wage shocks, formation of wage expectations, etc. As a result, the frequently observed negative wage elasticities can plausibly be attributed to specification error.

The ideal test of labor supply responses to transitory wage increases would use a context in which wages are relatively constant within a day but uncorrelated across days. In such a situation all dynamic optimization models predict a positive relationship between wages and hours worked, due to the negligible impact on life-cycle wealth of a one-day increase in wage (e.g., MaCurdy [1981, p. 1074]).

Such data are available for at least one group of workers: New York City cabdrivers. Drivers face wages that fluctuate on a daily basis due to demand shocks caused by weather, subway breakdowns, day-of-the-week effects, holidays, conventions, etc. Although rates per mile are set by law, on busy days drivers spend less time searching for customers and thus earn a higher hourly wage. These wages tend to be correlated within days and uncorrelated across days (i.e., transitory).

Another advantage of studying cabdrivers is that, unlike most workers, they choose the number of hours they work each day because drivers rent their cabs from a fleet for a fixed fee (or own them) and can drive as long as they like during a continuous twelve-hour shift. Furthermore, most analyses of labor supply measure hours (and sometimes income) by self-reports. For cabdrivers, better measures of hours and income are available from "trip sheets" the drivers fill out and from meters installed in cabs, which automatically record the fares.

Using these data, we investigate the relationship between wages and hours worked, and find little evidence for intertemporal substitution. Most of the elasticities we estimate are negative; drivers tend to quit early on high wage days and to drive longer hours on low wage days. Elasticities for inexperienced drivers are around -1 for each of the three samples of cabdrivers we used in our study. The results are robust to outliers and different specifications. There are several possible explanations for these negative elasticities. Some explanations can be ruled out, but others require more data to evaluate.

Assuming that the alternative explanations for negative

elasticities are not correct, two major conclusions can be drawn from the finding of negative elasticities. Both conclusions point to the importance of psychological factors that are not incorporated in conventional dynamic models of labor supply.

First, it is difficult to explain negative wage elasticities with a model that has more than a one-day time horizon for decision-making. Imagine, for example, that cabdrivers had an earnings target (an idea we return to below) beyond which they derived zero marginal utility of income. If applied at the daily level, such a target would produce wage elasticities of -1 because as the wage increased on a particular day, drivers would cut back their hours proportionately to earn a daily income that just meets the target (since exceeding it adds no utility). However, if a target were applied at even a two-day level—i.e., if drivers had a two-day earnings target—estimated elasticities would be positive for a wide range of plausible specifications. Drivers would intertemporally substitute between the two days, working long hours on the first day if it turned out to be high wage day, and cutting back on hours if it were a low-wage day. Thus, for plausible income utility functions, a one-day time horizon for labor supply decisions is necessary to explain strongly negative wage elasticities.

Second, if drivers take a one-day horizon, for elasticities to be substantially negative requires the marginal utility of income to drop substantially sharply around the level of average daily income. Analytically, this property is familiar as a high degree of income risk aversion. A psychological account of the source of this high risk aversion, which came from conversations with many cabdrivers, is that drivers drive as if they have an income target; when they get near the target, the probability of quitting for the day rises sharply (as if the marginal utility of income drops a lot).¹

1. For example, Weber [1958] wrote: "... raising the piece-rates has often had the result that not more but less has been accomplished in the same time, because the worker reacted to the increase not by increasing but by decreasing the amount of his work. A man, for instance, who at the rate of 1 mark per acre mowed $2\frac{1}{2}$ acres per day and earned $2\frac{1}{2}$ marks, when the rate was raised to 1.25 marks per acre mowed, not 3 acres, as he might easily have done, thus earning 3.75 marks, but only 2 acres, so that he could still earn the $2\frac{1}{2}$ marks to which he was accustomed." In their widely used microeconomics textbook Pindyck and Rubinfeld [1989, p. 503] write about a student who has a one-summer horizon: "In real life, a backward-bending labor supply curve might apply to a college student working during the summer to earn living expenses for the school year. As soon as a target level of earnings is reached, the student stops working and allocates more time to leisure activities. An increase in the wage rate will then lead to fewer hours worked because it enables the student to reach the target level of earnings faster."

Such a target might be set at a driver's average earnings level, at some round number such as \$150, or by a simple formula such as twice the daily fee for leasing the cab.

Both the idea that cabdrivers make labor supply decisions one day at a time, and that they seem to have a target wage or aspiration level, are consistent with much other research in psychology and economics. Indeed, this research motivated us to study the behavior of cabdrivers in the first place, because the one-day targeting hypothesis predicts negative elasticities and hence, directly competes with the standard theory.

Taking one day at a time is consistent with considerable research which suggests that people "bracket" decisions narrowly, simplifying decisions by isolating them from the entire stream of decisions they are embedded in [Read and Loewenstein 1996]. For example, people are risk averse to single plays of small gambles, even though they typically face many uncorrelated small risks over time that diversify away the risk of a single play. Another example, closely related to the cabdrivers' daily decisions, is betting on horse races. Bettors seem to record the betting activity for each day in a separate "mental account" [Thaler 1985]. Since the track takes a percentage of each bet, most bettors are behind by the end of the day. Studies show that they tend to shift bets toward long shots in the last race in an attempt to "break even" on the day [McGlothlin 1956]. This implies that the behavior on a given day depends much more on the outcome of earlier bets that same day than on the outcome of bets on previous days or on expectations of future days (in violation of a life-cycle theory of betting).

Narrow bracketing of decisions can produce other decision anomalies that are not based on risk taking. For example, Read and Loewenstein [1995] conducted an experimental study of variety-seeking among trick-or-treaters on Halloween. Children who were told to take any two pieces of candy at a single house always chose two different candies. Those who chose one candy at each of two adjacent houses (from the same set of options) typically chose the same candy at each house. Normatively, the children should diversify the portfolio of candy in their bag, but in fact they only diversify the candy from a single house. Decision isolation has also been observed in some strategic situations: Johnson et al. [1996] found that subjects in a three-stage "shrinking-pie" bargaining experiment often did not bother to look ahead and find out how much the "pie" they bargained over would shrink if their first-stage offers were rejected.

The notion that drivers are averse to falling below a target income is also consistent with many other findings. There is ample evidence from psychological studies that judgments and decisions depend on a comparison of potential outcomes against some aspiration level or reference point [Helson 1964; Kahneman and Tversky 1979; Tversky and Kahneman 1991]. For bettors at the track, for example, breaking even is a significant, clear reference point. In other situations, reference points could be determined by past income or consumption (e.g., Bowman, Minehart, and Rabin [1996]), by social comparison (e.g., Duesenberry [1949]), or by expectations for the future. Regardless of what the reference points are, the general finding is that people are “loss-averse”—they dislike achieving outcomes below a reference point about twice as much as they like exceeding the reference point by the same absolute amount.²

Benartzi and Thaler [1995] use the same combination of narrow bracketing and loss aversion that we use, to explain the equity premium puzzle—the tendency for stocks to offer much higher rates of returns than bonds over almost any moderately long time interval. In their model, the equity premium compensates stockholders for the risk of suffering a loss over a short horizon. They show that if investors evaluate the returns on their portfolios once a year (taking a narrow horizon), and have a piecewise-linear utility function which is twice as steep for losses as for gains, then investors will be roughly indifferent between stocks and bonds, which justifies the large difference in expected returns. If investors took a longer horizon, or cared less about losses, they would demand a smaller equity premium. Two papers in this issue [Thaler, Tversky, Kahneman, and Schwartz 1997; Gneezy and Potters 1997] demonstrate the same effect in experiments.

There is other field evidence of narrow bracketing and loss aversion in stock trading and consumer purchases. Investors who own stocks that have lost value hold them longer than they hold “winning” stocks, before selling [Odean 1996; Weber and Camerer forthcoming]. Purchases of consumer goods like orange juice fall a lot when prices are increased, compared with how much purchases rise when prices are cut (i.e., price elasticities are asym-

2. Other applications of loss aversion include Kahneman, Knetsch, and Thaler [1990] on “endowment effects” in consumer choice and contingent valuation of nonmarket goods, Samuelson and Zeckhauser [1988] on “status quo biases,” and Bowman et al. [1996] and Shea [1995] on anomalies in savings-consumption patterns.

metric; e.g., Hardie, Johnson, and Fader [1993]). These data suggest that like the trick-or-treaters mentioned above, investors and consumers isolate single decisions—selling one stock, or buying one product—from the more general decisions about the contents of their stock portfolio or shopping cart (contrary to portfolio theories in finance, and the economic theory of consumer choice). Note that losses loom largest when decisions are isolated because, otherwise, losses on a single stock or product can be combined with gains from other decisions in a single mental account. So the assumptions of narrow bracketing and loss aversion relative to a reference point are both needed to explain these phenomena.

II. EMPIRICAL ANALYSES

In this section we use data on trip sheets of New York City cabdrivers to explore the relationship between hours that drivers choose to work each day and the average daily wage. A trip sheet is a sequential list of trips that a driver took on a given day. For each trip the driver lists the time the fare was picked up and dropped off and the amount of the fare (excluding tip). Fares are set by the Taxi and Limousine Commission (TLC). For the first period we study (1988), the fares were \$1.15 per trip plus \$.15 for each 1/5 of a mile or 60 seconds of waiting time. For the second period we study (1990 and 1994) fares were \$1.50 per trip plus \$.25 each 1/5 of a mile or 75 seconds of waiting time. In both periods a \$.50 per-trip surcharge is added between 8 PM and 6 AM.

Our data consist of three samples of trip sheets. We describe each data set briefly here, and include longer descriptions as Appendix 1. The first data set, TRIP, came from a set of 192 trip sheets from the spring of 1994. We borrowed and copied these from a fleet company. Fleet companies are organizations that own many cabs (each affixed with a medallion which is required to operate it legally). They rent these cabs for twelve-hour shifts to drivers who, in our sample period, typically paid \$76 for a day shift and \$86 for a night shift. The driver also has to fill the cab up with gas at the end of the shift (costing about \$15). Drivers get most of their fares by “cruising” and looking for passengers. (Unlike many cities, trips to the airport are relatively rare—around one trip per day on average). Drivers keep all the fares including tips. The driver is free to keep the cab out as long as he wants, up to the twelve-hour limit. Drivers who return the cab

late are fined. When a driver returns the cab, the trip sheet is stamped with the number of trips that have been recorded on the cab's meter. This can then be used to determine how carefully the driver has filled in the trip sheet.

The measure of hours worked is obtained directly from the trip sheet. It is the difference between the time that the first passenger is picked up and the time that the last passenger is dropped off. Total revenue was calculated by adding up the fares listed on the trip sheet. The average hourly wage is total revenue divided by hours worked.

Many of the trip sheets were incomplete, since the number of trips listed by the cabdriver was much fewer than the number of trips recorded by the meter. Therefore, we exclude trip sheets that listed a number of trips that deviates by more than two from the metered number. This screen leaves us with 70 trip sheets from thirteen drivers (eight of whom drive on more than one day in the sample).

The advantage of the TRIP data set is that we can use the trip sheets to measure the within-day autocorrelation in hourly earnings as well as differences in earning across days. Even though taxi fares are fixed by the TLC, earnings differ from day to day because of differences in how "busy" drivers are; that is, whether they spend most of the day with passengers in their cab, or have to spend a lot of time searching for passengers.

The second and third data sets of trip sheets were obtained from the TLC.³ The TLC periodically samples trip sheets to satisfy various demands for information about drivers and earnings (e.g., when rate increases are proposed). In these two data sets, hours and the number of driver-listed trips are obtained from the trip sheets and the number of recorded trips, fares, and miles driven is obtained from the meter.

The TLC developed a screen to discard incomplete trip sheets. To pass this screen, the number of trips on the meter must exactly match the number of trips listed by the driver, and additional criteria must also be met (see Appendix 1 for details). Because the TLC provided us with the summary measures, but not the trip sheets themselves, we are unable to create an alternative screening procedure, so we use their screened data for our analyses.

The first of the TLC data sets, TLC1, is a summary of 1723

3. See NYTLC [1991, 1992] for descriptive analyses of the NYC taxi business based on these data sets.

trip sheets, collected mostly during October 29 to November 5, 1990. This data set includes three types of drivers: daily fleet drivers, lease-drivers who lease their cabs by the week or month, and others who own a medallion-bearing cab and drive it. Most owner-drivers rent their cab out to other drivers for some shifts, imposing constraints on when and how long they can drive. Those who do not rent out their cabs can drive whenever they want.

The screened data contain 1044 trip sheets and 484 drivers (234 of whom drove more than one day in the data). The main advantages of this sample are that it includes several observations for each of many drivers and contains a range of different types of drivers.

The second TLC data set, TLC2, is a summary of 750 trip sheets, mostly from November 1–3, 1988. This data set samples owner-drivers as well as drivers from mini-fleet companies (mini-fleets usually lease cabs to drivers weekly or monthly). We discard 38 trip sheets using the TLC screen, leaving us 712 trip sheets. The main differences between TLC2 and TLC1 are that no drivers appear more than once in the data in TLC2 and the fares set by the TLC in TLC2 are slightly lower.

The analyses reported in the body of the paper use only the screened samples of trip sheets for all three data sets. Appendix 3 reports sample statistics for the screened and “screened-out” data for TRIP and TLC1. (TLC2 is not compared because so few observations are screened out.) It also replicates the basic regressions reported in the paper, including the screened-out data. No substantive conclusions are changed.

To learn about important institutional details, we conducted a phone survey of fourteen owners and managers at fleet companies that rent cabs to drivers. The average fleet in New York operates 88 cabs, so the responses roughly summarize the behavior of over a thousand drivers. The institutional details they reported help make sense of the results derived from analysis of hours and income data.

Sample Characteristics

Table I presents means, medians, and standard deviations of the key variables. Cabdrivers work about 9.5 hours per day, take between 28 and 30 trips, and collect almost \$17 per hour in revenues (excluding tips). Average hourly wage is slightly lower in the TLC2 sample because of the lower rates imposed by the TLC during that time period. The distributions of hours and hourly wages

TABLE I
SUMMARY STATISTICS

	Mean	Median	Std. dev
<u>TRIP</u> ($n = 70$)			
Hours worked	9.16	9.38	1.39
Average wage	16.91	16.20	3.21
Total revenue	152.70	154.00	24.99
# Trips listed on sheet	30.17	30.00	5.48
# Trips counted by meter	30.70	30.00	5.72
High temperature for day	75.90	76.00	8.21
Correlation log wage and log hours = $-.503$. The standard deviation of log hours is $.159$, log wage is $.183$, and log total revenue is $.172$. The within-driver standard deviation of log revenue is $.155$ and across drivers standard deviation is $.017$.			
<u>TLC1</u> ($n = 1044$)			
Hours worked	9.62	9.67	2.88
Average wage	16.64	16.31	4.36
Total revenue	154.58	154.00	45.83
# Trips counted by meter	27.88	29.00	9.15
High temperature for day	65.16	64.00	8.59
Correlation log wage and log hours = $-.391$. The standard deviation of log hours is $.263$, log wage is $.351$, and log total revenue is $.347$. The within-driver standard deviation of log revenue is $.189$ and across drivers standard deviation is $.158$.			
<u>TLC2</u> ($n = 712$)			
Hours worked	9.38	9.25	2.96
Average wage	14.70	14.71	3.20
Total revenue	133.38	137.23	40.74
# Trips counted by meter	28.62	29.00	9.41
High temperature for day	49.29	49.00	2.01
Correlation log wage and log hours = $-.269$. The standard deviation of log hours is $.382$, log wage is $.259$, and log total revenue is $.400$.			

are presented in Appendix 2. In the TRIP data the average trip duration was 9.5 minutes, and the average fare was \$5.13.

One feature of the data is that the variation in hours worked and number of trips in the TRIP sample is substantially lower—about half as large—as in the TLC1 and TLC2 samples. Recall that a key difference is that TRIP consists of only fleet drivers who rent their cabs daily, while TLC1 consists of fleet, lease, and owner-drivers and the TLC2 consists of lease and owner-drivers. Figure II below is a distribution of hours broken up by driver-type for the TLC1 data. It is clear from the histograms that the differences in variation in the key variables across data sets (see Appendix 2) are driven by the differences in driver-types across the data sets.

Wage Variability within Days and between Days

In the empirical analyses below, we estimate labor supply functions using the daily number of hours as the dependent variable and the average wage the driver earned during that day as the independent variable (both in log form). The average wage is calculated by dividing daily total revenue by daily hours.⁴ However, this assumes that the decisions drivers make regarding when to stop driving depend on the average wage during the day, rather than fluctuations of the wage rate during the day.

Within-day fluctuations are important to consider because negatively autocorrelated intraday hourly wage rates could lead drivers who are actually driving according to the predictions of the standard theory to behave as if they were violating it. If autocorrelation is negative, on a day with a high wage early in the day, drivers will (rationally) quit early because high hourly wages are likely to be followed by low-wage hours. Conversely, on a day with low early wages, drivers will drive long hours expecting the wage to rise. If hourly autocorrelations are zero or positive, however, we can rule out this alternative explanation (unless drivers think the autocorrelation is negative when it is not).

To investigate how the hourly rate varied within the day, we used the trip-by-trip data available in the TRIP sample. Days were broken into hours, and the median hourly wage for all drivers during that day and hour were calculated. We then regressed the median hourly wage (across drivers driving that hour) on the previous hour's median wage, estimating an autocorrelation of .493 ($se = .092$).⁵ The second-order autocorrelation is even higher (.578), and the third- and fourth-order autocorrelations are also positive and significant. When hourly wage is regressed on two previous lags, both coefficients are greater than .40 and are significantly different from zero. If we divide days into first and second halves, the correlation between median wages in the two halves is .40.⁶ The patterns imply that when a day starts out as

4. This is similar to the method traditionally used in the labor supply literature—dividing yearly (or monthly) income by yearly (or monthly) hours to get the wage rate.

5. Weighting the median observations by the number of drivers used to construct that observation did not change the standard error, and changed the estimate only slightly, to .512.

6. The p -value of .15 for this correlation is higher than conventional levels, but note that the sample size for this correlation is only fourteen (because each observation is a day).

a high wage day, it will probably continue to be a high wage day. The fleet managers surveyed weakly agreed⁷ with these patterns, saying the within-day autocorrelation is positive or zero (none said it was negative).

Wages are significantly different across days ($p < .0001$ for TRIP and TLC1; too few days to permit a test for TLC2). The median (across drivers) of the average hourly wage for a day ranges from a low of \$13.93 to a high of \$20.62 in the TRIP data and a low of \$15.56 to a high of \$19.35 in the TLC1 data. Wages are also virtually uncorrelated across days. When we ran regressions of the mean or median wage on day t on the mean or median wage on day $t - 1$, the regression coefficient was $-.07$ and insignificant ($p > .7$).

Since wages are virtually uncorrelated across days, and fairly stable within days, they are ideal for calculating the labor supply response to a transitory change in wage.

Wage Elasticities

For each of the three data sets, we calculate the simple correlation between (log) hours and (log) wages. These statistics, provided in Table I, are $-.503$, $-.391$, and $-.269$. Figure I shows scatterplots of log hours and log wages in the three samples, which corroborate the negative correlations. Regressions of (log) hours on (log) wages are provided in Table II for the three data sets. TRIP and TLC1 include multiple observations for each driver, so either the standard errors are corrected to account for the panel nature of the data, or driver fixed effects are included.⁸

We also include two weather measures in the regression: the high temperature for the day and a dummy variable for rain (which does not vary in TLC1, since it did not rain in that time period). These variables control for shifts in labor supply that occur if driving on a rainy day is more difficult and driving on a

7. Fleet managers were asked whether "a driver who made more money than average in the first half of a shift" was likely to have a second half which was better than average (three agreed), worse than average (zero), or about the same as average (six). Expressing the target-income hypothesis, two fleet managers spontaneously said the second half earning were irrelevant "because drivers will quit early."

8. The fixed effects control for the possibility that drivers vary systematically in their work hours or their target income (see Section III), independent of the wage. There are not enough observations per driver to allow drivers' elasticities to vary. However, we estimated some individual-driver regressions using the TRIP sample for those drivers with many daily observations. Most of the wage elasticities were significantly negative.

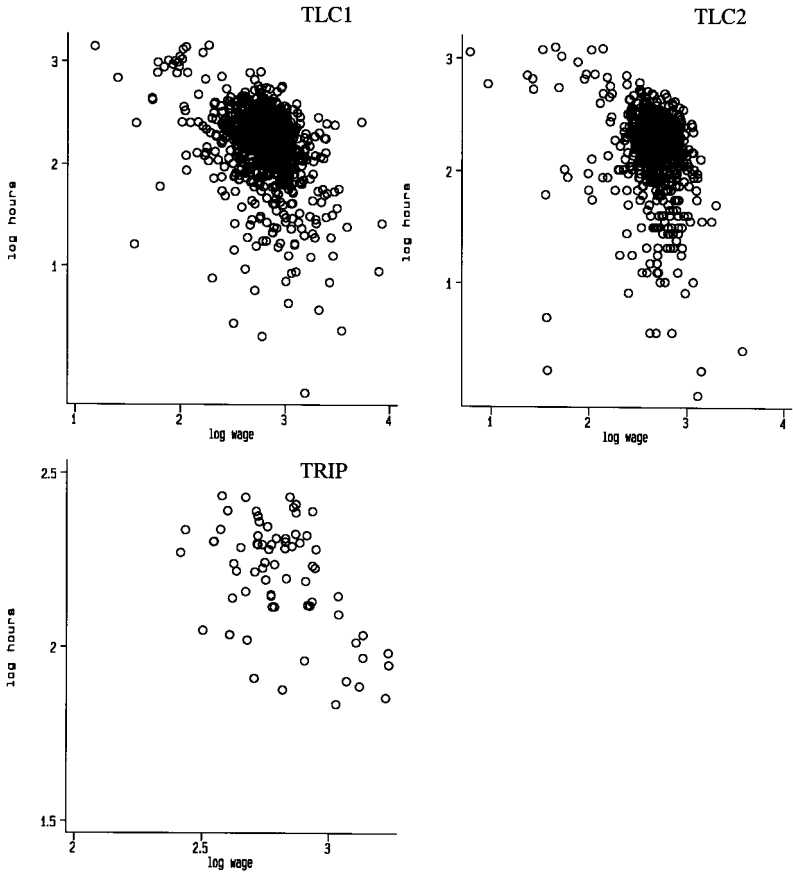


FIGURE I
Hours-Wage Relationships

warmer day has a higher opportunity cost (perhaps because forgone leisure is more pleasurable). Also included is a dummy variable for the shift driven and a dummy variable for a weekday versus weekend day (although all shifts are during the week in the TLC2 data).⁹

9. Shifts are described in detail in Appendix 1. Briefly, in the TRIP and TLC2 samples, the dummy indicates night shift (versus day or afternoon) and in the TLC1 sample there are two shift dummy variables (night and day, versus "other") reflecting the greater heterogeneity of driving arrangements in this sample. The estimates are changed very little if no shift designations are used. No additional

TABLE II
OLS LOG HOURS WORKED EQUATIONS

Sample	TRIP		TLC1		TLC2
Log hourly wage	-.411 (.169)	-.186 (.129)	-.501 (.063)	-.618 (.051)	-.355 (.051)
High temperature	.000 (.002)	-.000 (.002)	.001 (.002)	.002 (.002)	-.021 (.007)
Shift during week	-.057 (.019)	-.047 (.033)	-.004 (.035)	.030 (.042)	—
Rain	.002 (.035)	.015 (.035)	—	—	-.150 (.062)
Night shift dummy	.048 (.053)	-.049 (.049)	-.127 (.034)	-.294 (.047)	-.253 (.038)
Day shift dummy	—	—	.000 (.028)	.053 (.045)	—
Fixed effects	No	Yes	No	Yes	No
Adjusted R^2	.243	.484	.175	.318	.146
Sample size	70	65	1044	794	712
Number of drivers	13	8	484	234	712

Dependent variable is the log of hours worked. Standard errors are in parentheses and are corrected for the nonfixed effects estimates in columns 1 and 3 to account for the panel structure of the data. Explanatory variables are described in Appendix 1.

In TRIP the wage elasticities depend substantially on whether or not driver fixed effects are included in the model. In the first column (no driver fixed effects), the estimated wage elasticity is $-.411$ and is significantly different from zero. Including driver fixed effects, which are jointly significant, lowers the estimated elasticity to $-.186$, which is no longer significantly different from zero.¹⁰

improvement in fit is obtained if day of the week dummy variables are included rather than a weekday versus weekend dummy variable.

10. One way to make use of the large amount of screened-out data in TRIP is to impute missing hours for the incomplete trip sheets, by multiplying the driver-listed hours by the ratio of meter-recorded trips to the number of driver-listed trips. For example, if a driver listed only 16 trips in 5 hours of driving, but the meter recorded 24 trips, this method would impute 7.5 total hours of driving. This method yields OLS estimates of $-.549$ ($se = .156$, $n = 162$) and $-.276$ ($se = .071$, $n = 158$) for the TRIP sample without and with fixed effects. These estimates are slightly more negative and more precisely estimated than those for the screened sample, reported in Table II. Another method of imputation assumes that drivers stopped filling out their trip sheets when they got busy (so that the average wage during the missing hours is higher than during the listed hours). This method scales up the number of hours by a factor that is less than the ratio of meter-recorded trips to driver-recorded trips (since it assumes the hours-per-trip is smaller for the missing trips) and actually makes the estimates even more negative.

In the TLC1 and TLC2 samples, elasticities are strongly negative, more precisely estimated, and quite robust to including fixed effects (the estimates range from $-.355$ to $-.618$). In all three samples, analyses that reduce outlier influence (such as median regression) indicate that the results are not sensitive to outliers.

The difference between the wage elasticities in these samples and the fixed-effects estimate in the TRIP sample is a bit curious. However, recall that TRIP consists entirely of fleet drivers (who pay daily) while the TLC samples also includes weekly and monthly lease-drivers, and owner-drivers. Lease-drivers and owner-drivers have more flexibility in the number of hours they drive (since fleet drivers are constrained to drive no more than twelve hours). We report below (in Table V) that elasticities for the fleet drivers are substantially smaller in magnitude (less negative) than for lease- and owner-drivers. The TRIP sample, which is all fleet drivers, reflects this compositional difference in driver types.

Controlling for Measurement Error

Measurement error is a pervasive concern in studies of labor supply. Although the data on hours come from trip sheets rather than from memory, they may include recording errors.¹¹ If there is “classical” measurement error in hours (the errors are “white noise” and are uncorrelated with hours [Maddala 1992]), this leads to a predictable bias in the wage elasticity. Since the average hourly wage is computed by dividing daily revenue by reported hours, overstated hours will produce high hours-low wage observations and understated hours produce low hours-high wage observations, creating spuriously negative elasticities. This bias can be eliminated if we can find an instrument for wage that is uncorrelated with the measurement error in hours. We use summary statistics of the distribution of hourly wages of other drivers that drove on the same day and shift (the 25th, 50th, and

11. Measurement error in income may also occur due to the omission of tips. Suppose that true income equals income from fares times $(1 + t)$, where t is the average tip percentage. If $1 + t$ is independent of fares, when taking logs the measurement error will be independent of measured income, causing no bias in the wage elasticity. (Sherwin Rosen suggested that on high-demand days, frustrated passengers searching for cabs might add voluntary surcharges; e.g., waving money at cabdrivers. This would cause a bias: the highest hourly wages would be most understated; and the true elasticity would be even more negative than we estimate it to be.)

75th percentiles) as instruments for own wage. These instruments, that summarize the “wage” for the day, should be uncorrelated with a particular driver’s measurement error.

The first-stage regression of average wage on the 25th, 50th, and 75th percentiles of the other driver wage distribution is presented in the bottom half of Table III. The joint test of the null hypothesis that all coefficients are zero can be easily rejected. The top half of Table III reports estimated elasticities using these instruments, and including weather, shift, and weekday dummies as explanatory variables. The elasticities are less precisely estimated using the instrumental variables (as is common), but are even more negative. For TRIP and TLC1, estimates with and without fixed effects are included. The basic findings from Table II are maintained in the IV estimation: elasticities are negative and significantly different from zero, except in the TRIP sample when fixed effects are included.

The results in Table III are quite robust with respect to various specifications. We also estimated specifications that used as instruments: 1) the mean wage of other drivers on the same day and shift; 2) the 25th, 50th, and 75th percentiles of the other drivers on that day wage distribution, although not broken down by shift; and 3) percent of miles driven that are “live” (during which a passenger is in the cab). The basic results are unchanged when these other specifications are used.¹²

How Do Elasticities Vary with Experience?

Drivers may learn over time that driving more on high wage days and less on low wage days provides more income and more leisure. If so, the labor supply curve of experienced drivers would have a more positive wage elasticity than that of inexperienced drivers. There are good measures of driver experience in these data sets. In the TLC data sets, the TLC separated drivers into experience groups: for TLC1, those with greater or less than four years of experience and in TLC2, those with greater or less than three years of experience. These group measures are absent in the TRIP data. However, cabdriver licenses are issued with six-

12. In unreported regressions, we also tried using daily subway ridership as an instrument for wage. However, this instrument did not predict wages well in the first stage. We tried to obtain data on hotel occupancy or convention attendance but could not. Note that conventions are an ideal instrument because they are most likely to shift demand without also shifting the disutility of effort (and hence, the supply curve).

TABLE III
IV LOG HOURS WORKED EQUATIONS

Sample	TRIP		TLC1		TLC2
Log hourly wage	-.319 (.298)	.005 (.273)	-1.313 (.236)	-.926 (.259)	-.975 (.478)
High temperature	-.000 (.002)	-.001 (.002)	.002 (.002)	.002 (.002)	-.022 (.007)
Shift during week	-.054 (.023)	-.041 (.035)	-.016 (.042)	.028 (.044)	—
Rain	-.007 (.042)	-.001 (.041)	—	—	-.130 (.070)
Night shift dummy	.059 (.057)	-.036 (.053)	-.088 (.040)	-.242 (.064)	-.202 (.057)
Day shift dummy	—	—	-.030 (.038)	.068 (.048)	—
Fixed effects	No	Yes	No	Yes	No
Sample size	70	65	1044	794	712
Number of drivers	13	8	484	234	712

Dependent variable is the log of hours worked. Standard errors are in parentheses and are corrected for the nonfixed effects estimates in columns 1 and 3 to account for the panel structure of the data. Instruments for the log hourly wage include the summary statistics of the distribution of hourly (log) wages of other drivers on the same day and shift (the 25th, 50th, and 75th percentiles).

	First-stage regressions				
Median	.316 (.225)	.026 (.188)	-.385 (.394)	-.276 (.467)	1.292 (4.281)
25th percentile	.323 (.160)	.287 (.126)	.693 (.241)	.469 (.332)	-.373 (3.516)
75th percentile	.399 (.171)	.289 (.149)	.614 (.242)	.688 (.292)	.479 (1.699)
Adjusted R^2	.374	.642	.056	.206	.019
P -value for F -test of instruments for wage	.000	.004	.000	.000	.020

Dependent variable is the log of average hourly wage. Standard errors are in parentheses. Regressions also include weather and shift characteristics (dummy variable for rain, high temperature during the day, dummy variable for shift on a weekday, and time of shift dummy variables) as explanatory variables.

digit numbers (called hack numbers), in chronological order, so that lower numbers correspond to drivers who obtained their licenses earlier.¹³ Using their license numbers, we use a median

13. This is not a perfect measure of actual experience because, for example, some drivers with old licenses may only drive sporadically. However, licenses must be renewed each year so that sufficiently inactive drivers probably let their licenses expire.

TABLE IV
IV LOG HOURS WORKED EQUATIONS BY DRIVER EXPERIENCE LEVEL

Sample	TRIP		TLC1		TLC2	
	Low	High	Low	High	Low	High
Experience level	Low	High	Low	High	Low	High
Log hourly wage	-.841 (.290)	.613 (.357)	-.559 (.406)	-1.243 (.333)	-1.308 (.738)	2.220 (1.942)
Fixed effects	Yes	Yes	Yes	Yes	No	No
Sample size	26	39	319	458	320	375
<i>P</i> -value for difference in wage elasticity	.030		.666		.058	

Dependent variable is the log of hours worked. Standard errors are in parentheses. Regressions also include weather and shift characteristics (dummy variable for rain, high temperature during the day, dummy variable for shift on a weekday, and time of shift dummy variables) as explanatory variables. Instruments for the log hourly wage include the summary statistics of the distribution of hourly (log) wages of other drivers on the same day and shift (the 25th, 50th, and 75th percentiles).

split to divide drivers into low- and high-experience subsamples for the TRIP data.¹⁴

Table IV presents the wage elasticities estimated separately for low- and high-experience drivers. All regressions use instrumental variables, and all include fixed effects (except, of course for TLC2). In all three samples, the low-experience elasticity is strongly negative, generally close to -1 . The wage elasticity of the high-experience group is significantly larger in magnitude, for the TRIP and TLC2 samples ($p = .030$ and $.058$, respectively).¹⁵

How Do Elasticities Vary with Payment Structure?

The way drivers pay for their cabs might affect their responsiveness of hours to wages if, for example, the payment structure affects the horizon over which they plan. Alternatively, it might affect the degree to which they can significantly vary hours across days. The TLC1 sample contains data from three types of payment schemes: daily rental (fleet cabs), weekly or monthly rental (lease cabs), or owned. Table V presents elasticity estimates in

14. The number of observations in the low- and high-experienced samples for the TRIP data are not equal because the median split is done on drivers, not trip sheets, and there are different sample sizes for each driver.

15. An alternative approach is to use the median wage directly as a regressor, skipping the first-stage regression. This lowers the adjusted R^2 substantially (as is expected) but does not alter the sign or magnitude of the estimates reported in Table III systematically (TRIP and TLC2 estimates become more negative and TLC1 estimates become less negative). The large estimate and standard error on the high-experience TLC2 elasticity reported in Table IV do become smaller ($-.135$ and $.968$, respectively), but that does not change the conclusion that experience makes elasticities less negative.

TABLE V
 IV LOG HOURS WORKED EQUATIONS BY PAYMENT STRUCTURE TLC1 DATA

Type of cab	Fleet	Lease	Owned
Log hourly wage	-.197 (.252)	-.978 (.365)	-.867 (.487)
Fixed effects	Yes	Yes	Yes
Sample size	150	339	305

Dependent variable is the log of hours worked. Standard errors are in parentheses. Regressions also include weather and shift characteristics (dummy variable for rain, high temperature during the day, dummy variable for shift on a weekday, and time of shift dummy variables) as explanatory variables. Instruments for the log hourly wage include the summary statistics of the distribution of hourly (log) wages of other drivers on the same day and shift (the 25th, 50th, and 75th percentiles). Fleet cabs are rented daily; leased cabs are rented by the week or month; and owned cabs are owned by the drivers.

the three payment categories from the TLC1 sample. All regressions are estimated using instrumental variables and include driver-fixed effects.

All wage elasticities in Table V are negative. The elasticity that is smallest in magnitude, for fleet drivers, is not significantly different from zero. The lease- and owner-driver wage elasticities are approximately $-.9$ and are significantly different from zero. Part of the explanation for the lower elasticity for fleet drivers is a technical one. Since they are constrained to drive no more than twelve hours, the dependent variable is truncated, biasing the slope coefficient toward zero.

Could Drivers Earn More by Driving Differently?

One can simulate how income would change if drivers changed their driving behavior. Using the TLC1 data, we take the 234 drivers who had two or more days of data in our sample. For a specific driver i , call the hours and hourly wages on a specific day t , h_{it} and W_{it} , respectively, and call driver i 's mean hours over all the days in the sample h_i . By construction, the driver's actual total wages earned in our sample is $\sum_i h_{it} W_{it}$.

One comparison is to ask how much money that driver would have earned if he had driven h_i hours every day rather than varying the number of hours (i.e., if his labor supply curve of hours against wages was flat). Call this answer "fixed-hours earnings" (FHE), $\sum_i h_i W_{it}$.

Is FHE greater than actual earnings? We know that, on average, h_{it} and w_{it} are negatively correlated so that the difference between FHE and actual earnings will be positive in general. In fact, drivers would increase their net earnings by 5.0 percent on average (std.error = .04 percent) if they drove the same number

of hours (h_i) every day, rather than varying their hours every day. If we exclude drivers who would earn less by driving fixed hours (because their wage elasticity is positive), the improvement in earnings would average 7.8 percent. And note that if leisure utility is concave, fixed-hours driving will improve overall leisure utility too.

These increases in income arise from following the simplest possible advice—drive a constant number of hours each day. Suppose instead that we hold each driver's average hours fixed, but reallocate hours across days as if the wage elasticity was +1. Then the average increase in net income across all drivers is 10 percent. Across drivers who gain, the average increase is 15.6 percent.¹⁶

III. WHY MIGHT LABOR SUPPLY BE DOWNWARD-SLOPING?

Our results lend support to the common finding that elasticities are not strongly positive for temporary changes in wages. Indeed, wage elasticities estimated with instrumental variables are significantly negative in two out of three samples. Two additional effects we observe are that wage elasticities are significantly higher for experienced drivers in two of three samples, and wage elasticities are significantly more negative for lease- and owner-drivers than for fleet drivers. These two additional regularities, along with other patterns in the data, as well as information gleaned from our telephone survey of fleet managers, allow us to evaluate alternative explanations for the observed negative elasticities. We begin by discussing the explanation we favor, then evaluate three others suggested by colleagues, referees, and contentious friends.

Daily Income Targeting

As explained in the Introduction, one possible explanation for the negative hours elasticities is that cabdrivers take a one-day horizon, and set a target (or target range) and quit when the

16. Still another figure one can compute is the optimal reallocation of hours to earn the largest possible wage total. This calculation will yield a wage elasticity substantially larger than the +1 value used above. But such a calculation will require drivers to work 12-hour shifts (or longer for fleet and owner-drivers with 24-hour shifts) on all the high-wage days, and quit very early on low-wage days. This pattern will raise variation into leisure hours (which will lower overall utility if variation in leisure is undesirable). Without some accounting for the utility of forgone leisure, simply knowing how much more income the drivers would earn is not of much interest.

target is reached. This decision rule can be modeled by marginal utility of income declining substantially around the average daily income level. This explanation was suggested to us by several drivers in New York City, and also rang true to many of the fleet managers we surveyed. They were asked to choose which one of three sentences “best describes how many hours cabdrivers drive each day?” Six fleet managers chose “Drive until they make a certain amount of money.” Five chose the response “Fixed hours.” Only one chose the intertemporal substitution response “drive a lot when doing well; quit early on a bad day.” (One manager said “all of the above; it depends on the driver.”)

While daily income targeting may seem ad hoc to labor economists, it is, as we discussed in the Introduction, consistent with general principles of decision-making that have been observed in many other domains. In fact, the theory we use here is very similar to that used by Benartzi and Thaler [1995] in their paper about the equity premium puzzle and is implicit in evidence of disposition effects in stock trading and asymmetric price elasticities in consumer brand choice.

A utility function for daily income with a target reference point could result from various underlying psychological processes. For example, targeting is a simple decision rule: it requires drivers to keep track only of the income they have earned. This is computationally easier than tracking the ongoing balance of forgone leisure utility and marginal income utility—which depends on expected future wages—required for optimal intertemporal substitution. Working a fixed number of hours—“hours-targeting”—is equally simple, but drivers (especially inexperienced ones) may not realize that this alternative rule generates more income and more leisure. Note that a weekly or monthly earning target is much more difficult to implement because a driver would need to decide how much to earn on each day (given the wage, opportunity cost of time, etc., on that day). A daily earnings target produces a much simpler rule: simply drive until one earns the target.

Daily targets can also serve a second purpose: like many mental accounts, they help mitigate self-control problems (see Shefrin and Thaler [1992]).¹⁷ There are two kinds of self-control

17. The use of a short horizon and income target to avoid temptation suggests that these features can be thought of as a self-imposed liquidity constraint, but could also be empirically distinguished from liquidity constraint imposed by limited wealth and borrowing power.

problems drivers might face. First, driving a cab is tedious and tiring and, unlike many jobs, work hours are not rigidly set. Drivers are free to quit any time they want. A daily income goal, like an author imposing a daily goal of written pages, establishes an output-based guideline of when to quit. A weekly or monthly target would leave open the temptation to make up for today's shortfall tomorrow, or next week, and so on, in an endless cycle.

Drivers could also keep themselves from quitting too early by setting daily targets that vary positively with wages early in the day (i.e., they plan to work longer hours after a few good hours and allow themselves to quit early after a few bad hours). Such a wage-dependent targeting rule creates intertemporal substitution but also creates a second self-control problem: drivers must save the windfall of cash they earn from driving long hours on a high-wage day so they can afford to quit early on low-wage days. But a drive home through Manhattan with \$200–\$300 in cash from a good day could be an obstacle course of temptations for many drivers. Given these two self-control problems, substituting over a weekly or monthly horizon may be too difficult, so daily targeting results. Of course, like most self-control strategies, it yields less income and leisure than a person with perfect self-control would earn.

A strong form of the target income hypothesis, in which the target is constant across days, and is the same for all drivers, can be easily rejected. This hypothesis predicts that daily income should not vary much across days, but it clearly does (see Table I). And the fact that (log) daily income has more variance within-drivers than across-drivers (see Table I again) implies that targets vary more across days than across drivers. The constant-target hypothesis also predicts the log hours-log wage relation will be linear, but adding a quadratic term improves fit significantly.

While the constant-target hypothesis can be rejected, income targeting in some form is useful for explaining two features of the data. First, for drivers with a one-day horizon and additively separable income and leisure utility, income utility must be quite concave around the average income level to explain elasticities as extremely negative as -1 , which are evident in the inexperienced-driver regressions (Table IV).¹⁸ Strong concavity is, of

18. Assume a one-day horizon, no nonwage income, wage = w , hours = h , income $y = hw$, and leisure $L = 24 - h$, and an additively separable utility func-

course, a possible feature of any utility function. The idea that workers are “loss-averse” around an income target (they dislike falling short much more than they like exceeding it) is one simple explanation of extreme concavity that is consistent with much other evidence.

Second, there are fewer low-hours days among fleet drivers who pay daily (in the TLC1 sample, see Figure II). The reluctance to work short days is consistent with the hypothesis that fleet drivers use their daily lease fee as one reference point and are particularly reluctant to quit before reaching it (compared with lease-drivers who pay weekly or monthly, and owner-drivers).

The daily income-target hypothesis also seems to account for the effect of experience rather naturally; experienced drivers who have larger elasticities either learn over time to take a longer horizon (and to resist the temptations of quitting early and squandering cash from good days), or to adopt the simple rule of driving a fixed number of hours each day. (Similarly, we suspect that experienced gamblers are less likely to allow within-day outcomes to influence their subsequent choices. They learn “not to count the money while they’re sitting at the table.”) Alternatively, some drivers may just lack these qualities. They will have less leisure and income and will be selected out of the experienced-driver pool. Either way, experienced drivers will have more positive wage elasticities.

Liquidity Constraints

Negative elasticities could occur because cabdrivers face strongly binding liquidity constraints. Liquidity-constrained drivers who must earn a certain amount of money each day must drive long hours when wages are low. This explanation seems unlikely for two reasons.

First, according to our fleet manager survey, almost all lease-drivers pay their weekly or monthly fees in advance. Most of the

tion $v(y) + u(L)$ with $v(\cdot)$ and $u(\cdot)$ both concave. Assuming workers maximize utility and differentiating gives the elasticity equation $(dh/dw)(w/h) = (1 - yr(y))/(yr(y) + hr(L))$, where $r(y) = -v''(y)/v'(y)$ and $r(L) = -u''(L)/u'(L)$ are risk-aversion coefficients. For $u(L)$ concave ($r(L) > 0$), the elasticity becomes negative for $r(y) > 1/y$ (e.g., more concave than log utility). The elasticity becomes increasingly negative as $r(y)$ gets larger, but does not reach -1 unless $r(y)$ become infinite (corresponding to a kink at the income target reference point; cf. Bowman et al. [1996]). If leisure utility is convex, or if leisure and daily income are strong complements, then it is easier to generate negative elasticities (then a wage increase raises income, holding hours fixed, which triggers an increase in leisure utility and causes an optimizing worker to cut hours and consume more leisure).

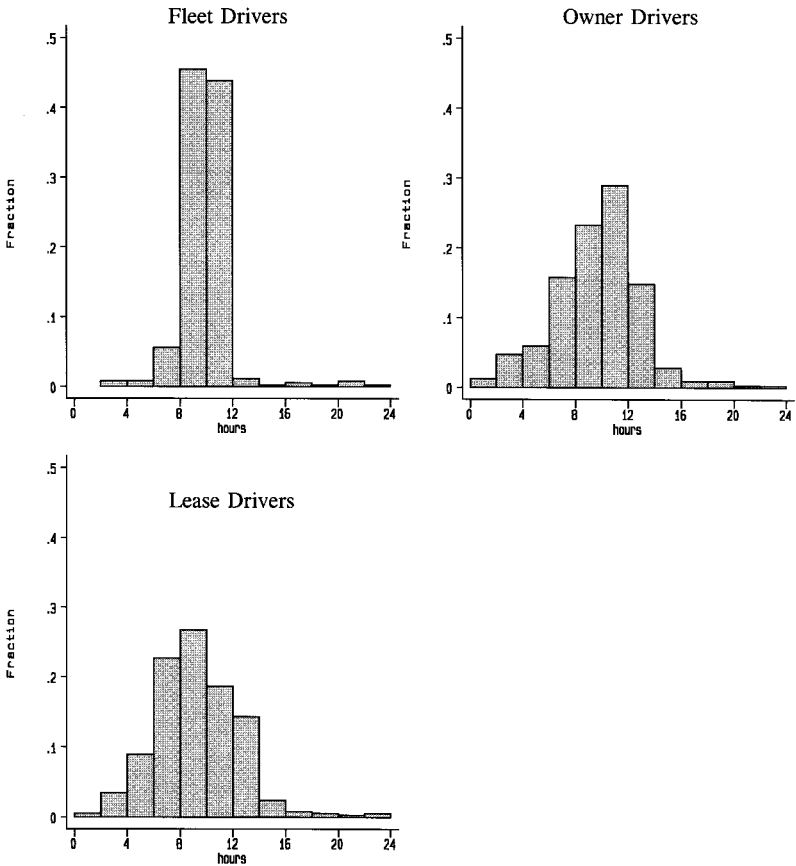


FIGURE II
Histograms of Hours Worked by Driver Ownership Class
(TLC1 only)

fleet drivers pay at the end of the day but most fleet managers said drivers could sometimes pay later. Since lease-drivers pay in advance and fleet drivers can pay late, most drivers do not need to drive long hours on low-wage days to scrape together enough cash to pay the lease fee immediately at the end of their shift.

Second, the liquidity constraint argument implies that people who are not liquidity constrained—such as those with substantial wealth—should not display negative elasticities. At the time the data were generated, cab medallions were worth

about \$130,000, so owner-drivers presumably have substantially more wealth or borrowing power than nonowners. Therefore, medallion ownership is a weak proxy for wealth. If daily liquidity constraints are responsible for the negative wage elasticities, drivers who own their cabs would have larger elasticities than drivers who rent from fleets or lease. The empirical results in Table V show the opposite.

Breaks

The trip sheets used to measure work hours do not distinguish between idle time spent searching for fares, and conscious breaks that might be considered leisure. If drivers are taking lots of breaks on low-wage days, for example, that could explain why they appear to work longer hours on those days than on high-wage days. Perhaps if we could subtract these leisure breaks from hours worked, the true wage elasticity would be more positive.

We do not have good data on the amount of break time drivers take, but there are three reasons to think self-administered breaks do not explain all three regularities. First, in early analysis using the TRIP sample, breaks of more than 30 minutes were removed when calculating hours. The results were similar to those reported here. Second, various assumptions about how the amount of break time varies with wages help bound the effect that excluding breaks would have. For a plausible range of assumptions, the true wage elasticity will not be positive if the measured elasticity is negative.¹⁹ Third, to explain the increased elasticities of experienced drivers requires the assumption that

19. Call measured hours m , true (unobserved) hours t , and breaks b . By definition $t = m - b$. Taking derivatives, $dt/dw = dm/dw - db/dw$. Noting that dm/dw appears to be negative in our analyses, we can ask how dt/dw would change for plausible values of a break response db/dw . If breaks are fixed in length across days (e.g., breaks are taken for meals or coffee), then $db/dw = 0$, and $dt/dw = dm/dw < 0$. Another possibility is that breaks respond to wages like nonwork leisure does. Defining nonwork leisure $L = 24 - m$, if $db/dw = dL/dw$, then $db/dw = dm/dw$, so that $dt/dw = 2(dm/dw) < 0$. The opposite possibility is that breaks and nonwork leisure are perfect substitutes (drivers do not care whether they take breaks on the job or after work at home), and unresponsive to wage, so $db/dw = -dL/dw$. Then $db/dw = -dm/dw$ so $dt/dw = 0$. These three simple assumptions show that for values of db/dw in the interval $[-|dL/dw|, |dL/dw|]$, dt/dw remains negative or zero. For dt/dw to be positive requires that breaks respond more strongly to wages than leisure L does, and in the opposite direction: drivers must really like taking breaks on slow (low-wage) days and dislike them on busy days, though they exhibit the opposite pattern of leisure preferences. This is conceivable (and could be tested with better data), but no more plausible than the other three assumptions, which produce nonpositive dt/dw .

the length of breaks they take responds less strongly to wages than for inexperienced drivers, which is possible but not supported by any particular intuition or evidence.

Increasing Disutility of Effort

Since fares are fixed within each sample, a high-wage day is a busy day in which a driver picked up many fares or drove them further. Drivers may get tired faster on these high-wage days and quit earlier due to fatigue.

The hypothesis implicit in this explanation is not just that driving with a passenger in your cab is hard work, but that carrying a passenger is harder work than searching for one. Almost all of the managers in our survey said the opposite. The fleet managers were asked to compare two hypothetical drivers: (A) "a driver who worked 10 hours, found fares very quickly, drove 30 trips and spent little time cruising looking for passengers" and (B), "a driver who worked 10 hours, drove 20 trips and had a hard time finding fares so he spent a lot of time cruising looking for passengers." Ten fleet managers said the 20-trip driver would be "more tired at the end of the day." Only one said the busy 30-trip driver would be more tired. (Two managers said the two drivers would be equally tired or did not know.) This makes sense, given the logistics of searching for passengers in Manhattan. Driving to a specific destination probably requires less attention than driving while searching for a potential passenger who is trying to hail a cab, on either side of the street, and preparing to swerve across traffic to reach the passenger.

The earning-money-is-tiring hypothesis also does not easily explain the effect of experience, unless one assumes that inexperienced drivers get relatively more tired carrying passengers and experienced drivers get relatively more tired searching for passengers. The opposite effect could easily be true: if experienced drivers learn the easiest places to find fares, then searching for passengers becomes relatively less tiring for them.

Participation

The hours equation is estimated using only days on which cabdrivers worked positive hours. If unobserved factors affected drivers' decisions about whether to work at all (or "participate") and those factors also affected their hours decisions, the wage elasticity will be biased [Heckman 1979]. The sign of the bias will be opposite of the sign of the correlation between the error terms

in the hours and participation equations. If unobserved shocks to participation and hours are positively correlated, for example, the wage elasticity will be downward-biased. One way to control for this "selection bias" is to collect data on participation. Unfortunately, we do not have these data. However, there are several reasons to think that selection bias is not severe enough to explain the substantial negative elasticities. First, including driver fixed effects in the hours equation mitigates the omitted variable problem that leads to the correlation between the hours and participation error terms. Second, in the survey eleven of the fourteen fleet managers said drivers usually have a regular schedule of shifts each week. When drivers skip days, about half the fleet managers said those drivers had to pay their fees anyway or suffered some penalty so they have a large incentive to stick to their schedule. Also, a driver cannot always participate on an unscheduled day even if he decides to. Cabs are not always available because medallion owners tightly schedule them, to maximize the lease fees they collect. While owner-drivers are not strictly subject to a regular schedule, most of them rent their cabs to another driver or two; the remaining days effectively constitute a regular schedule for themselves. Because of the regularity in the drivers' schedules, there is not that much variation in unobserved factors that affect participation, and there should be little selection bias.

IV. DISCUSSION AND CONCLUSIONS

Dynamic theories of labor supply predict a positive labor supply response to transitory fluctuations in wages. Previous studies have not been able to measure this elasticity precisely, and the measured sign is often negative, contradicting the theoretical prediction. These analyses, however, have been plagued by a wide variety of estimation problems.

Many of these estimation problems are avoided by estimating labor supply functions for taxi drivers. Drivers have flexible self-determined work hours and face wages that are highly correlated within days, but only weakly correlated between days, (so fluctuations are transitory). The fact that our analyses yield negative wage elasticities suggests that elasticities of intertemporal substitution around zero (or at least, not strongly positive) may represent a real behavioral regularity. Further support for this assertion comes from analyses of labor supply of farmers [Berg 1961; Orde-Brown 1946] and self-employed proprietors

[Wales 1973] who, like cabdrivers, set their own hours and often have negative measured wage elasticities. These data suggest that it may be worthwhile to search for negative wage elasticities in other jobs in which workers pay a fixed fee to work, earn variable wages, and set their own work hours—such as fishing, some kinds of sales, and panhandling.

Of course, cabdrivers, farmers, and small-business proprietors are not representative of the working population. Besides some demographic differences, all three groups have self-selected onto occupations with low variable wages, long hours and (in the case of farmers and cabdrivers) relatively high rates of accidents and fatalities. However, there is no reason to think their planning horizons are uniquely short. Indeed, many cabdrivers are recent immigrants who, by immigrating, are effectively making long-term investments in economic and educational opportunity for themselves and their children.

Because evidence of negative labor supply responses to transitory wage changes is so much at odds with conventional economic wisdom, these results should be treated with caution. Further analyses need to be conducted with other data sets (as in Mulligan [1995]) before reaching the conclusion that negative wage elasticities are more than an artifact of measurement or the special circumstances of cabdrivers. If replicated in further analyses, however, evidence of negative wage elasticities calls into question the validity of the life-cycle approach to labor supply.

APPENDIX 1: DESCRIPTION OF DATA SETS

Trip Sheet Data

Data Set 1: TRIP

We collected 192 trip sheets from a fleet company in New York City that rents cabs daily to drivers. This sample consists of 27 cabdrivers who drove during the days April 24, 1994, to May 14, 1994. A trip sheet is a sequential list of trips that a driver took on a given day. For each trip the driver lists the time the fare was picked up and dropped off and the amount of the fare (excluding tip). The company uses these trip sheets for insurance purposes (they are not used for taxes). When a driver returns the cab, the trip sheet is stamped with the number of trips that have been recorded by the meter in the cab.

Our measure of hours worked is obtained directly from the trip sheet. It is the difference between the time that the first passenger is picked up and the time that the last passenger is dropped off. We calculated total revenue by adding up the fares listed on the trip sheet. Average hourly wage is total revenue divided by hours worked.

Not all trip sheets we obtained were complete, because the number of trips listed by the cabdriver is sometimes much less than the number of trips recorded by the meter. Because we are calculating hours and total daily revenue from the trip sheet, we need to screen out incomplete trip sheets (because hours and revenues will be too low). We decided to use only trip sheets where the number of trips listed by the driver is within two of the number of trips stamped by the meter. Using this to screen the trip sheets, we are left with 70 trip sheets and 13 drivers. Eight of these drivers appear more than once in the screened data.

There are important differences between the data retained in the screened sample and those that are not used in the analysis (the screened-out data). In Appendix 3 we provide summary statistics for the key variables for both samples. As expected, the number of trips listed by the driver in the screened sample is much greater than in the screened-out sample. This causes hours worked in the screened sample to be greater than in the screened-out sample. However, the average wage (for the trips listed) does not differ between the two samples. This is some evidence that whether or not the cabdriver fills out the trip sheet completely is not related to how "busy" the cabdriver is. In Appendix 4 we also present the basic regressions from the paper when the entire sample of trip sheets is used rather than only the screened sample.

For our screened sample, drivers either worked the afternoon or evening shift. We defined the afternoon shift to be those drivers who picked up their cabs before 4:30 PM and the evening shift as those picking up their cabs after 4:30 PM (all drivers in our sample picked up their cabs between 1:00 PM and 7:25 PM). Approximately 30 percent of the trip sheets are for the afternoon shift. The results are not sensitive to whether a shift is defined.

For the 70 shifts, the average trip duration was 9.5 minutes, and the average time searching for the next fare was also 9.5 minutes. The average fare per trip was \$5.13. The percent of time that a driver spent with a passenger in the cab was 51.7.

There is no direct information on the experience of the driv-

ers. In the analysis we use hack numbers, which are issued in consecutive order by the TLC. We use a median split on the hack number to separate the drivers into the “low” or “high” experience group.

To try to control for factors other than wages that might affect hours, we collected measures of the weather on the days in our sample from *The New York Times*. It rained during approximately one-third of the shifts, and the high temperature averaged almost 76 degrees. Seventy-three percent of the shifts were during the week.

TLC Data:

We use two data sets of trip sheets collected by the New York City Taxi and Limousine Commission (TLC1 and TLC2). A trip sheet is a sequential list of trips that a driver took on a given shift. For each trip the driver lists the time the fare was picked up and dropped off, and the amount of the fare. On each trip sheet the driver also stamps the following output from the meter: number of trips the meter logged at the start and end of the driver’s shift (the difference is number of trips taken by the driver), number of miles at the start and end of the shift, number of miles “live” (with a passenger), and total revenue this shift (excluding tips).

Our measure of hours worked is obtained directly from the trip sheet. It is the difference between the time that the first passenger is picked up and the time that the last passenger is dropped off. Our measure of total revenue is obtained directly from the meter (we do not have revenues from the trip sheet). We calculate the average hourly wage by dividing total revenue from the meter by the number of hours worked from the trip sheet.

Not all trip sheets were filled out completely, because the number of trips listed by the cabdriver is sometimes much less than the number of trips recorded by the meter. Because the TLC calculates hours from the trip sheet, a screen is needed to eliminate these incomplete trip sheets. Also, the TLC has indicated that the meters malfunction occasionally, recording negative numbers of trips or negative revenues. The TLC developed a screen to discard trip sheets. To pass this screen, the number of trips on the meter must exactly match the number of trips listed by the driver and the percent of “live miles” (percent of miles driven when driver has a passenger) is between 20 and 91.

We were not given the trip sheets themselves, but only the

summary measures such as hours driven, revenue collected, shift, the number of trips, and whether the trip sheet passed the screen. Because we do not have the trip sheets themselves, we cannot devise an alternative screening procedure. Therefore, we use the TLC screen for the analysis in the paper. However, in Appendix 3, we present sample statistics for the screened and the screened-out sample and present the basic regression results in Appendix 4 for the unscreened (full) sample.

Data Set 2: TLC1

This data set consists of 1723 completed trip sheets collected for a study by the New York City Taxi and Limousine Commission (TLC). The shifts occurred mostly during the time period, October 29, 1990, to November 5, 1990. The screen developed by the TLC eliminates 658 trips sheets, and we eliminate 21 additional trip sheets due to missing hack numbers (we need hack numbers to correct the standard errors in the nonfixed-effects model and to estimate the fixed-effects models). Summary statistics of the screened-out sample are given in Appendix 3 (note that the number of observations is 646 rather than 658 because 12 observations are omitted because of missing data on hack number or date driven).

In the screened sample, here are 1044 trip sheets logged by 484 drivers. Of the 1044 trip sheets, 34 percent are from fleet companies, 35.5 percent are leases, and 30.5 percent are from owner-drivers. The NYC TLC estimates that of all shifts driven in 1990, 22 percent are from fleet companies, 30 percent are owner-drivers, and 40 percent are leases (8 percent are other). Therefore, this sample overrepresents fleet company shifts.

The TLC provided measures of experience for the drivers in this sample. Approximately 45 percent of the shifts in the sample have drivers with less than four years of experience.

We obtained from the TLC variables that indicated what time the driver began driving and what shift they had designated for that driver ("day," "night," or "other"). However, we realized that the TLC's designations were not consistent across drivers. For example, there might be many drivers that began driving at AM, most of which were labeled as the "day" shift. However, some drivers that also began driving at AM were labeled "other" shift. We decided to make the shift designations consistent so that all

drivers beginning at the same time were labeled as the same shift. (To do this, we used the majority designation and assigned it to everyone.) Drivers beginning their day: between AM and AM are “day” shift (33.6 percent of trip sheets); between 3PM and 10PM are “night” shift (28.7 percent of trip sheets); “other” is a residual shift category (37.7 percent trip sheets). All the analyses in the paper are run using these shift designations. However, the results are not sensitive to our particular definition of shift. The results are qualitatively unchanged if we use the original TLC definitions (the only difference is in the IV fixed-effects models for TLC1—there is no experience effect). Furthermore, if no shift designation is used at all, the results are identical to those presented in the paper.

We obtained measures of the weather from *The New York Times* for the days in this sample. It did not rain on any of the days in the sample. The high temperature averaged about 65 degrees. Thirty-four percent of the shifts were on the weekend.

Data Set 3: TLC2

This data set consists of 750 trip sheets taken from mini-fleet and owner-drivers. Mini-fleets are smaller operations than fleets and usually lease cabs to drivers weekly or monthly. We cannot identify which trips sheets come from mini-fleets and which are owner-drivers. There is only one observation per driver, mostly from November 1, 2, or 3, 1988. The screen used by the TLC eliminates 38 trip sheets (screen is described above), leaving 712 trip sheets for our analysis. Summary statistics for the screened and screened-out sample are provided in Appendix 3. The regression results do not change at all (mostly because so few trip sheets are eliminated with the screen).

We obtained shift and experience measures from the TLC. Although we do not have the time drivers began their shifts, the TLC designated the shift either “day” or “night.” Fifteen percent of the screened sample are night shift, and 85 percent are day shift. All shifts in this sample are during the week. Forty-six percent of the shifts are with drivers with fewer than three years of experience.

We obtained measures of the weather from *The New York Times*. It rained on approximately 5 percent of the shifts, and the high temperature averaged just over 49 degrees.

APPENDIX 2: DISTRIBUTION OF HOURS AND WAGES SCREENED SAMPLE

	TRIP <i>n</i> = 70	TLC1 <i>n</i> = 1044	TLC2 <i>n</i> = 712
		Hours	
Minimum	6.27	0.78	1.00
5%	6.60	4.42	4.25
10%	7.06	6.18	5.75
25%	8.29	8.09	7.88
50%	9.38	9.67	9.25
75%	10.10	11.08	11.00
90%	10.91	12.50	12.50
95%	11.14	13.77	14.00
Maximum	11.41	23.43	22.25
		Average hourly wage	
Minimum	11.20	3.28	2.17
5%	12.75	10.88	9.61
10%	13.32	12.44	11.38
25%	14.97	14.32	12.99
50%	16.20	16.31	14.71
75%	18.45	18.36	16.45
90%	21.92	21.05	18.39
95%	22.95	23.63	19.48
Maximum	25.43	50.56	35.60

APPENDIX 3: COMPARISON OF SCREENED DATA WITH SCREENED-OUT DATA

TRIP	Screened			Screened out		
	Mean	Median	Std. dev	Mean	Median	Std. dev
Hours worked	9.16	9.38	1.39	6.94	7.24	2.90
Average wage	16.91	16.20	3.21	17.41	17.10	4.67
Total revenue	152.70	154.00	24.99	114.00	123.88	47.69
# Trips listed on sheet	30.17	30.00	5.48	22.81	23.00	9.59
# Trips counted by meter	30.70	30.00	5.72	34.19	36.00	6.89
Sample size		70			122	
Correlation (log hours, log wages)		-.502			-.431	
TLC1	Screened			Screened out		
	Mean	Median	Std. dev	Mean	Median	Std. dev
Hours worked	9.62	9.67	2.88	9.91	9.66	3.44
Average wage	16.64	16.31	4.36	17.88	16.64	8.36
Total revenue	154.58	154.00	45.83	162.13	161.00	58.39
# Trips counted by meter	27.88	29.00	9.15	30.84	31.00	13.45
Sample size		1044			646	

APPENDIX 3: CONTINUED

TLC1	Screened			Screened out		
	Mean	Median	Std. dev	Mean	Median	Std. dev
Correlation (log hours, log wages)		-.391			-.487	
TLC2	Screened			Screened out		
	Mean	Median	Std. dev	Mean	Median	Std. dev
Hours worked	9.38	9.25	2.96	10.03	10.13	1.40
Average wage	14.70	14.71	3.20	10.19	10.36	2.26
Total revenue	133.38	137.23	40.74	100.62	104.17	22.92
# Trips counted by meter	28.62	29.00	9.41	19.76	19.50	7.17
Sample size		712			38	
Correlation (log hours, log wages)		-.269			-.234	

APPENDIX 4: LOG HOURS WORKED EQUATIONS USING FULL SAMPLE

Sample	TRIP		TLC1	
<u>OLS results</u>				
Log hourly wage	-1.402 (.753)	.157 (.113)	-.410 (.053)	-.468 (.028)
Fixed effects	No	Yes	No	Yes
Adjusted R^2	.198	.882	.197	.232
Sample size	192	183	1690	1316
<u>IV results</u>				
Log hourly wage	-.609 (.439)	.190 (.244)	-1.164 (.387)	-1.305 (.273)
Fixed effects	No	Yes	No	Yes
Sample size	192	183	1690	1316
<u>IV by experience results</u>				
Log hourly wage	.127 (.406)	.281 (.242)	-.373 (.319)	-1.194 (.412)
Fixed effects	Yes	Yes	Yes	Yes
Sample size	91	92	564	732

Dependent variable is the log of hours worked. Standard errors are in parentheses and are corrected (for the nonfixed-effects models) to account for the panel structure of the data. All regressions also include weather and shift characteristics (dummy variable for rain, high temperature during the day, dummy variable for shift on a weekday, and time of shift dummy variables) as explanatory variables. Instruments for the log hourly wage include the summary statistics of the distribution of hourly (log) wages of other drivers on the same day and shift (the 25th, 50th, and 75th percentiles).

REFERENCES

- Altonji, Joseph G., "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," *Journal of Political Economy*, XCIV (1986), s176-s215.
- Benartzi, Shlomo, and Richard Thaler, "Myopic Loss Aversion and the Equity Premium Puzzle," *Quarterly Journal of Economics*, CX (1995), 73-92.
- Berg, Elliot J., "Backward-Sloping Labor Supply Functions in Dual Economies—The Africa Case," *Quarterly Journal of Economics*, LXXV (1961), 468-92.
- Bowman, David, Debby Minehart, and Matthew Rabin, "Loss Aversion in a Savings Model," University of California at Berkeley working paper, 1996.
- Browning, Martin, Angus Deaton, and Margaret Irish, "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle," *Econometrica*, LIII (1985), 503-43.
- Duesenberry, J. *Income, Saving, and the Theory of Consumer Behavior* (Cambridge, MA: Harvard University Press, 1949).
- Gneezy, Uri, and Jan Potters, "An Experiment on Risk Taking and Evaluation Periods," *Quarterly Journal of Economics* CXII (1997), 631-645.
- Hardie, Bruce G. S., Eric J. Johnson, and Peter S. Fader, "Modeling Loss Aversion and Reference-Dependence Effects on Brand Choice," *Marketing Science*, XII (1993), 378-94.
- Heckman, James, "Sample Selection Bias as a Specification Error," *Econometrica*, XLVII (1979), 153-61.
- Helson, Harry, *Adaptation-Level Theory* (New York, NY: Harper and Row, 1964).
- Johnson, Eric J., Colin F. Camerer, Talia Rymon, and Sankar Sen, "Limited Computation and Fairness in Sequential Bargaining Experiments," University of Pennsylvania Department of Marketing Working Paper, 1996.
- Kahneman, Daniel, Jack Knetsch, and Richard Thaler, "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, XCVIII (1990), 1325-48.
- Kahneman, Daniel, and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, XLVII (1979), 263-91.
- Laisney, Francois, Winfried Pohlmeier, and Matthias Staat, "Estimation of Labor Supply Functions Using Panel Data: A Survey," in Matyas and Sevestre, eds., *The Economics of Panel Data: Handbook of Theory and Applications* (Dordrecht, The Netherlands: Kluwer, 1992).
- Lucas, Robert E., Jr., and Leonard A. Rapping, "Real Wages, Employment, and Inflation," *Journal of Political Economy*, LXXVII (1969), 721-54.
- MaCurdy, Thomas E., "An Empirical Model of Labor Supply in a Life-Cycle Setting," *Journal of Political Economy*, LXXXIX (1981), 1059-85.
- Maddala, G. S., *Introduction to Econometrics*, 2nd edition (New York, NY: Macmillan Publishing Company, 1992).
- Mankiw, N. Gregory, Julio J. Rotemberg, and Lawrence H. Summers, "Intertemporal Substitution in Macroeconomics," *Quarterly Journal of Economics*, C (1985), 225-51.
- McGlothlin, William H. "Stability of Choices among Uncertain Alternatives," *American Journal of Psychology*, LXIX (1956), 604-15.
- Mulligan, Casey, "The Intertemporal Substitution of Work—What Does the Evidence Say?" University of Chicago Population Research Center working paper 95-11, June 1995.
- NYC Taxi and Limousine Commission, "Taxi Trip and Fare Data: A Compendium," October 29, 1991. NYC Taxi and Limousine Commission, "The New York City Taxicab Fact Book," May 1992.
- Odean, Terry, "Are Investors Reluctant to Realize Their Losses?" University of California-Berkeley Working Paper, 1996.
- Orde-Brown, G., *Labour Conditions in East Africa* (London: Colonial Office, H.M.S.O., 1946).
- Pencavel, John, "Labor Supply of Men: A Survey," in O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics, Volume I* (Amsterdam, The Netherlands: North-Holland, 1986), pp. 3-102.
- Pindyck, Robert S., and Daniel L. Rubinfeld, *Microeconomics* (New York: Macmillan, 1989).
- Read, D., and G. Loewenstein, "The Diversification Bias: Explaining the Differ-

- ence between Prospective and Real-Time Taste for Variety," *Journal of Experimental Psychology: Applied*, I (1995), 34–49.
- Read, D. and G. Loewenstein, "Temporal Bracketing of Choice: Discrepancies between Simultaneous and Sequential Choice," Carnegie Mellon University, Department of Social and Decision Sciences Working Paper, 1996.
- Samuelson, William, and Richard Zeckhauser, "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty*, I (1988), 39–60.
- Shea, John, "Union Contracts and the Life-Cycle/Permanent-Income Hypothesis," *American Economic Review*, LXXXV (1995), 186–200.
- Shefrin, Hersh M., and Richard H. Thaler, "Mental Accounting, Saving, and Self-Control," in G. Loewenstein and J. Elster, eds., *Choice Over Time* (New York: Russell Sage Foundation Press, 1992).
- Thaler, Richard, "Mental Accounting and Consumer Choice," *Marketing Science*, IV (1985), 199–214.
- Thaler, Richard, Amos Tversky, Daniel Kahneman, and Alan Schwartz, "How Myopic Loss-Averse Investors Learn from Experience," *Quarterly Journal of Economics*, CXII (1997), 647–661.
- Tversky, Amos, and Daniel Kahneman, "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics*, CVI (1991), 1039–61.
- Wales, Terence J., "Estimation of a Labor Supply Curve for Self-Employed Business Proprietors," *International Economic Review*, XIV (1973), 69–80.
- Weber, Martin, and Colin F. Camerer, "The Disposition Effect in Securities Trading: An Experimental Analysis," *Journal of Economic Behavior and Organization*, forthcoming.
- Weber, Max, *The Protestant Ethic and the Spirit of Capitalism* (New York, NY: Charles Scribner & Sons, 1958).