Essays on Consumption and Labor Supply

by

Dmitri K. Koustas

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Card, Co-chair
Professor David Romer, Co-chair
Professor Yuriy Gorodnichenko
Professor Amir Kermani
Professor Patrick Kline

Summer 2018

Essays on Consumption and Labor Supply

Abstract

Essays on Consumption and Labor Supply

by

Dmitri K. Koustas

Doctor of Philosophy in Economics

University of California, Berkeley

Professor David Card, Co-chair

Professor David Romer, Co-chair

This thesis investigates topics in consumption and labor supply using "big" data from bank accounts and credit cards from personal financial records. Chapter 1 discusses how these data allow researchers to examine economic activity in gig economy jobs, a sector has been difficult to measure due to lack of data. Chapter 2 focuses on one popular gig economy industry, ridesharing, to explore whether flexible work can help workers better smooth their consumption. Chapter 3 examines how these data can be used to measure the marginal propensity to consume out of permanent shocks by exploiting changes in gasoline prices.

To Y.A.

Contents

List of Figures

List of Tables

Acknowledgments

I thank my family for being so supportive for the many years I was in school. I thank my advisors David Card and David Romer and my committee members, Yuriy Gorodnichenko, Patrick Kline and Amir Kermani, who provided invaluable advice and support throughout the entire project. This research uses anonymized data from a large financial aggregation and bill-paying computer and smartphone application. I am grateful to the company for making these data available for research, and to Michael Gelman, Yuriy Gorodnichenko, Shachar Kariv, Matthew Shapiro, Dan Silverman and Steven Tadelis for their support. This work has benefited from the comments of Yukiko Asai, Christopher Carroll, Andrew Garin, Julien Lafortune, Piyush Panigrahi, David Schoenholzer and Matthew White and from seminar participants at Airbnb, Amherst College, Deutsche Bundesbank, Federal Reserve Board, Chicago Fed, Dallas Fed, San Francisco Fed, University of California Berkeley, University of California San Diego, University of Chicago Booth School of Business, University of Chicago Harris School of Public Policy, University of Texas Austin, University of Maryland, University of Tokyo, Uber, and Waseda University.

Introduction

Thesis Overview

My thesis is composed of three chapters addressing topics in household consumption and labor supply. For each of the topics I explore, I argue that empirical research progress has been constrained, and in some cases stalled, by limitations of previously available data. In the last few years, the “big data” and “information technology” revolutions have come together to make incredibly detailed data on households available for research. Each chapter in my thesis uses one such dataset—data from a personal finance application introduced in Gelman, Kariv, Shapiro, Silverman, and Tadelis (2014). These new data allow me to make insights in ways that simply were not possible just a few years ago when data like these had not yet become available for research. I begin with a overview of the possibilities of big data for consumption and household finance research, before describing the papers that comprise my thesis in more detail.

Possibilities of Big Data for Research on Consumption and Household Finance

For most of the history of applied economics research, the primary source of microdata has come from household surveys. By far, the datasets most commonly used in household consumption and household finance research in particular are the Consumer Expenditure Survey (CEX), the Survey of Consumer Finances (SCF), and the Panel Study of Income Dynamics (PSID). To provide a brief overview, the Interview Survey of the CEX asks households to recall spending over the previous quarter. Households can have up to four observations (one year of data). There is also a Diary Survey, where households record daily spending. While recording expenditure in diaries reduces the scope for recall error, the panel dimension of the segment is short (14 days), limiting its use for many applications. The SCF is the only one of these surveys specifically designed to capture assets. The survey is a repeated cross-section conducted every three years, making it best suited for low frequency analyses of broad trends. The PSID was not designed to be a survey on consumption or assets; however, it has been used heavily in the literature due to its long panel (some households have been in the data continuously since 1968). The commonly used measures of consumption in the PSID are food spending and vehicle miles traveled.[1] Researchers have also imputed total consumption in the PSID based on food consumption and household demographics. This “superior” measure of consumption

[1]Consumption categories were expanded in the PSID beginning in 2001. With a few notable exceptions (e.g. Saporta-Eksten, 2014), these expanded data have been less commonly used, perhaps because the PSID also became biennial during this time.

was proposed by Skinner (1987), and is still heavily used twenty years later, most notably in Blundell, Pistaferri, and Preston (2008), Attanasio and Pistaferri (2014) and Kaplan, Violante, and Weidner (2014).

Numerous issues have been documented with the quality of consumption and asset information in these survey datasets. Pistaferri (2015) discusses longstanding issues with consumption measures in the CEX, most notably serious under-reporting that has been increasing over time, particularly at the upper end of the income distribution. Juster, Smith, and Stafford (1999) compare wealth in the SCF, PSID, and Flow of Funds, writing that the "size of measurement error in wealth is quite disturbing especially since most economic models deal most directly with wealth change" (274). Some researchers have addressed these issues by explicitly modeling sources of error or making other statistical adjustments to the data (Pischke, 1995; Alan, Attanasio, and Browning, 2009; Attanasio, Battistin, and Ichimura, 2007). With few exceptions, these adjustments are usually not adopted by other researchers. The limitations of survey data have led one prominent researcher to declare the existing data unsuitable for certain types of commonly used empirical research designs, such as GMM estimation and log linearizing Euler equations (Carroll, 2011).

While these datasets have also been used to understand joint consumption-leisure decisions (e.g. Blundell, Pistaferri, and Saporta-Eksten, 2016), in many papers in labor economics the empirical focus tends to be on labor supply behavior in particular. One reason this is the case is that data on hours and labor supply are best captured by other datasets without consumption or asset data, most notably the Current Population Survey (CPS). To name one example, Card, Chetty, and Weber (2007) test implications of credit constraints using administrative data from Austria on severance pay and job search.

Recently, researchers have become aware of limitations of traditional survey datasets in capturing key areas of labor market activity. An important example is the growth of new forms of work arrangements and independent contracting, such as the so-called "gig" economy. There is general agreement that the sector has grown exponentially, particularly in the transportation and accommodation industries. However, the gig economy is virtually missing from government survey data (Katz and Krueger, 2016; Abraham, Sandusky, Haltiwanger, and Spletzer, 2017). This is because government surveys were designed to capture traditional employment relationships. Even the recently implemented Contingent Worker Supplement (CWS) to the Current Population Survey (CPS), which targets alternative work, prioritizes data collection for *main* jobs.

Despite these documented issues, survey datasets have been heavily used for research in these areas for so long mainly because there have been few better alternatives. Recently, digital transaction data from bank accounts and credit cards have started to become available for research.[2] Digital transaction data have been collected by banks for some time now, and were used as early as Gross and Souleles (2002). However, only lately have credit or debit card usage become commonplace, so that researchers can now study a reasonably representative sample of American households.[3] The two main examples of these data used in research are data from the J.P. Morgan Chase Institute (JPMCI), which follows an anonymized sample of Chase Bank customers,[4] and personal financial software used by households to aggregate financial accounts.

In my thesis, I use data from the personal financial software. The specifics of these particular data are discussed in detail in Gelman, Kariv, Shapiro, Silverman, and Tadelis (2014). The data are compared to existing sources like the Census in Gelman, Kariv, Shapiro, Silverman, and Tadelis (2014), and to the CEX in Gelman, Gorodnichenko, Kariv, Koustas, Shapiro, Silverman, and Tadelis (2016). A very similar dataset is discussed in Baker (forthcoming) and used in Kueng (2016). These newly available datasets have been shown to cover a heterogeneous sample of the U.S. population, and aggregates line up surprisingly well with aggregates from existing survey and census data. Outside of the U.S., Olafsson and Pagel (2016) use data from a popular personal finance application in Iceland.

For the first time, these data make it possible to examine spending and saving decisions at the frequency at which these decisions take place. Expenditure is available daily (in some cases, the spending includes a time stamp down to the second).[5] In addition, paychecks come in at the frequency of pay (typically weekly, biweekly and monthly). In contrast, it is not possible to identify pay frequency in any of the main survey datasets discussed above. Knowing the frequency of pay is actually quite important for many applications, such as for properly defining "hand-to-mouth" households in Kaplan and Violante (2014). Moreover, research using survey datasets implicitly aggregate income and consumption at the different frequencies in which

[2]Another important data source in recent years are administrative data. With the exception of Scandanavian data, (e.g. Chetty, Friedman, Leth-Petersen, Nielsen, and Olsen, 2014), these datasets tend to only have limited information on household balance sheets outside of income. Yet another dataset, particularly for consumption research, is scanner data (mainly from AC Nielsen or Symphony IRI). These data do not contain any information on household assets or durable purchases, and income is self-reported, in interval values, with up to a two year lag.

[3]One notable excluded population remains—the "unbanked" population who do not have bank accounts or credit cards.

[4]See, for example, Ganong and Noel (2017).

[5]The process of pending and posting transactions can complicate pinning down the precise timing of spending. Pending transactions may differ from final spending (they could include credit card holds at gas stations or hotels, and spending before tips, for instance), while the final posted transaction may only post days after the actual spending occurred.

the data are collected, which can sometimes be different for different components of the household balance sheet, even within the same dataset. In the CEX Interview for instance, consumption is recorded quarterly, while income is recorded on an annual basis.[6]

Working with high frequency data also introduces some new challenges, particularly for research on household consumption. At low frequency, consumption and *expenditure* are usually considered to align. At higher frequency, this is not necessarily the case, because of shopping and inventory behavior that can lead to true consumption being spread out across multiple periods. Some implications of the mismatch between expenditure and consumption have recently been studied in Coibion, Gorodnichenko, and Koustas (2017). One simple way to deal with the problem is to aggregate the data to lower frequency, such as the month or quarter, which is something I do in my work as a robustness check. Research on high frequency expenditure data is still in its infancy, and should be an exciting area for future research.

One strength of the studies using survey data is that these studies are potentially replicable by any researcher, since these datasets are usually publicly available. Moreover, there is a large community of researchers with a deep familiarity with the variables and data. On the other hand, almost all of the app datasets are proprietary, or have very large barriers to entry. With the advent of new, proprietary datasets, there are usually few other economists that have worked on these data in the past, if any at all, and the barriers to entry are extremely large—running regressions in big data can can take months, for instance. Questions of external validity also arise within specialized samples. This makes replication research using different, but similar, datasets even more important in the future.

While these limitations are important to address and, if possible, correct for, the bottom line remains that millions of Americans are in these datasets, and the population looks broadly representative. Perhaps the most compelling reason to use these data is that there are no other datasets currently available that have high-quality, high-frequency measures of spending, income and assets in one place. For many questions, the returns to using these data are extremely large because of insufficient sample sizes, aggregation, and measurement error in available survey data. In the next section, I discuss how these data allow me to shed light on three different research areas in ways that were not previously possible.

[6]Time aggregation has long been known to introduce spurious correlations in time series (Working, 1960). This fact tends to be ignored in almost all the consumption literature for reasons of tractability. Crawley (2018) shows that Blundell, Pistaferri, and Preston (2008)'s widely cited estimates of partial insurance to transitory shocks of 5 percent are severely downward biased by time aggregation; when a correction is implemented, estimates of partial insurance to transitory shocks rise 19 percentage points, to 24 percent.

Three Papers Using Big Data

The previous section highlighted opportunities (and raised some challenges) of using big data in research. In the chapters of my thesis, I make use of data from a large, personal finance service to make progress that I argue could not be made using previously available data.

My first two chapters focus on a topic that has recently received considerable attention in the media and among public policymakers: the "gig" economy. As discussed earlier, the gig economy is not being captured in survey datasets on labor markets. Some progress has been made using data coming directly from online platforms. For instance, Uber, the largest gig economy company, has made data available to a handful of researchers. However, an important limitation of data from just one platform is that there is no information about economic activity outside of the specific platform, whether it be on another gig economy platform, or in a primary job outside of the gig economy. This is a limitation for studying the gig economy because most individuals who participate in the gig economy are multi-job holders

In my first chapter, "What Do Big Data Tell Us About the Participation Decision for Gig Economy Jobs?" I discuss how data from the personal finance app allow researchers to examine economic activity in and outside gig jobs in unprecedented detail. In the paper, I document the evolution of gig income, non-gig income and other components of the household balance sheet surrounding the participation decision for gig economy jobs. This simple analysis reveals striking pretrends in income and assets. In addition to providing insight into the reasons why households enter the gig economy, I discuss how these findings have potentially important implications for the external validity of previous studies focusing on gig economy activity only.

My second chapter, "Consumption Insurance and Multiple Jobs: Evidence from Rideshare Drivers" more closely examines rideshare drivers, the single largest occupation in the gig economy. In this chapter, I ask whether this increased flexibility helps workers to smooth income shocks. I analyze information on rideshare income, outside income, spending, and liquid assets for a sample of rideshare drivers in the app data. My key finding is that, after households add ridesharing as an additional source of income, the sensitivity of spending to main income falls by 82 percent, suggesting substantial increases in consumption smoothing. Matching these empirical findings to a structural intertemporal labor supply model with credit and labor frictions implies benefits from flexible second jobs of over $1,800 per year. The results suggest the value of leisure is relatively low for this group of workers, which has important implications for understanding the welfare costs of income fluctuations.

The recent trends in the labor force have drawn the attention of policymakers. In some cases, "gig" work has been fully banned, like Uber in much of Europe and Japan, or Airbnb in places in the U.S. like Santa Monica. There is also a trend towards increased regulation, like a recent registration system set up in San Francisco, and calls for minimum wages (Reich and Parrott, 2018). Full accounting of the costs and benefits of new work arrangements are important to inform good public policy. Taken together, the first two chapters of my thesis provide new insights into the costs and benefits of alternative work.

My third chapter, "The Response of Consumer Spending to Changes in Gasoline Prices" is joint with Michael Gelman, Shachar Kariv, Yuriy Gorodnichenko, Matthew Shapiro, Dan Silverman, and Steve Tadelis, and focuses on a different topic: Our key goal is estimating a structural parameter, the Marginal Propensity to Consume (MPC) out of "permanent" shocks. The paper uses the differential impact across consumers of the sudden, large drop in gasoline prices in 2014 for identification. We precisely estimate an MPC out of changes in gasoline prices of approximately one. We are able to construct a "hand-to-mouth" measure following Kaplan and Violante (2014) that, for the first time, can correctly take into account the frequency of the paycycle. When we use this measure to compare our results over status as a hand-to-mouth consumer, we find no differences in the MPC, consistent with the permanent income hypothesis. We also compare our results to what one would find using the CEX, which is much lower frequency and asks households to recall their spending over the previous month. We find that analysis of the CEX produces estimates that are extremely noisy, again highlighting the value of the app data over traditional survey datasets for estimation.

Chapter 1

What Do Big Data Tell Us About Why People Take Gig Economy Jobs?

1.1 Introduction

Why do households take gig economy jobs? There are now several studies examining labor supply of individuals of a particular gig economy company, but little is known about the economic activity of these individuals outside of the gig economy, or even on other gig economy platforms.[1] New government surveys have been designed to better capture the contingent workforce and the gig economy, but these surveys each have their own limitations.[2] In this study, I follow a panel of gig economy workers in data from a large, online personal financial aggregator and bill-paying application for smartphones and computers. While these data have some drawbacks, they are currently one of the only datasets available that contain the complete balance sheet of households who participate in gig economy jobs.[3]

Using an event study analysis, I find that households who participate in the gig economy have outside income and liquid assets that are deteriorating rapidly before they start a gig economy job. There are two main hypotheses consistent with these findings: a voluntary running-down of assets while waiting to gear up for gig work, or financial distress due to outside shocks. The latter explanation can have important implications for studies focusing on gig economy activity only. I discuss two examples: bias from not observing outside shocks, and estimating labor supply elasticities.

1.2 The Gig Economy in Big Data

This paper employs a unique, transaction-level dataset from a large financial aggregator and bill-paying application (henceforth, the "app").[4] A particular strength of

[1]Studies include but are not limited to: Hall and Krueger (2016); Chen and Sheldon (2011); Chen, Chevalier, Rossi, and Oehlsen (2017); Angrist, Caldwell, and Hall (2017); Cohen, Hahn, Hall, Levitt, and Metcalfe (2016); Hall, Horton, and Knoepfle (2017). Hall and Krueger (2016) provide some evidence from outside the gig economy from two surveys of approximately 600 rideshare drivers. However, non-response rate for the survey was quite high at 90 percent. At least one study using proprietary data has explicitly modeled the effects of cross-platform substitution (Angrist, Caldwell, and Hall, 2017), however nearly all studies ignore outside activity.

[2]Most notably, the 2017 Contingent Worker Supplement to the Current Population Survey, known as the CWS, asks only about a person's main job (defined as the job with most hours worked), not about additional sources of income.

[3]Tax data provide a complete picture of outside income, see for instance Abraham, Sandusky, Haltiwanger, and Spletzer (2017). However, tax data are lower frequency and do not contain information on assets.

[4]These same data have previously been used to study the high frequency responses of households to shocks such as the government shutdown (Gelman, Kariv, Shapiro, Silverman, and Tadelis, 2015),

the data I use in this paper is the comprehensive coverage of accounts across different financial providers.[5] Users of the app can link almost any financial account, including bank accounts, credit card accounts, and utility bills. Each day, the app automatically logs into web portals for a user's accounts and obtains the user's account balances and daily transactions. Spending transactions are available beginning December 2012. Bank account balances are available beginning August 2013. The app had approximately 1.4 million active users in the U.S. in 2013.

Baker (forthcoming) discusses benefits and caveats of data like these in detail, and so I will only briefly address issues specific to the data and context that they are used here. While the app data provide rich, high frequency data on consumption, income and assets, these data only have limited demographic information attached.[6] The app data notably exclude "unbanked" households. While this is a potential concern in other contexts, it is largely not relevant for the gig economy because most gig companies require a bank account for direct deposit.

Another potential concern is non-random selection into the app. For instance, if users of the app are more financially responsible in that they are more likely to find sources of extra income to smooth income shocks, this could be a threat to external validity. While I cannot rule out some sources of unobserved selection, in other work, I deal directly with the endogeneity of the participation decision, including exploiting Uber's entry into a city as a source of exogenous variation in participation, and using coworkers at a main employer as a control group facing similar economic shocks (Koustas, 2018).

For the purposes of this paper, I focus on the "on-demand" gig economy. I follow a sample of workers at gig firms that are known for having an easy sign up process, and where households have a reasonable expectation of earning money on any given day. I exclude companies that require specialized knowledge (the so-called

anticipated income (Gelman, Kariv, Shapiro, Silverman, and Tadelis, 2014), and the 2014 fall in gasoline prices (Gelman, Gorodnichenko, Kariv, Koustas, Shapiro, Silverman, and Tadelis, 2016).

[5]For instance, other transaction data that have been used to study this sector can only provide coverage for accounts at Chase Bank (Farrell and Greig, 2016a,b).

[6]This is not always the case with these types of data. Baker (forthcoming) has demographic information that he uses to reweight the data to be more representative of the general population, for instance. Gelman, Kariv, Shapiro, Silverman, and Tadelis (2014) surveyed a sample of app users and found the population to be heterogeneous and broadly in line with US demographics. For demographics specific to the population of gig users, other work using tax data (Jackson, Looney, and Ramnath, 2017), and survey and internal data from a popular ride-sharing platform (Hall and Krueger, 2016), can be referenced. The population is more likely to be male and young (although there are more women and older workers when compared to taxi drivers). Among all gig workers filing Schedule SE in 2014, 46 percent were married, 44 percent had children and the mean (median) age was 40 (38) (Jackson, Looney, and Ramnath, 2017).

"expert economy.") I identify sources of gig economy in the app data by searching for income into checking accounts associated with major firms.[7] The companies that are in my sample for this paper are Caviar, Doordash, Grubhub, Handy, Instacart, Lugg, Lyft, Postmates, Sidecar, Taskrabbit, Urbansitter, and Uber, with the two rideshare companies, Uber and Lyft, comprising the majority of the sample (approximately 80 percent).

The process involved in cleaning the data to make it ready for analysis is discussed in more detail in Appendix A. The final sample used in this paper consists of approximately 20,000 users with gig economy income.

1.3 Evolution of Household Balance Sheets Around Starting a Gig Job

In this section, I ask, "How do household balance sheets evolve around starting a gig economy job?" To examine the evolution of key variables around starting a first gig job, I make use of an event-study framework. The event study framework provides a non-parametric way of exploring the evolution of key variables around starting a gig job, controlling for individual heterogeneity in baseline levels of an outcome, as well as seasonality and trends. I do not imply the event study coefficients are causal effects. On the contrary, part of the point of this exercise is to test for endogeneity by examining the pretrends. The event-study specification I use is standard and given as follows:

$$y_{it} = \sum_{k \in K} \beta_k D_{it}^k + \alpha_i + \alpha_t + \epsilon_{it} \tag{1.1}$$

where y_{it} is an outcome variable of interest, a_i is an individual fixed effect, and α_t is a time period fixed effect. $D_{it}^k = \mathbb{I}\{t = E_i + k\}$ is a indicator for time to first gig pay, E_i, with negative k indicating a future event date, and positive k indicating the event occurred k periods in the past.

I run my specifications at the weekly frequency, omitting the indicator for the period two weeks before first gig earnings. Since households are typically paid one week after working, this assures that the the β_k coefficients are relative to the period before the household first started working in the gig economy. I follow households for one quarter pre and post starting a gig job. Periods beyond one quarter are binned

[7] One issue is that not all gig income is identifiable from transaction strings (if they pay into Paypal, for instance).

together, identifying the average effect for periods beyond one quarter. The sample is restricted to be balanced 4 weeks pre and post the event.[8]

Results

I begin with a graphical presentation of three key event study results: gross gig income, gross gig income less automobile spending,[9] and non-gig income (Figure 1.1). A 95 percent confidence interval is shaded around the main estimates.

The solid red line in the figure shows gross gig income. Gross gig income reaches about $200 in the first weeks after starting a gig job, before declining by about $100. Note that gross gig income includes weeks with $0 from not working. Considerable debate has focused on the costs associated with gig work, the largest being gasoline spending. The dashed red line shows the coefficients for gig income, net of gasoline spending. The increase in gig income net of gasoline spending is about $20 less per week.[10]

The black line shows the coefficients for non-gig income. Non-gig income begins to fall about 12 weeks before the household starts a gig job. This decline accelerates until bottoming out about one month after the household starts a gig job, and only partially recovers over the next month.

Results for additional variables of interest are shown in Table 1.1. Column (1) shows the probability of any payroll income at all in the current week or the next week. For this column only, the sample is restricted to users for whom I ever-observe payroll income.[11] Relative to the period before starting in the gig economy, the share with payroll income during the week is 5.8 percentage points higher. One month after starting in the gig economy, the total drop in the share with payroll income is 8.9 percentage points. Column (2) shows net balances (total liquid assets in bank account and checking accounts, net of credit card debt). Net balances show a very striking pretrend, declining by over $600 and barely recovering in the post-period. Column (3) shows total income, and columns (4) and (5) show spending net of auto expenses in levels and in logs, respectively. While total income is falling in the

[8]Outside of this window, the sample will be unbalanced and the composition of the sample may change, requiring caution in interpreting coefficients. I have experimented with longer balanced panels and find results to be nearly identical over the window (albeit less precise).

[9]See Gelman, Gorodnichenko, Kariv, Koustas, Shapiro, Silverman, and Tadelis (2016) for a discussion of how gasoline spending is identified in the data.

[10]Gas spending is approximately $5 more when focusing on ridesharing jobs only (not shown).

[11]Payroll income is classified based on key identifiers in the transaction strings, such as “direct deposit” or “salary” or income in excess of $100 every 14,15 or 16 days. This is likely to understate the true share of the population with income, if they receive income without these identifiers, or income at different frequencies.

Figure 1.1: Net gig income and outside income around starting a gig job

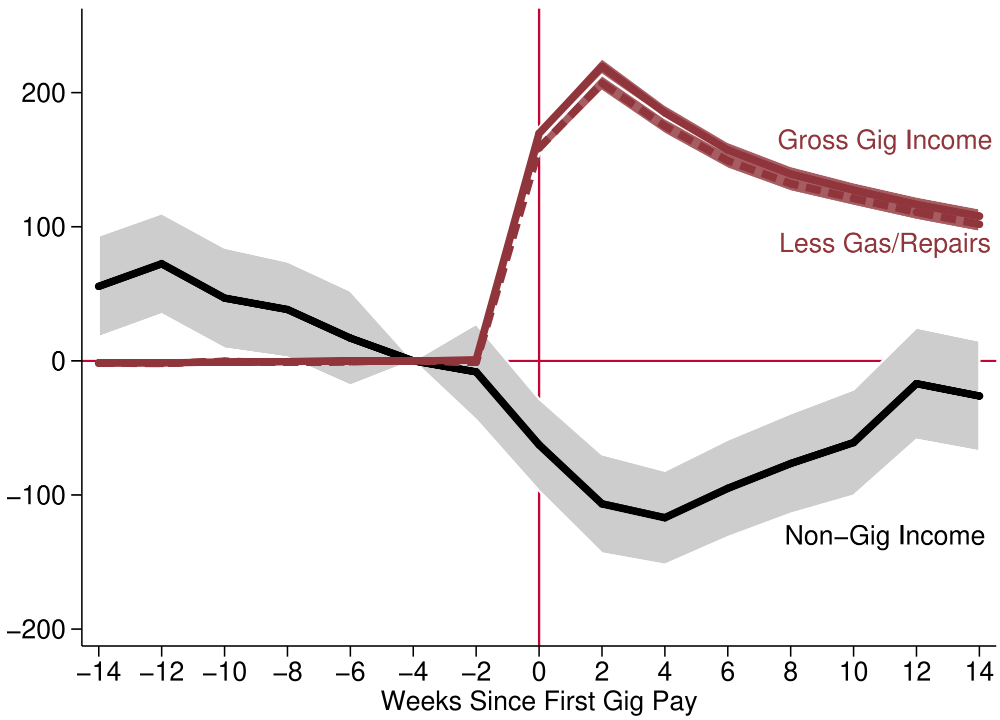


Figure plots the event-study coefficients from Equation 2.12, for three different outcomes of interest. "Gross Gig income" is total income from all gig economy sources. "Less Gasoline" removes gasoline spending from gross gig income. "Non-gig income" is the total non-transfer credits flowing into a user's accounts. The x-axis shows "Weeks Since First Gig Pay." "0" indicates the first week any gig pay is observed. Negative values indicate weeks before first gig pay is received, and positive values indicate weeks after first gig income is received. The y-axis is in dollars to accommodate zero values in the outcomes. 95% confidence intervals are shaded around the estimates.

pre-period, it nearly recovers in the post-period when gig income is combined with other sources of income. Column (4) and (5) show some increase in spending in the immediate months after starting a gig job, and no statistically significant long-run increase in household spending.

1.4 Discussion

This simple analysis above reveals striking pretrends in income and assets. There are a number of potential explanations that are consistent with these findings. One interpretation is that the decline comes from gearing up for gig work, and the bounce back reflects better juggling gig and non-gig work, or simply learning that gig work is not as appealing as anticipated. A second interpretation is that a gig worker is facing outside shocks. This latter explanation has potentially important implications for the external validity of previous studies focusing on gig economy activity only. Below, I discuss two examples.

Example 1. Bias from Not Observing Outside Shocks

Quite simply, large, persistent outside shocks that happen at the same time as households start in the gig economy are likely to confound most analyses of the gig economy. For example, suppose one were to run the naive differences-in-differences estimator:

$$Y_{it} = \beta Post_{it} + \alpha_i + \alpha_t + \epsilon_{it}$$

where $Post_{it}$ is an indicator for being in the period after starting a gig economy job. If households select in after a large income or asset shock, $\text{Cov}(Post, Shock) > 0$. Suppose the shock would lower Y_{it}; then $\text{Cov}(Y_{it}, Shock) < 0$, biasing β downwards.

Studies using proprietary company data will only be able to focus on the post period, since no data is available beforehand. Figure 1.1 shows that outside income bounces back from a shock, which could explain high rates of attrition in the gig economy. Events happening outside the gig economy can also explain low take up rates of experimental incentives within the gig economy.[12]

[12] In Angrist, Caldwell, and Hall (2017), for instance, low take up of lease offers is interpreted as "lease-aversion," but could also be due to improvements in outside options as a function of time.

Table 1.1: Selected Event Study Coefficients

	(1)	(2)	(3)	(4)	(5)
	1{Payroll}	Net Balances	Tot. Income	Tot. Spend ($)	Tot. Spend (Ln)
≤ -14	0.0584***	636.1***	102.3***	39.72**	0.0191
	(0.00646)	(87.12)	(18.05)	(14.14)	(0.0127)
-12	0.0479***	286.2***	91.79***	14.05	0.0227
	(0.00576)	(56.90)	(23.99)	(17.92)	(0.0159)
-8	0.0385***	211.6***	81.11***	7.564	0.00766
	(0.00522)	(43.94)	(23.44)	(17.50)	(0.0155)
-4	0.0142***	98.48***	19.60	-10.08	-0.0161
	(0.00429)	(28.63)	(22.00)	(16.77)	(0.0147)
0	-0.0165***	-54.79	164.5***	36.03*	0.114***
	(0.00455)	(30.34)	(21.96)	(16.98)	(0.0140)
+4	-0.0302***	55.10	72.60***	38.54*	0.120***
	(0.00535)	(42.40)	(21.80)	(16.44)	(0.0146)
+8	-0.0306***	10.74	62.96**	3.337	0.0780***
	(0.00607)	(52.60)	(23.21)	(17.11)	(0.0153)
+12	-0.0217***	-5.847	117.7***	20.01	0.0798***
	(0.00647)	(63.50)	(25.75)	(18.16)	(0.0157)
$\geq +14$	-0.0132	67.93	74.48***	-1.035	0.0229
	(0.00748)	(96.22)	(18.61)	(14.70)	(0.0135)
Dep. Mean	.63	-114.32	1138.26	1017.13	6.78

Table shows selected event-study coefficients from Equation 2.12. Each row in the table is a week around first receiving gig income. Rows indexed with a negative value indicate weeks before first gig pay is received, and rows indexed with positive values indicate weeks after first gig income is received. See text for descriptions of the outcome variables. Standard errors clustered on user in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Example 2. Estimating Labor Supply Elasticities

The empirical results in Section 1.3 imply that persistent wealth shocks and credit constraints affect this population. This has important implications for key structural parameters. Consider the Frisch labor supply elasticity, commonly estimated by regressing changes in log hours on changes in log wages. On one hand, the gig economy appears to provide a perfect opportunity to estimate labor supply elasticities, because hours are flexible, there is considerable variation in wages, and experiments can be designed to provide exogenous variation in wages. As a result, a number of papers have attempted to estimate labor supply elasticities in the gig economy (Hall, Horton, and Knoepfle, 2017; Angrist, Caldwell, and Hall, 2017; Chen and Sheldon, 2011; Chen, Chevalier, Rossi, and Oehlsen, 2017).

An empirical specification regressing log hours on log wages is motivated by log-linearizing the first order condition for hours in the standard utility maximization framework under the assumptions of perfect capital markets and transitory wage shocks. With the presence of borrowing constraints, this is no longer the case. For instance, assume instantaneous utility $U(c,h) = \frac{c^{1-\rho}}{1-\rho} - \alpha\frac{h^{1+1/\eta}}{1+1/\eta}$, where c is consumption, and h is hours, and the household cannot borrow. The structural model underlying hours changes in this case is given by:[13]

$$\Delta \log h_t = \text{constant} + \eta(\Delta \log w_t - \frac{\mu_{t-1}}{\lambda_{t-1}}) + \epsilon_t$$

where λ_t, the marginal utility of wealth in period t, μ_t is the marginal utility of borrowing in period t, and ϵ_t is an error term. μ_t will be positive if credit constraints bind, and 0 otherwise. Since wage growth and μ_t are positively correlated, the Frisch labor supply elasticity will be *downward* biased in the presence of borrowing constraints. Intuitively, if constrained households cannot borrow across periods, they will want to work more even when wages are temporarily low. Simulations from Domeij and Floden (2006) suggest a bias of around 50% in a credit-constrained sample. Interestingly, the labor supply elasticities estimated on gig data are already on the large side for studies using microdata (two prominent studies estimate labor supply elasticities of 1.2 and 1.7), suggesting true labor supply elasticities could be above 2.

[13] See Domeij and Floden (2006).

1.5 Conclusion

While it was already well-known that the gig economy serves as a source of secondary income for many households, there has been little evidence to date on the evolution of outside income and assets of gig economy households. The personal finance data used in this study shows that participating households appear to be facing a decline in income and a significant running down of assets before entering the gig economy. Implications of financial distress have largely been ignored in studies using proprietary company data, but is likely to matter given the large magnitudes. Future work in this area should do more to consider the implications of shocks to outside income and wealth as well as credit constraints.

Chapter 2

Consumption Insurance and Multiple Jobs: Evidence from Rideshare Drivers

2.1 Introduction

Since pioneering work of Friedman (1957) on household consumption, the literature has explored factors that hamper optimal adjustment to economic shocks, including credit market imperfections, adjustment costs of assets and consumption commitments (Zeldes, 1989; Chetty and Szeidl, 2007; Kaplan and Violante, 2014). Adjusting hours is another consumption-smoothing mechanism (Card, 1994; Low, 2005; Swanson, 2012; Heathcote, Storesletten, and Violante, 2014; Blundell, Pistaferri, and Saporta-Eksten, 2016), but it can be difficult to adjust hours due to frictions. For example, many workers face employer-determined scheduling. Exploiting the sudden rise of the rideshare industry noted for its hours flexibility, this paper considers how reducing frictions on hours can affect households' ability to smooth consumption in response to income shocks.

The main objective of this paper is to estimate causal impacts of reducing labor supply frictions on consumption and labor supply. To this end, I exploit a natural experiment that altered flexibility in work arrangements and reduced search and transaction costs. Specifically, recent technological change has allowed more tasks to be performed outside of traditional, arms-length employment relationships. I focus on one notable example: the rideshare industry.[1] Rideshare platforms entered different geographic markets in a staggered fashion beginning in 2012. Figure 2.1 shows the overall rise in rideshare employment over the period 2012-2016. Strikingly, the number of active rideshare drivers now exceeds taxi drivers and chauffeurs in the U.S.[2]

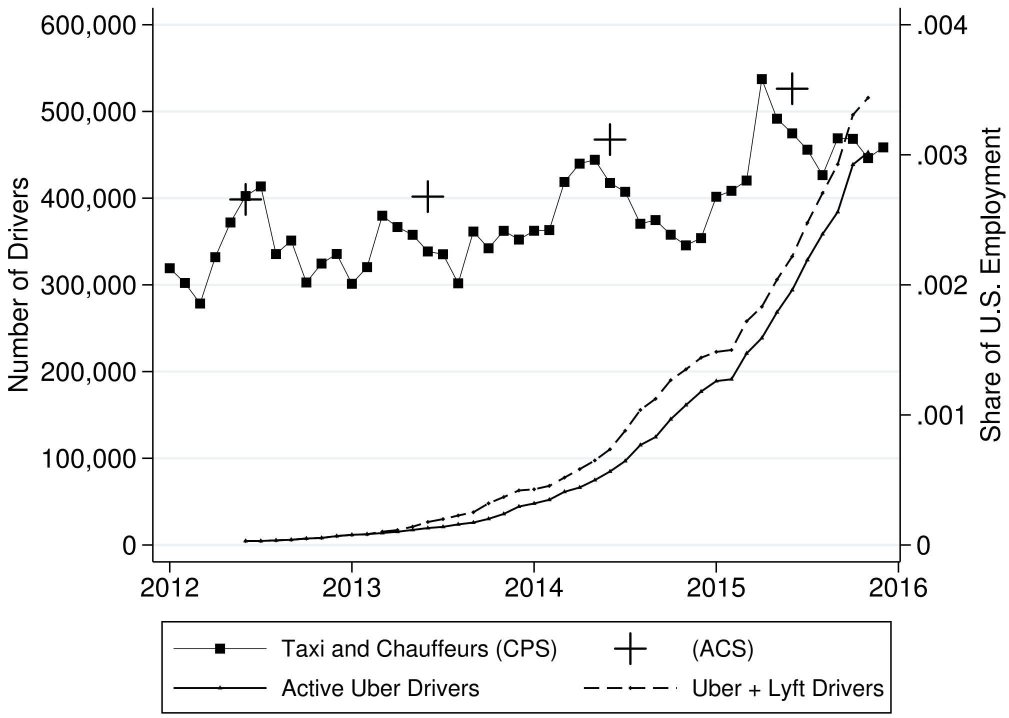
Why has rideshare employment been so popular? One reason may be that households were constrained in their hours choices before rideshare entry. I begin by sketching a consumption model with endogenous labor supply, credit constraints and frictions on adjusting hours. Households in the model are risk-averse and value leisure. In the steady state, the household is on its labor supply curve, but earnings can vary every pay period in response to the employer's demand. During times of weak demand, households are free to take on second jobs in a spot market, but must pay a fixed cost. I model the introduction of ridesharing as an exogenous fall in this cost.

The model has intuitive predictions: selection into second jobs is decreasing in assets and the fixed cost of taking a second job. After the cost of taking a second job falls, consumption smoothing increases, as households are able to better adjust to shocks using a second job. Consumption increases and steady-state assets fall as

[1] The industry is popularly known by the names of the two major platforms, Uber and Lyft.

[2] On taxi and chauffeur employment, Appendix B.1 shows that rideshare drivers are not captured in U.S. survey data.

Figure 2.1: Total Taxi and Rideshare Drivers



"Taxi and Chauffeurs (CPS)" is the weighted count of currently employed individuals with the occupation code "Taxi and Chauffeurs" (occupation code [*peio1ocd*] 9140) in the Current Population Survey Basic Monthly Files. "Taxi and Chauffeurs (ACS)" is the comparable statistic from the American Community Survey. The ACS occurs throughout the year, and so I assign ACS estimates to mid-year. "Active Uber Drivers" is from Hall and Krueger (2016), Figure 1. "Uber + Lyft" drivers is "Active Uber Drivers" multiplied by 1 plus the ratio of Lyft drivers that do not also drive for Uber to Uber drivers in the given month, where this ratio comes from the app data.

households maintain a smaller buffer stock of savings. One natural prediction is that adding a second job helps insure against negative shocks only—households will not work in a second job in response to positive shocks.

I next take the model's predictions to the data. This paper relies on "big data" from over 1.5 million households from a large personal financial management aggregator for smartphones and computers.[3] In the data, I am able to identify approximately 18,000 ever-rideshare drivers between December 2012 to November 2016, which lines up with the largest rise in rideshare employment. This sample is approximately 1 percent of ever-rideshare drivers over this period.

A key feature of these data is that both outside spending and non-rideshare income are observed. This is a crucial advantage over other research on rideshare labor supply using proprietary company data (Hall and Krueger, 2016; Chen and Sheldon, 2011; Chen, Chevalier, Rossi, and Oehlsen, 2017; **?**; Cohen, Hahn, Hall, Levitt, and Metcalfe, 2016) because most rideshare drivers are working temporarily while between jobs or as a second job. As a result, these studies do not observe the primary source of other income, a severe limitation for understanding household labor supply.[4] Having access to total spending and income allows me to identify high-frequency payroll shocks that induce workers to participate in ridesharing. The size of payroll volatility is large and survey evidence suggests that households have trouble smoothing these shocks.[5]

Following Blundell, Pistaferri, and Preston (2008), I consider a key summary measure of household "insurance": the elasticity of consumption spending to income. I estimate the elasticity of consumption to payroll income to be approximately 1/3 over the entire sample of workers. To examine the consumption smoothing benefits of rideshare work, I consider how this measure of partial insurance changes around starting rideshare. My research design addresses two main endogeneity issues. First, since movements in main income may be endogenous to movements in consumption,

[3] These data have been previously used in Gelman (2016); Gelman, Kariv, Shapiro, Silverman, and Tadelis (2014); Gelman, Gorodnichenko, Kariv, Koustas, Shapiro, Silverman, and Tadelis (2016); Gelman, Kariv, Shapiro, Silverman, and Tadelis (2015) and are similar to but distinct from the data used in Baker (forthcoming) and Ganong and Noel (2017).

[4] For instance, Chen, Chevalier, Rossi, and Oehlsen (2017)'s identification strategy does not consider changes in reservation wages over time induced by changes in outside income, which I will show to be a very important determinant of rideshare labor supply.

[5] These shocks are highly volatile with "fat" tails, consistent with Guvenen, Karahan, Ozkan, and Song (2016). Using high-frequency data similar to this paper, Farrell and Greig (2016b) find a typical monthly swing in income around \$500, half of which comes from payroll income. Monthly swings in income of \$500 are not trivial, especially considering that 46 percent of households report being unable to smooth a \$400 shock with cash on hand (Board of Governors of the Federal Reserve System, 2017).

I instrument individual income with income fluctuations common to all coworkers of a given individual at his or her primary job, which should pick up income variation due to changes in firm demand. Second, even though rideshare entry is exogenous, the decision of exactly when to drive for rideshare is still endogenous. This decision is dealt with using a second instrumental variable exploiting cross-city variation in Uber's launch.[6]

The main findings are as follows: Compared to coworkers at their firms, rideshare drivers have a very similar income process and responsiveness of spending to payroll income before they start rideshare. However, rideshare drivers tend to have lower income and much higher debt and credit utilization, suggesting they are more likely to be borrowing constrained. After a household begins ridesharing, total spending (net of auto expenses) rises by about 3-5 percent and the excess sensitivity of spending to main payroll income falls by over 80 percent in my IV specification. Focusing on the differential response of spending to negative income deviations, which is a key prediction of the model, shows that all of the gain in consumption smoothing comes from the response to negative shocks, with no changes in the response to positive shocks.

I next consider how much households would be willing to pay for access to flexible jobs. I calibrate and estimate via Simulated Method of Moments (SMM) a structural intertemporal labor supply model with credit constraints and hours frictions. My structural estimates imply fixed costs of adjusting hours of over $500 per pay period. This cost cannot be directly mapped to a measure of welfare losses, however, because in many cases households would simply choose not to participate in second jobs. Willingness to pay can be calculated from a consumption equivalence variation exercise similar to Lucas (2003), determining the amount of consumption that would make a household indifferent between costly hours adjustment and flexible jobs. Aggregating over the distribution of shocks and the asset distribution in the pre-period implies households would be willing to pay on average around $1,800 per year for flexible labor supply. To put this number in perspective, willingness to pay for flexible labor supply is about two-thirds of the value from completely eliminating borrowing constraints and has similar welfare implications as eliminating the bottom 25 percent of negative income shocks. These results suggest that income fluctuations can be quite costly in the presence of credit constraints and hours frictions.

The paper proceeds as follows: I briefly discuss related literature in the next subsection. Section 2.2 sets the stage with an illustrative intertemporal labor supply

[6] A similar research design has also been used to study transportation demand (Hall, Palsson, and Price, 2016), motor-vehicle fatalities (Dills and Mulholland, 2017; Brazil and Kirk, 2016), and the effect of Uber on the taxi industry (Berger, Chen, and Frey, 2017).

model with credit constraints, hours uncertainty, and frictional labor supply adjustment. The model's predictions are next taken to the data. Section 2.3 provides a detailed discussion of data used in this paper, as well as the construction of the estimation sample and a control group of matched coworkers. In Section 2.4, I discuss the main research design, results of which appear in Section 2.5. Structural estimation is performed in Section 2.6. Section 2.7 concludes with a summary and discussion of policy implications.

Related Literature

Much of the consumption smoothing literature tends to be either focused on life-cycle dynamics (e.g. Bodie, Merton, and Samuelson, 1992; Blundell, Meghir, and Neves, 1993; Low, 2005; Gourinchas and Parker, 2002; Heathcote, Storesletten, and Violante, 2014) and/or a handful of income events that are easy to identify in existing survey data, such as wage changes (Pistaferri, 2003; Heathcote, Storesletten, and Violante, 2014; Blundell, Pistaferri, and Preston, 2008; Blundell, Pistaferri, and Saporta-Eksten, 2016), tax rebates (Johnson, Parker, and Souleles, 2006a; Kaplan and Violante, 2014), or unemployment (e.g. Ganong and Noel, 2017; Gruber, 1997; Card, Chetty, and Weber, 2007). These events tend to be low-frequency and may or may not be expected. One view of these income fluctuations is that they are not very costly because households have a high value of leisure time (Hagedorn and Manovskii, 2008), households have important complementarities between consumption and leisure (Aguiar and Hurst, 2005, 2007, 2013), or households can adjust on other margins, such as by varying shopping intensity (Coibion, Gorodnichenko, and Hong, 2015). An alternative and not mutually exclusive view is that considerable frictions prevent optimal adjustment. A large body of work in macroeconomics and finance has focused on the implications of credit market imperfections, starting from seminal work by Zeldes (1989), Deaton (1991), and Aiyagari (1994). Kaplan and Violante (2014) argue that rational households will want to invest some of their portfolio in frictional, high-return assets, generating excess sensitivity to transitory shocks.

Few studies have investigated the costs of high-frequency income fluctuations, mainly due to reasons of data availability. In many empirical treatments, an identifying assumption for transitory shocks is that they have no impact on contemporaneous consumption spending (e.g. Blundell, Pistaferri, and Saporta-Eksten, 2016). On the other hand, financial diaries collected by Morduch and Schneider (2017) document households struggling with paycheck volatility, both predictable (lower demand at work during a local college football game) and unpredictable (checks not arriving on time), by cutting food and decreasing other spending. For the sample of rideshare drivers considered in this paper, revealed preference suggests that high-frequency shocks have at least some costs. If high-frequency volatility were costless, we would not see households take up rideshare employment once it becomes available.

This paper views rideshare employment from the perspective of the key workhorse model in labor economics, where households also smooth shocks by adjusting labor supply (MaCurdy, 1981; Card, 1994; Altonji, 1993). In a series of recent papers, Swanson (2012, 2014, 2015) provides a general treatment of endogenous labor supply in dynamic consumption models, showing that labor supply has important implications for risk aversion and asset pricing. Heathcote, Storesletten, and Violante (2014) find that about 20 percent of shocks over the lifecycle are smoothed by adjusting labor supply. When credit market frictions are added to this class of models, a key result is that they will lead to more volatile consumption, but consumption volatility will be mitigated by an increase in hours worked (Domeij and Floden, 2006).

In the standard version of this model, households choose labor supply freely. In reality, households often do not have control over their hours in their main jobs, particularly in high-frequency. Literature examining optimal adjustment in practice has mainly focused on either dual earning households or unemployment search intensity. Research on the "added worker effect" goes back to Mincer (1962), and has most recently been advanced by Blundell, Pistaferri, and Saporta-Eksten (2016), who take advantage of the new consumption modules in the PSID. Mankart and Oikonomou (2017) consider dual earners in a search and matching model of the labor market. Card, Chetty, and Weber (2007) examine the effect of cash on hand on job search intensity. An earlier literature focused on participation frictions. Cogan (1981) investigates the effect of extensive margin frictions in a static setting. Card (1990) considers a minimum hours threshold of work, the idea being that a portion of the budget constraint is simply unobtainable. Early empirical work focused on testing for the existence of hours constraints (Ham, 1986; Altonji and Paxson, 1992; Paxson and Sicherman, 1996). Taking on a second job, also called "moonlighting" or "multi-job holding" (Shishko and Rostker, 1976; Krishnan, 1990; Paxson and Sicherman, 1996; Zhao, 2015), has been viewed as evidence for hours constraints in main jobs.[7] In practice, second jobs will face search and other costs of work, just as in any first job. I will show that while some rideshare drivers experience an unemployment event, most are adding ridesharing as a *second* job. One interpretation consistent with this literature is that these households are mitigating hours frictions in their main job.

[7] An alternative hypothesis is that households value task heterogeneity (Renna and Oaxaca, 2006). While this may be true for some second jobs, this is unlikely to the driving force behind becoming a rideshare driver.

This paper is also related to a recent literature investigating whether workers value flexibility in their work arrangements. Using experimental variation in schedules offered to call center workers, Mas and Pallais (forthcoming) find that jobseekers are not willing to pay for added flexibility in their employment contracts but are willing to pay substantially to avoid evening and weekend work. Looking at Uber drivers, Chen, Chevalier, Rossi, and Oehlsen (2017) use variation in household reservation wages at different times of the day and days of the week to calculate a willingness to pay for flexible work. They find that households would be willing to pay about $150 per week for the flexibility an Uber job provides, mainly because the household can increase hours on evenings and weekends. The model in this paper provides some insight on this contradiction: risk-averse households want stable first jobs, but value flexible second jobs that can be used to mitigate negative shocks in the main job.

This paper differs from earlier work in its focus on multijob holding, which has proven harder to capture in survey data, and by exploiting a source of experimental variation in the availability of flexible second jobs. In addition, I provide insight on consumption smoothing in high-frequency, and on time-series properties of biweekly income processes.

2.2 Consumption Model with Frictional Second Jobs

This section outlines an illustrative intertemporal labor supply model with credit and hours frictions. While these ingredients have been considered separately in other contexts, I believe this is the first paper to consider these frictions together in a dynamic setting.

The model setup is as follows: Households receive an exogenous income stream from a main employer. In addition, households have the choice to participate in a spot market for a second job. This second job spot market is frictional due to a fixed cost of participation that generates a notched budget set, and a household may decide not to work because of this cost. Note that it is not necessary to have a cost per se to get non-participation in a second job. As long as the household has a positive reservation wage, this alone will generate non-participation if the second job wage is lower than the reservation wage. The difference is that with costs of taking second jobs, the household also has reservation hours—a minimum number of hours the household would want to work in order to be better off than not working at all. In addition, households are assumed to have access to savings technology but have limitations on borrowing.

One way to rationalize this labor market is with implicit or long-term contracts with partial insurance (Abowd and Card, 1987; Beaudry and DiNardo, 1991; Lamadon, 2016). Because it is not the main object of interest for this paper, I leave the contracting process unmodeled, and assume that there is no dynamic incentive to deviate from the contract to seek another permanent job. This could be satisfied if households are on their labor supply curve on average, households are in the steady state, shocks are transitory, and/or there are switching costs across main jobs.

The next subsection lays out a general statement of the household problem. I make a number of standard assumptions, but relax a common assumption that consumption and hours require an interior solution. I establish a number of predictions that will then be tested in the data. In Section 2.6, I consider a numerical solution of the model with additional structure.

The Household Problem

At period t, a representative household chooses consumption, c_t, and *second hours*, $h_{2,t}$, to maximize a stream of future expected utility, discounted by a factor $\beta \in (0, 1)$. Instantaneous utility, $u(c_t, h_t)$, depends on consumption and total hours. Denote derivatives of the utility function with respect to control variable x by u_x. $u(c_t, h_t)$ is assumed to have standard properties: $u_c > 0, u_h < 0$, with $\lim_{c \to 0} u_c = \infty$.

Households have an exogenous main job earnings process, $e_{1,t}$. Main job earnings are a function of main job hours, $h_{1,t}$ (e.g. overtime work can be paid at a higher wage). $h_{1,t}$ could be zero (unemployment). Households can borrow and save but asset positions are subject to a borrowing constraint, $\underline{A}$. Denote A_t as start of period assets (assets carried forward at the end of period $t-1$), and R_t the realized gross interest rate on assets between the end of period $t-1$ and start of period t.

In addition, the household can access a spot market for a second job and earn an hourly wage of $w_{2,t}$. Second hours choices must be non-negative, and total hours, $h_t = h_{1,t} + h_{2,t}$, must not exceed H, the hours endowment. If the household participates in the second job, the household must pay a fixed cost, κ. This fixed cost can be thought of as actual outlays, such as childcare costs and transportation costs, and/or time costs like those from commuting and job search.[8] The exogenous variables $h_{1,t}$, $w_{2,t}$, and R_t are assumed to follow a finite-dimensional Markov process.

[8] Cogan (1981) separately considers both fixed and time costs in a static setting. Denote time costs of work τ. The period budget constraint with participation would be $R_{t+1}A_t + e_{1,t+1}(h_{1,t+1}) + w_{2,t+1} \cdot (h_{2,t+1} - \tau) - \kappa - c_{t+1}$, with total hours $h_t = h_{1,t} + h_{2,t} + \tau \leq H$. Time costs introduce a substitution effect in addition to an income effect. In practice, these costs will be difficult to separately identify from κ when I take the model to the data unless the substitution effect is large, which is why I only focus on one summary measure.

Assume the value function for the household's optimization problem exists and satisfies the recursive Bellman equation. The model is summarized by the following equations:

$$V_t(A_t) = \max_{\{c_t, h_{2,t}\}} u(c_t, h_t) + \beta \mathbb{E}_t \big[V_{t+1}(A_{t+1}) \big] \tag{2.1}$$

subject to:

$$A_{t+1} = \begin{cases} R_{t+1} A_t + e_1\big(h_{1,t+1}\big) - c_{t+1} & \text{if } h_{2,t+1} = 0 \\ R_{t+1} A_t + e_1\big(h_{1,t+1}\big) + w_{2,t+1} h_{2,t+1} - \kappa - c_{t+1} & \text{if } h_{2,t+1} > 0 \end{cases} \tag{2.2}$$

$$\underline{A} \le A_t \tag{2.3}$$
$$h_t = h_{1,t} + h_{2,t} \le H \tag{2.4}$$
$$0 \le c_t, h_{1,t}, h_{2,t} \tag{2.5}$$

Many models from the literature can be nested in this setup. In particular, as the second-hours labor supply elasticity goes to zero, or κ becomes sufficiently large, or $w_2 \to 0$, the model becomes the standard Deaton (1991)-Carroll (1997)-style consumption model with an externally-imposed borrowing constraint, income uncertainty and labor supply that is perfectly inelastic. If $\kappa \to 0$, $w_{1,t} = w_{2,t}$, and the borrowing constraint is relaxed, the model essentially becomes the workhorse intertemporal labor supply model from the labor economics literature.

Model Solution and Dynamics

The solution in any period t is a series of policy functions for consumption and second hours that define optimal behavior. Let $c_t^* \equiv c_t^*(A_t; \Theta_t)$ and $h_t^* \equiv h_{2,t}^*(A_t; \Theta_t)$ denote the time t policy functions for consumption and second hours, respectively, as a function of the state at time t, where Θ_t is a Markov state vector governing the exogenous variables.

Optimal consumption and hours behavior between two periods is governed by the Euler equations. There are two Euler equations, one for consumption and one for second hours. The Euler equation for consumption is standard, and given by:

$$\begin{aligned} u_c(c_t^*, h_t^*) &= \beta \mathbb{E}_t \big[R_{t+1} V_{t+1}'(A_{t+1}) \big] + \mu_t^A \\ &= \beta \mathbb{E}_t \big[R_{t+1} u_c(c_{t+1}^*, h_{t+1}^*) \big] + \mu_t^A \end{aligned} \tag{2.6}$$

where μ_t^A, is the Lagrange multipliers on constraint (2.3). The first line is an immediate result of combining the first order conditions on consumption and assets. The second line follows from the Envelope condition.

Consider the impact of the borrowing constraint on consumption. If the household is borrowing constrained in t, it must be that they expect the marginal utility of consumption to be lower tomorrow. The gain in marginal utility from relaxing the borrowing constraint is given by μ_t^A. Suppose the household has no access to a second job. Then consumption can only be smoothed via assets. If the borrowing constraint binds, then the household will be forced to consume less today. Utility could be improved if the household had access to a technology that could move consumption back to today.

Now, suppose the household can use labor supply to increase consumption today. Whether the borrowing constraint binds or not, *if the household participates*, the household will optimally choose to supply labor until the marginal rate of substitution between hours and consumption is exactly equal to the second wage. This intratemporal condition is given by:

$$w_{2,t} = -\frac{u_{h_2}(c_t^*, h_t^*) + \mu_t^0 - \mu_t^H}{u_c(c_t^*, h_t^*)} \tag{2.7}$$

where the link between the marginal rate of substitution may be distorted if the household is overemployed in the main job ($\mu_t^0 > 0$) or hits the upper bound on available hours ($\mu_t^H > 0$). If the household has a higher marginal utility of consumption today, for instance, because the borrowing constraint binds, the denominator in (2.7) will be relatively high, requiring more hours to maintain equality with the wage offer. The probability constraints will bind in the future will also affect decisions today. Suppose the household gets hit with a negative shock. This means fewer assets are carried forward. If the shock has any persistence, the household expects a similar situation tomorrow. The probability that the credit constraint will bind in the future has increased. This will have an added effect of decreasing consumption and therefore increasing the hours response today. If hours are higher because the borrowing constraint binds, this will be a "second-best" response, because of the extra disutility of work. These dynamics establish the first key prediction.

Prediction 1. *Conditional on participation, second-job hours are increasing for negative deviations in main job hours from the steady state. A optimizing household will seek to increase labor supply by more when credit constrained.*

As in Swanson (2012), assume the model has a nonstochastic steady state, defined as $x_t = x_{t+k} = \overline{x}\ \forall k > t$, where $x \in \{c_t, h_{2,t}, A_t, h_{1,t}, w_{1,t}, w_{2,t}, \Theta_t\}$ and further assume $\overline{w_2} < \overline{w_1}$. For the household's main job to be the optimal contract, at the steady state, the household desires no additional hours at the first job wage, $h_2^*(\overline{A}|_{w_{2,t}=\overline{w}_1}; \overline{\Theta}) = 0$. Steady-state consumption and assets will depend on the primitives in the model.

At the steady state, the model has an interesting asymmetry. Because the household is assumed to be on their labor supply curve in the steady state, any increase in income will increase consumption and lower desired second hours. Since main hours are exogenous and second hours are bounded below by zero, the household will have no ability to adjust hours downward in response to positive realized deviations in main hours. If there is no change in w_1 (no overtime) or the increase in w_1 is insufficient to keep the household on their labor supply curve, then μ_t^0 will bind and the household will be overemployed. In either case, second jobs will only be relevant for negative income and wealth shocks. This establishes my next testable prediction of the model:

Prediction 2. *Second jobs are only a mechanism to smooth negative shocks to main earnings at the steady state.*

Next, I turn to the fixed cost, κ, and its effect on participation in second jobs. The fixed cost enters the Euler equation inside assets, and indirectly inside the decision rules for consumption and second hours. $\partial V(A_{t+1})/\partial\kappa \leq 0$ is immediate, since a higher κ lowers lifetime utility on any path where the household ever works in a second job and has no effect otherwise.

While the household will unambiguously have lower utility due to the costs of work, whether the costs raise or lower consumption and hours is ambiguous. To see this, note that if the household chooses to pay the fixed cost to participate, it has to work more in order to recoup the cost (income effect). It may instead be optimal for the household to choose not to work and pay the costs and to have lower consumption and assets today. The policy rule for second hours will involve a participation decision governed by the reservation wage. The reservation wage, $w_{2,t}^R$, is defined as the minimum wage the household would be willing to accept in order to participate in the second job spot market.[9] The household will participate only

[9]In the standard case without a fixed cost, the reservation wage is the marginal rate of substitution between leisure and consumption at zero hours of work. With a fixed cost, the reservation wage is given by the marginal rate of substitution at reservation hours, where reservation hours are the minimum number of hours required to work for the household to be indifferent between working and not working. Reservation hours are determined from the following cost minimization problem:

$$\min_{c_t,h_{2,t}} R_tA_{t-1} + e_{1,t} + w_{2,t}^R h_{2,t} - \kappa - c_t - A_t|_{h_{2,t}=0} \tag{2.8}$$

subject to

$$w_{2,t}^R = -\frac{u_{h_2}(c_t, h_{1,t} + h_{2,t})}{u_c(c_t, h_{1,t} + h_{2,t})} \tag{2.9}$$

$$u(c_t, h_{1,t} + h_{2,t}) \geq u(c_t(A_t|_{h_{2,t}=0}; \Theta_t), h_{1,t}) \tag{2.10}$$

if the reservation wage is less than the prevailing second job wage, i.e. $w_{2,t}^R < w_{2,t}$. The second job policy decision, or labor supply function, can be summarized by the following three cases:

$$h_{2,t}^* = \begin{cases} 0 & \text{if } w_{2,t} \leq w_{2,t}^R \\ h_{2,t} & \text{if } w_{2,t}^R < w_{2,t} \text{ and } h_{2,t} < H \\ H & \text{if } h_{2,t} > H \end{cases} \tag{2.11}$$

The first case in (2.11) depicts non-participation in the second job. This can occur for three reasons: 1) the household is at its steady state, 2) the household is overemployed and wishes to work less in the current period, or 3) the household wants to increase hours but the costs of work are sufficiently prohibitive. In the second case in Equation (2.11), the household becomes a multi-job holder. The household bears the costs of work, and chooses second-job hours to equate the marginal rate of substitution between work and consumption equal to the wage. The final case is when the household still wants to work more, but is limited by the upper bounds on their time.

This labor market for a given biweekly pay period is illustrated in Figure 2.2. In each panel, the household is assumed to start the period with a different draw of main income. The budget constraint is shown by the thick black lines, and the thin dashed curves represent indifference curves. In Panel (a), the household works 62 hours at the main job. Point A is the (total hours, consumption) bundle at 0 second job hours. The household could also choose to pay the fixed cost and take a second job. In this case, the optimal decision will be to work and consume at point B and save the amount given by the vertical distance between the budget constraint and point B. By definition of the optimum, the marginal rate of substitution at this point is just equal to the second job wage, illustrated by the thick dashed line through point B. Point C is the point where the household is indifferent between not working, and working and paying the fixed cost. The reservation wage is represented by the absolute value of the slope of the thick dashed line that passes through C. Because the reservation wage is higher than the second job wage, the household will optimally behave by choosing not to work, point A.

Panels (b) and (c) illustrate two cases with second job participation. In these panels, the household has a smaller main earnings draw than in Panel (a). In both

where $c_t(A_t|_{h_{2,t}=0};\Theta_t)$ is the consumption solution with no second hours that satisfies the Euler equation, $A_t|_{h_{2,t}=0} = R_t A_{t-1} + e_{1,t} - c_t(A_t|_{h_{2,t}=0};\Theta_t)$ are the optimal assets brought forward in this case, and $u(c_t(A_t|_{h_{2,t}=0};\Theta_t), h_{1,t})$ is the current period utility from not participating in the second job at time t given the state. Evaluated at the solution, Equation (2.8) will be just tangent to the indifference curve at reservation hours with slope equal to the reservation wage.

cases, the marginal rate of substitution is greater than the reservation wage, so the household chooses to participate. In Panel (b), the household will achieve lower current period utility, because the household will optimally choose to save a substantial amount of current period earnings and carry it forward to the next period. In Panel (c), the household achieves higher utility in the current period as well.

Define "underemployment" when it is not optimal to work only because of the costs of work. For a given $\kappa > 0$, define a "small" shock as a shock for which it is not optimal for the household to pay the fixed cost of participation. Accordingly, the household will not use labor supply to insure against the shock and will only smooth by adjusting consumption/ assets. As assets fall (for instance, following a series of negative shocks), or if the shock is "large", the household may reach a point where $w_{2,t} > w^R_{2,t}$. If κ falls, the reservation wage for taking a fixed cost falls, reducing the region of inaction. This is the third key prediction of the model.

Prediction 3. *Costs of work lead to a region of inaction for "small" shocks. Accordingly, eliminating fixed costs will increase labor supply responses to these shocks and reduce consumption volatility.*

To summarize the model: households smooth shocks by adjusting consumption (hence assets) and second job hours. Credit constraints limit the ability to smooth via borrowing, and will make labor supply adjustments more important. Labor supply frictions (modeled here as fixed costs of second job employment) will hamper the ability to adjust via labor supply. Reductions in costs of second employment will lead to increased consumption and second hours, and a reduction in consumption volatility, as the ability to smooth "small" income shocks increases. In the next sections of the paper, I take these model predictions to the data.

2.3 Data

Sources of data on rideshare drivers are currently limited. This paper employs a unique, transaction-level dataset from a large financial aggregation and bill-paying computer and smartphone application (henceforth, the "app").[10] A strength of these data is that they include high frequency income, spending and assets.

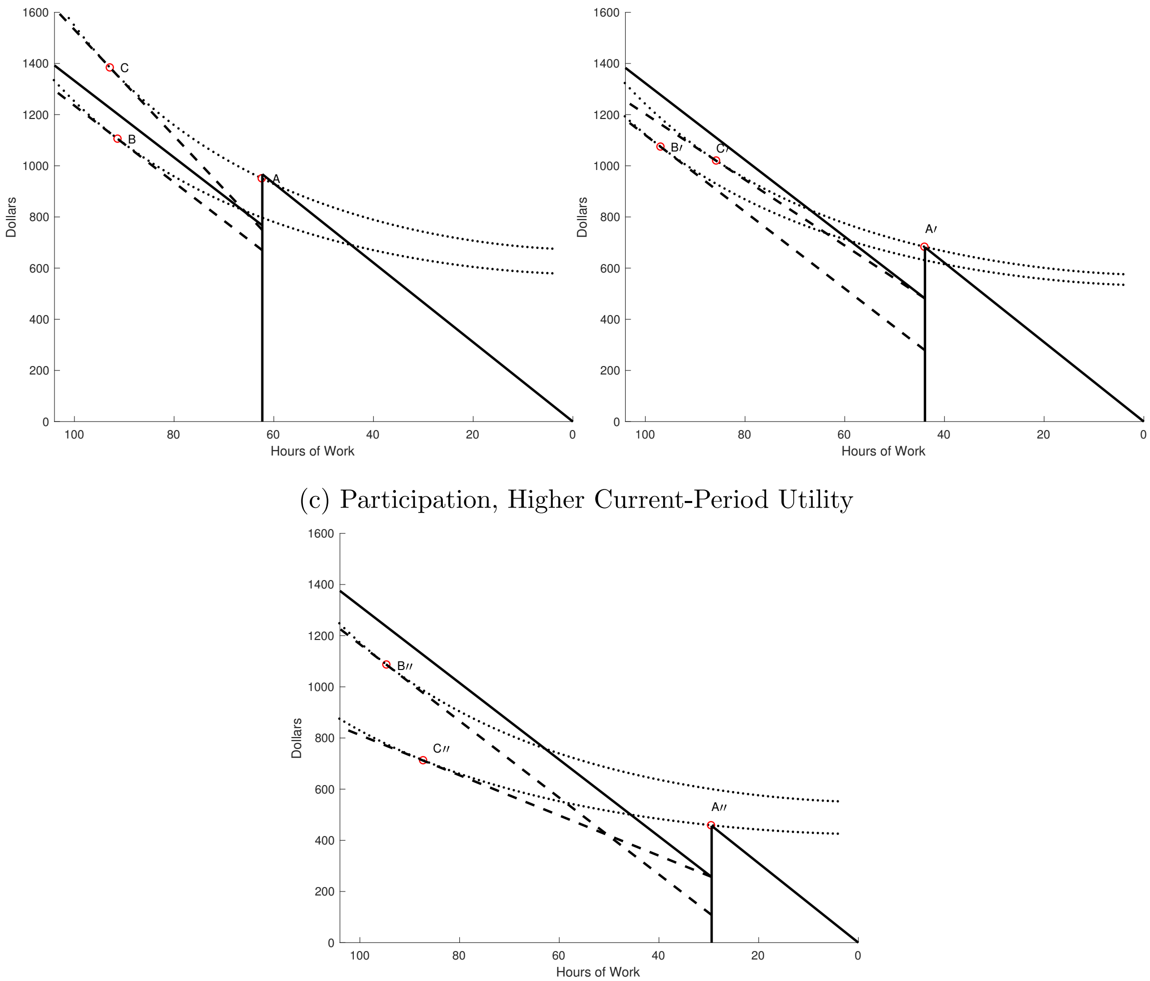
[10] These same data have previously been used to study the high frequency responses of households to shocks such as the government shutdown (Gelman, Kariv, Shapiro, Silverman, and Tadelis, 2015), anticipated income (Gelman, Kariv, Shapiro, Silverman, and Tadelis, 2014), and the 2014 fall in gasoline prices (Gelman, Gorodnichenko, Kariv, Koustas, Shapiro, Silverman, and Tadelis, 2016); similar data have been used to survey the gig economy (Farrell and Greig, 2016b,a) and understand the consumption response to debt (Baker, forthcoming).

Figure 2.2: The Participation Decision

(a) Non-Participation

(b) Participation, Lower Current-Period Utility

(c) Participation, Higher Current-Period Utility



Figures show consumption-leisure tradeoffs within a period. In each panel, the household receives a different draw of main income. The household is assumed to start the period with no assets, an assumption made only for these figures. The budget constraint is illustrated by the thick black lines. Thin dashed curves trace out indifference curves. The thick dashed lines are the lines tangent to the indifference curve at non-participation in the second job (the reservation wage) or at the (hours,consumption) bundle satisfying the intratemporal condition after paying the fixed cost of participation.

Users of the app can link almost any financial account, including bank accounts, credit card accounts, and utility bills. Each day, the app automatically logs into web portals for a user's accounts and obtains the user's account balances and daily transactions. Spending transactions are available beginning December 2012. Asset data are available beginning August 2013. The app had approximately 1.4 million active users in the U.S. in 2013, and approximately 18,000 households have any rideshare income over the period 2012-2016.

A number of sample restrictions need to be made before proceeding with analysis. The most important restriction is that households must be observed for a period before starting rideshare (the "pre period"). A household can sign up for the app after starting ridesharing, which would mean they would only be observed in a "post period." Since rideshare work can sometimes have gaps between weeks worked, I choose 6 weeks of lead time as my cutoff. Many households sign up for rideshare and appear to drop out soon thereafter: perhaps they have learned that it is unsuitable for them. I restrict the sample to households that have some attachment to ridesharing. The sample is restricted to households with 6 or more weeks between first and last observed rideshare pay. The sample is comprised of 10,316 rideshare drivers after this restriction is made.

I also focus on a second subsample of workers with stable, biweekly employment (i.e. paid every 14 days), with identifiable main employers. This subsample is important for two reasons. First, most households in the U.S. are paid biweekly. Second, because I can identify employers, I am able to find coworkers of rideshare drivers at their same primary (non-rideshare) employers. These matched coworkers are useful for identifying common trends. Further, coworkers can be used to identify a firm-level income process that is a plausibly-exogenous source of variation in main payroll income, driven for instance by firm productivity shocks.

This subsample includes 2,217 drivers. While the sample size is considerably smaller, this does not imply that only 22 percent have stable jobs over the period around starting rideshare. First, I only keep households with regular biweekly income, therefore dropping weekly and monthly earners. In addition, I cannot identify employers when a transaction string in the app data does not contain a keyword associated with payroll income. Appendix B.3 describes further how I construct the main samples used in estimation. Appendix B.5 contains results for additional sample restrictions, such as focusing on all workers with non-zero payroll income within a month, instead of just strict biweekly earners.

One obvious concern with data from an online app like this one is that the data are not representative or the sample is composed of households that are more financially sophisticated than the general population. Gelman, Kariv, Shapiro, Silverman, and Tadelis (2014) discuss the demographics of the sample in more detail and find that the population in the app data is heterogeneous and broadly in line with US demographics. In addition, Gelman, Gorodnichenko, Kariv, Koustas, Shapiro, Silverman, and Tadelis (2016) compare the consumption behavior of the whole sample to consumers in the CEX and find that total consumption and gasoline spending in the app and CEX line up closely. The app data notably exclude poor "unbanked" households. While this is a potential concern in other contexts, it is not relevant for this study, because rideshare drivers must have late model cars and receive direct deposit, making a bank account a virtual requirement for becoming a rideshare driver. Effects of app usage on household behavior are another concern. Baker (forthcoming) compares "active" users—users who use the app frequently—with "passive users"—users who sign up but do not use the app frequently. He finds these users have similar consumption behavior, suggesting that intensity of app usage does not change consumer behavior on average.

Descriptive Statistics

Table 2.1 provides descriptive statistics on variables well-measured in the app data: income, assets and spending. I provide statistics on the cross-sectional averages and medians of total spending, total income, payroll income, bank balances, credit card (CC) balances, credit card utilization rates (credit card balances divided by credit limits), and net balances (bank balance net of credit card balance). For spending and income, I calculate each individuals' time-series standard deviation and report the cross-sectional averages and medians over all individuals. I also provide descriptive statistics on the census region of the closest city where Uber operates.[11]

Columns (1) and (2) of Table 2.1 show the descriptive statistics, pre- and post-starting rideshare, for the full sample of ever-rideshare drivers satisfying the sample selection criteria discussed in the previous subsection. Columns (3) and (4) show the same weekly values for the biweekly job-holder subsample with identifiable main employers.

Comparing log spending and its standard deviation across rows, the subsample of regular biweekly earners have higher and less volatile spending than the full sample. Comparing pre and post, spending rises and the standard deviation of spending falls for both groups. Average and median total income is higher for the biweekly earners, mainly due to differences in payroll income. One important difference between the full sample and the subsample is how payroll income changes over the pre and the post period: While payroll income declines for the full sample in the post period,

[11] User location is determined from the most common city where gasoline spending occurs. In a small number of cases where location cannot be scraped from gasoline transaction strings, I default to the location of the user's IP address

Table 2.1: Descriptive Statistics - Full Sample, Weekly Values

	(1) Full Sample-Pre		(2) Full Sample-Post		(3) Biweekly Subsample-Pre		(4) Biweekly Subsample-Post	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Spending	1191.88	954.49	1247.10	1000.39	1253.93	1054.40	1329.18	1118.18
SD$_t$	999.08	810.21	1017.21	819.33	1005.38	844.01	1038.15	852.53
Log Spending	6.48	6.49	6.52	6.54	6.61	6.62	6.68	6.72
SD$_t$	0.95	0.91	0.95	0.89	0.90	0.85	0.89	0.81
Income	1116.50	899.13	1170.15	938.49	1205.19	1014.16	1280.50	1082.18
Payroll Income	415.75	320.60	396.36	259.04	679.46	586.94	680.88	583.93
Bank Balance	3143.01	1049.06	3444.19	1059.13	3346.31	1142.84	3700.18	1261.28
CC Balance	4862.59	2540.27	6288.41	3534.24	4983.54	2650.15	6186.96	3439.24
CC Utilization Rate	0.59	0.63	0.61	0.66	0.58	0.62	0.58	0.63
Net Balance	-1454.08	-954.44	-2518.71	-1799.77	-1289.84	-950.86	-2093.13	-1580.91
Northeast	0.11		0.11		0.11		0.11	
Midwest	0.08		0.08		0.09		0.09	
South	0.34		0.34		0.36		0.36	
West, Excl. CA	0.08		0.08		0.08		0.08	
CA	0.21		0.21		0.18		0.18	
Observations	10316		10316		2217		2217	

Notes: Column (1) and (2) show average and median weekly values, pre- and post- starting rideshare, respectively, for the full sample of ever-rideshare drivers satisfying the sample selection criteria discussed in the text. Columns (2) and (3) shows the average and median weekly values, pre- and post- starting rideshare, respectively, for the subsample of biweekly earners with no break in employment 6 weeks prior to starting rideshare and 4 weeks after starting rideshare and with identifiable main employers. SD_t refers to the household's time-series standard deviation of the variable from one row above. All dollar values and log dollar values are winsorized at the 1% level.

Table 2.2: Descriptive Statistics - Biweekly Earners, Biweekly Values

	(1) Biweekly Subsample-Pre		(2) Biweekly Subsample-Post		(3) Control Coworkers	
	Mean	Median	Mean	Median	Mean	Median
Spending	2593.42	2123.82	2744.65	2271.38	2456.40	1870.04
SD_t	1459.03	1065.33	1482.30	1076.15	1411.03	968.75
Log Spending	7.54	7.52	7.60	7.60	7.37	7.40
SD_t	0.51	0.49	0.54	0.50	0.58	0.52
Income	2660.45	2234.05	2844.68	2383.15	2635.54	2040.62
Log Income	7.57	7.58	7.65	7.64	7.50	7.53
SD_t	0.50	0.49	0.49	0.47	0.49	0.46
Payroll Income	1321.55	1192.19	1355.74	1217.65	1459.81	1234.61
Log Payroll Income	7.01	7.05	7.03	7.08	7.01	7.08
SD_t	0.23	0.19	0.21	0.17	0.23	0.18
Bank Balance	3458.97	887.77	3803.53	921.97	5538.64	1035.65
CC Balance	5009.36	2662.98	6041.40	3311.69	4469.93	2178.24
CC Utilization Rate	0.49	0.46	0.47	0.43	0.40	0.33
Net Balance	-1149.86	-1234.56	-1858.37	-1757.76	1987.03	-490.21
Northeast	0.14		0.14		0.13	
Midwest	0.11		0.11		0.12	
South	0.41		0.41		0.39	
West, Excl. CA	0.10		0.10		0.12	
CA	0.22		0.22		0.15	
Observations	2217		2217		468365	

Columns (1) and (2) show the average and median biweekly values, pre- and post- starting rideshare, respectively, for the subsample of biweekly earners with no break in employment 6 weeks prior to starting rideshare and 4 weeks after starting rideshare and with identifiable main employers. Spending is calculated in the 14 days following receipt of income, while asset values are for the day prior to receiving income. Column (3) shows results for matched coworkers at the same payroll employers ever-rideshare drivers work at in the six months prior to starting rideshare. SD_t refers to the household's time-series standard deviation of the variable from one row above. All dollar values and log dollar values are winsorized at the 1% level.

there is no significant change in average or median payroll income for the biweekly earners. The typical rideshare driver is in debt between \$950 and \$2,519 across the periods, and credit card balances rise for both groups over time. Overall, net balances (bank balances net of credit card balances) deteriorate over the periods.

Table 2.2 provides descriptive statistics for the biweekly job-holder subsample where asset values are taken the day before receipt of biweekly pay and spending is calculated in the 14 days following receipt of biweekly pay. Column (3) of this table provides the descriptive statistics for 468,365 matched coworkers at the same employers. Appendix B.3 describes how the sample of matched coworkers is constructed. Because payroll income is never zero for the groups in Table 2.2, I also include log values of income in the table.

Interestingly, log spending of ever-rideshare drivers is 12-17 percent higher than coworkers in the pre-period. Log payroll income is similar to the coworkers, but log total income is 5-7 percent higher for ever rideshare drivers comparing pre-period values, suggesting ever-rideshare drivers may be more likely to earn income outside of the main job than coworkers. Note that both the ever-rideshare drivers and their coworkers have large volatility in income and spending—the average and median time-series standard deviation of biweekly log payroll range between 17 and 23 percent.

Asset and credit variables show the most striking differences. Ever-rideshare drivers with regular biweekly earnings in the pre-period have less in the bank than coworkers, and more credit card debit. Median credit card utilization in the pre period is 13 percentage points higher than coworkers. Median net balances of ever-rideshare drivers on the day before the paycheck are *negative* \$1,235 in the pre-period and -\$1,758 in the post-period, compared to -\$490 for the median coworker. This finding is in line with Prediction 1 of the model, which stated that second labor supply will be increasing as assets decline.

While the app data provide rich, high frequency data on consumption, income and assets, there is only limited demographic information available. Other work using tax data (Jackson, Looney, and Ramnath, 2017), and survey and internal data from a popular ride-sharing platform (Hall and Krueger, 2016), can be referenced for more detailed demographics specific to rideshare drivers. Briefly, the population is more likely to be male and young (although there are more women and older workers when compared to taxi drivers). In 2014, 46 percent were married, 44 percent had children and the mean (median) age was 40 (38) (Jackson, Looney, and Ramnath, 2017).

2.4 Research Design

This section presents my research design. In Section 2.4, I introduce an event-study framework to examine the evolution of key variables around starting rideshare. I next discuss my research design for the subsample of continuously employed bi-weekly earners in Section 2.4. I then turn to consumption smoothing. I present a differences-in-differences framework measuring how the "partial insurance" parameter from Blundell, Pistaferri, and Preston (2008) changes around starting rideshare. An instrumental variables strategy for addressing endogeneity of main income using firm shocks and endogenous selection into rideshare using Uber's launch is discussed in Section 2.4. Finally, a number of measurement concerns that are particular to high-frequency data, such as shopping and inventory behavior, are discussed in Section 2.4.

As in other related studies, the main estimates will be "Treatment-on-the-Treated" (TOT) effects. Only those households that have a benefit from ridesharing employment will select in. On one hand, we might expect the treatment effects of ridesharing to be largest for the treated since they chose to join rideshare. On the other hand, there are potentially positive treatment effects in the rest of the population as well. Information frictions and other frictions prevent participation, such as credit frictions that prevent a car purchase. Finally, the benefits may be time-varying: households that do not select in now could be hit with an income shock at a later time that could make them participate.

Event Study Framework

Event studies provide a non-parametric way of exploring the evolution of key variables around starting rideshare.[12] The event-study specification I use is standard

[12] There are actually three different events of interest: (1) Uber entry into the local geographic market, (2) signing up for Uber, and (3) first rideshare income. While (1) is the true exogenous event, in reality, information about Uber entry will take time to spread to households, which makes Uber entry not a precise event date. The typical driver in my sample begins driving for rideshare over 1 year following Uber's entry into a market. Event (2) can be proxied with account verification deposits discussed in Appendix B.3. Receiving an account verification suggests the household has applied for driving for Uber. A gap between verification and first income could be indicative of multiple things. First, it could suggest that households are waiting to use rideshare for the first time, perhaps due to an expected future income shock unobserved to the econometrician. Alternatively, there may be a small amount of uncertainty about whether a car or a background check might be approved (although according to Uber, denials are relatively rare). In either case, there is an expected probability of future income, and so consumption theory would suggest that consumption would rise on this expectation. However, what I find is that spending jumps only upon receipt of income, not upon income verification. This motivates my use of receipt of income as the event date.

and given as follows:

$$y_{it} = \sum_{k \in K} \beta_k D_{it}^k + \alpha_i + \alpha_t + \epsilon_{it} \tag{2.12}$$

where y_{it} is the dependent variable of interest. a_i is an individual fixed effect, and α_t is a time fixed effect (calendar week for the weekly sample, actual paydate for the biweekly sample). $D_{it}^k = \mathbb{I}\{t = E_i + k\}$ is a dummy indicating time to first rideshare pay, E_i, with negative k indicating a future event date, and positive k indicating the event occurred k periods in the past. In specifications that are run at the weekly frequency, I omit the indicator for the time period two weeks before first rideshare earnings (1 week prior is the week the household would have worked in order to receive income in period 0). The β_k coefficients are then relative to the week before the household started working as a rideshare driver.

In specifications with a control group, the control group has $D_{it}^k = 0$ for all k. The control group adds precision to the estimates of α_t. If the control group is on a different trend than the ever-rideshare drivers, we can discern this by comparing the estimated pretrends with and without including the control group.

By construction, the sample is balanced 6 weeks prior and 4 weeks post the event. Outside of this window, the sample can become unbalanced. When the sample is not balanced, interpreting the coefficients when few observations are identifying them must be done with caution. As is conventional, the standard errors are clustered at the unit which receives the "treatment." In my baseline specification, this is the household. In other specifications, I will cluster on either firm and city, depending on the source of variation.

Consumption Responses to Income Pre- and Post- Rideshare

The sensitivity of rideshare income and consumption spending are important moments from the model. Of course, rideshare income is only observed in the post period. Although other sources of second income can be observed earlier, other second jobs are difficult to observe in the data, particularly if pay is received in cash or check, rather than direct deposit. On the other hand, household spending can be observed in both the pre and post periods. This allows me to use a differences-in-differences research design for spending. For the specifications that follow, I focus on the sample of biweekly earners with no break in their employment, so that estimates capture responses to the typical biweekly earnings process faced by most workers. My differences-in-differences research design for this sample is discussed next. Threats to identification are discussed in section 2.4.

Consumption and Income Smoothing

Following Blundell, Pistaferri, and Preston (2008), a summary measure of "consumption insurance" is a household's consumption response to income deviations. Consider the following specification:

$$\text{Log Spending}_{i,t} = \delta_1 \text{Log Main Pay}_{i,t} + \gamma_i + \gamma_t + \epsilon_{it} \tag{2.13}$$

where in my case Log $\text{Spending}_{i,t}$ is log total spending net of auto expenditures,[13] Log Main $\text{Pay}_{i,t}$ is log payroll earnings from the main job, and γ_i and γ_t are individual and time fixed effects, respectively. By construction, my sample contains households who have non-zero spending in a biweekly period, so the dependent variable is well defined.

Because this is a log-log specification with individual and time fixed effects, we can interpret δ_1 as the elasticity of spending with respect to changes in main payroll income. This specification can be motivated by a log-linearized version of the consumption Euler equation (see Blundell, Pistaferri, and Preston, 2008). In the reduced-form, δ_1 tells us about the degree of partial "insurance" from income volatility. A value of "0" implies full insurance for payroll income volatility, while a value of "1" implies no insurance.

In Blundell, Pistaferri, and Preston (2008), the main mechanism driving changes in consumption smoothing behavior is postulated to come from two sources: changes in the time-series properties of shocks and increased credit intermediation. The authors estimate elasticities of consumption with respect to transitory shocks and permanent shocks of 0.05 and 0.64, respectively. The key idea in my paper is that changes in households' ability to adjust their labor supply due to technological change will lead to increased consumption smoothing (Prediction 3 of the model.)

As a way to capture this mechanism, consider a differences-in-differences version of Specification (2.13), that includes an indicator for when a household starts driving for rideshare, Post $\text{Rideshare}_{i,t}$, and an interaction between Log Main $\text{Pay}_{i,t}$ and Post $\text{Rideshare}_{i,t}$, as follows:

$$\begin{aligned}\text{Log Spending}_{i,t} = {} & \delta_1 \text{Log Main Pay}_{i,t} + \delta_2 \text{Post Rideshare}_{i,t} \\ & + \delta_3 \text{Log Main Pay}_{i,t} \times \text{Post Rideshare}_{i,t} + \gamma_i + \gamma_t + e_{i,t}\end{aligned} \tag{2.14}$$

The δ_3 coefficient is interpretable as the change in "insurance value" after starting ridesharing. In specifications with a control group of coworkers, δ_1, which tells us

[13] Auto expenditures include spending on gasoline and at auto body repair shops. These expenditures are identified using a machine learning classification algorithm discussed in Appendix B.3.

about smoothing in the pre-period, will also be identified off of the control group. The control group might have a differential spending response to payroll income because they have more assets or a different earnings process. We can explicitly test for this by interacting Main $\text{Pay}_{i,t}$ with an indicator for being an ever rideshare driver. Denote this indicator Ever $\text{Rideshare}_{i,t}$.

Prediction 2 of the model stated that households will become more insured against *negative* shocks. To test this prediction, we can augment Specification (2.14) with an indicator for receiving a negative income deviation. My procedure to identify negative deviations involves two steps. Step 1 is to residualize the covariates from individual and year fixed effects—denote these residuals $\widetilde{\text{Main Pay}}_{i,t}$. These are within-year deviations, picking up seasonal variation in demand, etc. Using the residualized variables in a regression will yield the same coefficients as Specification (2.14) (Frisch-Waugh-Lovell theorem). I next identify negative residuals, $\text{Neg}_{i,t} = \mathbb{I}\{\widetilde{\text{Main Pay}}_{i,t} < 0\}$. Step 2 places these residuals in the main specification. The expanded differences-in-differences specification is shown below:

$$\begin{aligned} \widetilde{\text{Log Spending}}_{i,t} = {} & \delta_1 \widetilde{\text{Main Pay}}_{i,t} + \theta_1 \widetilde{\text{Main Pay}}_{i,t} \times \text{Neg}_{i,t} \\ & + \delta_2 \widetilde{\text{Post Rideshare}}_{i,t} + \theta_2 \widetilde{\text{Post Rideshare}}_{i,t} \times \text{Neg}_{i,t} \\ & + \delta_3 \widetilde{\text{Main Pay}}_{i,t} \times \widetilde{\text{Post Rideshare}}_{i,t} + \\ & \theta_3 \widetilde{\text{Main Pay}}_{i,t} \times \widetilde{\text{Post Rideshare}}_{i,t} \times \text{Neg}_{i,t} + \theta_4 \text{Neg}_{i,t} + e_{i,t} \end{aligned} \tag{2.15}$$

δ_1 (θ_1) now captures the household's response to positive (negative) deviations, and δ_3 (θ_3) estimates how the response to positive (negative) shocks changes in the post period. Because the regressors are generated from a first step, I calculate clustered bootstrapped standard errors over the two steps.

Identification

There are two main endogeneity issues that must be addressed before interpreting the results as the causal effect of starting rideshare: endogeneity in main job income and endogenous selection into rideshare. These concerns and an instrumental variables design to address them are discussed next.

Endogenous Main Income

Specifications (2.13)-(2.15) examine main income as a right hand side variable. One concern is that the household income process may have an endogenous component that is correlated with the dependent variable. In the standard intertemporal labor supply model, for instance, hours in the main job are assumed to be under control at all times. Even if this is not the case, it is reasonable to assume that households still exert at least some control over weeks worked at some points in the year, e.g. they can choose to go on vacation. In this example, consumption spending will increase, but income will stay flat if the household has paid vacations, and may fall if overtime or other earning opportunities are foregone. In addition, some endogenous reasons impacting hours in the first job, such as taking a sick day or going on vacation, will also spill over to the second job.

My instrumental variables design to deal with this issues uses the firm component of main income as an instrument for individual income. Consider the following specification for log main pay additively separable in firm and individual components:

$$\text{Log Main Pay}_{it} = \beta \text{Log Main Pay}_{J(i),t} + \alpha_i + \alpha_t + \zeta_{i,t} \tag{2.16}$$

where J indexes firms. Log Main Pay$_{J(i),t}$ is the firm component of the current period income, which can be driven by productivity shocks, α_i is an individual fixed effect capturing individual ability, α_t is a common aggregate movement, and $\zeta_{i,t}$ is the individual's idiosyncratic component specific to the pay period.

In practice, I consider a leave-out mean Log Main Pay$_{J(-i),t}$ so that the instrument is not biased by the individual's earnings. Figure 2.3 illustrates the first stage by comparing an individual's own income deviations against this leave out mean. The regression coefficient is 0.404 and the F-statistic is over 200. This confirms that common movements in firm income are a large source of variation in biweekly earnings.

Endogenous Participation Decision

Even though the decision to drive for rideshare is first determined by rideshare entry into the market, the household can choose to drive anytime thereafter. While non-time-varying level differences can be controlled for via household fixed effects, idiosyncratic variation in assets and income likely influence the decision of when to start driving as well as consumption levels and consumption smoothing. My research design to deal with this issue exploits the staggered geographic entry of rideshare platforms into different markets.

Consider an expected, idiosyncratic, and permanent decline in income at time $t+1$ that induces a household to drive for rideshare at time t. The counterfactual is lower future consumption. If the household drives for rideshare and only makes up 50 percent of the earnings losses, we would attribute a negative treatment effect to starting rideshare. Pretrends in the event study results can be used to assess the scope of endogeneity in the participation decision. For instance, we might expect

Figure 2.3: Individual Income Residuals v Average Coworker Residuals

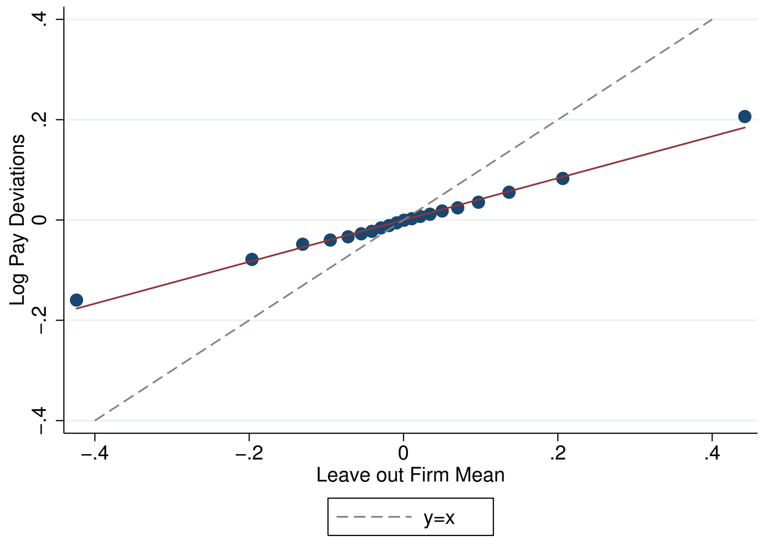


Figure shows income, residualized from year and individual fixed effects, against the average of coworker earnings, also residualized from year and individual fixed effects.

to see an "Ashenfelter" dip, where a variable trends downward in the weeks before starting rideshare. In the event study results, the treatment effect might instead require comparing post outcomes to an earlier period, rather than the period right before starting rideshare. If the household chooses to start driving simultaneous to an income shock, or preceding an expected income shock, then this will not be discernible from the pretrends. Endogenous participation is also a problem in Specifications (2.14)-(2.15) because if the household starts the period with lower assets, they will have higher MPC's out of current income. This will bias the estimates towards finding smaller consumption-smoothing benefits (or even less smoothing in the post period), particularly in short panels.

Suppose the household has the following model for the rideshare participation decision:

$$\text{Post Rideshare}_{it} = \beta \text{Uber Launch}_{c(i),t} + \alpha_i + \alpha_t + e_{it} \tag{2.17}$$

where Post Rideshare$_{it}$ = 1 following a household's decision to start rideshare employment, and 0 otherwise, Uber Launch$_{c(i),t}$ = 1 following Uber's entry into the local market, c, and zero otherwise, α_i is an individual fixed effect reflecting individual distaste for driving, distance to the market, etc, and α_t is a time fixed effect capturing aggregate determinants of entry. e_{it} may contain current or future income shocks which may be in the information set of the household at time t but unobservable to the econometrician.

Equation (2.17) introduces my instrument for the decision to drive: rideshare launch into the market. The maps in Figure B.11 show the spatial time series of when the largest rideshare platform, Uber, launched in a new city. While my empirical results use total rideshare earnings, here I use Uber's launch dates. As shown in Figure 2.1, Uber is the largest rideshare platform and operates in the most markets. Moreover, conditional on being an Uber drivers, 93.3 percent of rideshare earnings come from Uber. Conditional on being a Lyft driver, 33 percent of earnings come from Uber. The exclusion restriction is that rideshare decision to enter the market is exogenous to household consumption decisions. Market size and idiosyncratic reasons such as the friendliness of local governments were key factors in Uber's entry choices. These reasons are unlikely to be related to individual consumption growth paths, suggesting the instrument is valid.[14]

In specifications with a control group, the indicator for Uber's launch, Uber $\text{Launch}_{c(i),t}$, will be a weak instrument, given the probability of driving for rideshare is relatively low, less than 1-2 percent. Because whether someone takes up the treatment is observed, I proceed by instead studying the interaction with an indicator for being an ever-rideshare driver: Rideshare $\text{Launch}_{c(i),t} \times$ Ever $\text{Rideshare}_{i,t}$. This will not yield an ATE because the decision to drive for Uber is not random. However, the decision of *when* to drive for Uber is now exogenous. When the control group consists of coworkers, they only help to identify the time-fixed effects. As with any specification comparing means pre and post, a parallel trends assumption must hold for the identification strategy to be valid.

We can combine the instruments for both income and Uber's launch to instrument Specification (2.14). The first stage is given by:

$$\begin{bmatrix} \text{Log Main Pay}_{it} \\ \text{Post Rideshare}_{it} \\ \text{Log Main Pay}_{it} \times \text{Post Rideshare}_{it} \end{bmatrix} = Z\delta + e_{i,t} \tag{2.19}$$

[14] Since both Post Rideshare_{it} and Uber $\text{Launch}_{c(i),t}$ are binary, the estimated coefficient in Equation (2.17) is a Wald estimator. In a simple regression framework, this estimator is as follows:

$$\beta^{Wald} = \frac{\mathbb{E}[\text{y}_{it}|\text{Uber Launch}_{it} = 1] - \mathbb{E}[\text{y}_{it}|\text{Uber Launch}_{it} = 0]}{\mathbb{E}[\text{Start Rideshare}_{it}|\text{Uber Launch}_{it} = 1] - 0} \tag{2.18}$$

Here, the numerator is the difference in means of the dependent variable pre- and post- Uber's entry into the market and the denominator is the average share of possible periods working for rideshare. Because I consider a regression with individual and time fixed effects, identification comes from across households in other cities that have yet to receive access to ridesharing.

where Z is a block diagonal matrix with the instrument set on each of the diagonal elements equal to [$\text{UberLaunch}_{c(i),t}$, $\text{UberLaunch}_{c(i),t} \times \text{Ever Rideshare}_{i,t}$, $\text{Coworker Earnings}_{J(-i),t}$, $\text{Coworker Earnings}_{J(-i),t} \times \text{UberLaunch}_{c(i),t}$, $\text{Coworker Earnings}_{J(-i),t} \times \text{Uber Launch}_{c(i),t} \times \text{Ever Rideshare}_{i,t}$, α_i, α_t], δ is a column vector of stacked coefficients specific to each endogenous variable, and $e_{i,t}$ is a vector of independent error terms for each endogenous variable

Measurement Issues

The *consumption*-smoothing benefit of flexible labor supply is the key object of estimation in this paper. The app data do not contain "true" consumption, but rather expenditures. Suppose current period expenditure is given by the following accounting identity:

$$E_{it} = C_{it} + D_{it} + \zeta_{it}$$

where E_{it} is expenditures, C_{it} is true consumption, D_{it} represents spending which generate a flow value of utility but is not instantaneous consumption, like durables purchases, inventory purchases and bill pay, and assume ζ_{it} is all other expenditures, some of which may or may not be consumption (such as taxes). Note, each component of expenditures will also affect assets in the same way as consumption.

To test whether a relationship between inventory/shopping behavior is driving results, I consider two robustness checks: (1) including leads and lags of spending in the main regressions and (2) aggregating to lower frequencies following Coibion, Gorodnichenko, and Koustas (2017). If these robustness checks have a limited effect on my main parameters of interest, this suggests a very weak link between inventory behavior and income (i.e. $\text{cov}(D_{it}, \text{Log Main Pay}_{it} \approx 0)$). Other factors, such as whether an item is on sale, likely play bigger roles. In this case, if D_{it} and ζ_{it} are orthogonal to true consumption, they will effectively be measurement error. As is well known, measurement error in the dependent variable will inflate standard errors.

In addition, I do not observe hours, only income. Total rideshare earnings can include bonuses and incentive pay from rideshare providers. This will also be treated as measurement error in my framework. Many of these bonuses are incentives to sign up for ridesharing and fade away over time, and so they should not affect long-run outcomes.

2.5 Empirical Results

The main empirical results are presented in this section. The event-study results for the full sample are found in Section 2.5. Section 2.5 focuses on the sample of continuously-employed biweekly earners. I present the event-study results for this subsample, before exploring consumption smoothing results in Section 2.5.

All Ever-Rideshare Drivers

I begin with a graphical presentation of the event study results for the full sample of ever-rideshare drivers. In all event study figures (Figures 2.4-2.6), the dashed vertical lines indicate that the area between the coefficients are estimated on a balanced sample. 95 percent confidence intervals are shown in dashed gray lines around the main estimates.

Panel (a) of Figure 2.4 shows the event-study results for gasoline spending, measured in dollars. Recall, the coefficients are all relative to the period two weeks before first rideshare pay. Gasoline spending begins to rise 1 week before the first rideshare pay. This happens because first rideshare pay is received with a lag of one week after starting to work. Gasoline spending peaks one week later at a $19 increase, and then declines over time. Gas prices fluctuate a great deal over this period, but assuming an average gasoline price of $2.50, this is about 7.6 gallons of gasoline. The average car in the US at this time received about 21.5 miles to the gallon,[15] implying that the average rideshare driver drove about 160 miles in the week.

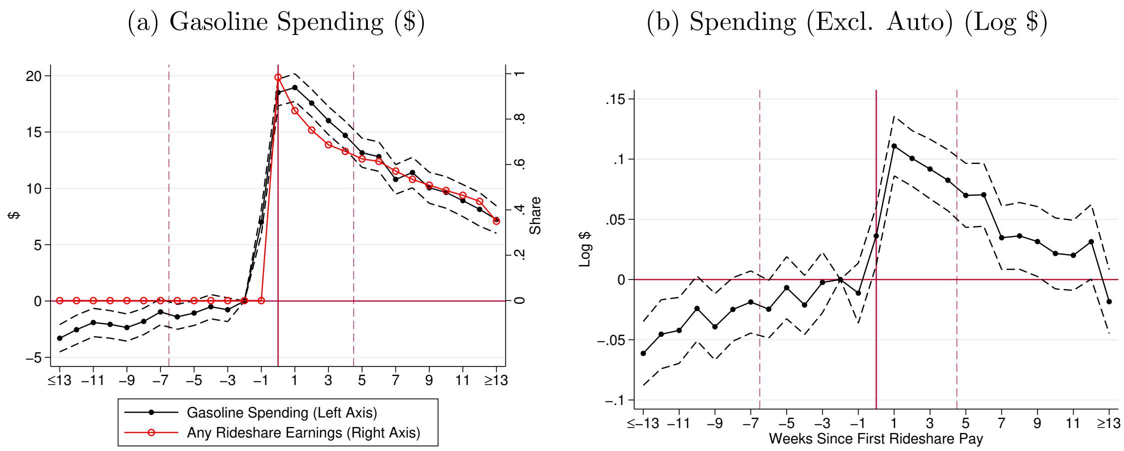
In this figure, I overlay the probability of receiving rideshare pay in any week (red line with hollow marker). The decline in gas spending lines up closely with the decrease in the probability of working in rideshare in that week. About 1 month later, only around 60 percent are working. Recall that I restrict the sample to having last observed rideshare pay at least 6 weeks after the first payment, so this decline is not driven by quitters.

Panel (b) of Figure 2.4 plots the event-study coefficients for log total spending excluding automobile expenditures. In contrast to gasoline consumption, non-gasoline consumption appears not to jump until the week with receipt of income.[16] The pre-event coefficients shows a small positive pretrend over the 3 months prior to starting rideshare. Two weeks after receiving the first rideshare pay, spending increases by about 10 percent. However, the benefits fade over time. The underlying reason will be made clear after examining the income process.

[15] "Table 4-23: Average Fuel Efficiency of U.S. Light Duty Vehicles." Bureau of Transportation Statistics. Link. Last Accessed 11/13/2017.

[16] I have also separately examined results using the account verification date discussed in Appendix B.3 as the event date (not shown). If we examine households that have at least one week gap between income verification and first income, consumption does not jump until the week income is received. This appears consistent with a large literature on expected income shocks (e.g. Johnson, Parker, and Souleles, 2006a).

Figure 2.4: Event Study: Spending



Panel (a) plots the event-study coefficients for gasoline spending in dollars. Panel (b) plots the event-study coefficients from Specification 2.12 for log total spending, excluding gasoline and auto-repair spending. The area between the dashed vertical lines indicates the coefficients are estimated on a balanced sample. 95% confidence intervals are shown in dashed lines around the main estimates. Dependent variables are winsorized at the 1% level.

Before turning to income, I break down spending into component categories. I run separate event studies for different categories of goods. Since most categories have many zeros, the dependent variable is in dollars. The event-study coefficients for 6 weeks pre and 1 week post are shown in Table 2.3. I choose to report these coefficients because 6 weeks pre is the earliest period for which the sample is balanced, and 1 week post is the period where spending peaks, and so it is interesting to see where households are spending this money. The table shows total spending rises by 74.42 one week after first rideshare pay. The negative coefficients 6 weeks pre on services and parts implies spending increases around starting rideshare by about $6 per week on average. In the post period, the household increases fast food spending by about $2.8 per week on average. While this might be partly a non-separability (because of increased work schedules, the household substitutes towards fast food), grocery and restaurant spending also increase, suggesting households increase overall food spending. In addition, clothing and electronics spending are higher relative to two weeks before starting rideshare.

In the next set of figures, I focus on the income process. Because the typical income process is biweekly, I aggregate income over two week periods. Panel (a) of Figure 2.5 focuses on total income, in levels, to accommodate $0 income. Panels (b)

	6 weeks pre	1 week post
Total	-14.91	74.42***
	(16.39)	(17.31)
Gasoline	-1.144**	19.08***
	(0.560)	(0.617)
Service & Parts	-5.900***	0.101
	(1.505)	(1.765)
Fast Food	0.484	2.763***
	(0.295)	(0.296)
Groceries	0.973	3.860***
	(1.018)	(1.050)
Restaurants/Bars	0.425	5.241***
	(0.880)	(0.907)
Personal Care/Services	-1.562	5.214
	(2.921)	(3.301)
Clothing	2.669**	4.349***
	(1.042)	(1.111)
Electronics	0.412	3.444*
	(1.836)	(1.897)

Table 2.3: Event Study Results, By Spending Category
The table shows coefficients from the main event study where the dependent variable is indicated in the rows. Units are given in dollars. The time-frame of aggregate is week. Results are relative to pay periods 2 weeks before starting rideshare. Standard errors clustered on individual in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

and (c) focus on payroll income, excluding rideshare income. The dependent variable in Panel (b) is an indicator for any payroll income, while Panel (c) focuses on payroll income in dollars (again, so that zeros are included).

While Panel (a) shows increases in total income, Panels (b) and (c) show large, persistent declines in main payroll income in the period surrounding rideshare takeup. Panel (b) shows that two months prior to starting rideshare, the percentage with payroll income was about 3 percentage points higher than the period right before starting rideshare. One quarter after starting rideshare, the share with payroll income is about 6 percentage points lower. Thus, the total decline in the share having any payroll income around starting rideshare is around 9 percentage points. While we cannot know for sure whether the income losses are voluntary or not, income appears to begin falling many weeks prior to rideshare takeup. This suggests that the household is not substituting away from the main job specifically to take up a rideshare job. Panel (c) shows payroll income in dollars, and therefore accounts for changes in income coming from both the intensive and extensive margins. One month prior to starting rideshare, income was about $50 higher; the fall in average payroll income mirrors the decline in the probability of working in Panel (b). Comparing one quarter pre rideshare takeup with one quarter post takeup, he total decline in payroll income is $174. On this figure, I also overlay average rideshare earnings *minus* auto expenses (gasoline and car service/repair). Average rideshare earnings net of expenses peak at about $250, and decline to $126 per week after a quarter. When compared to the income results, this suggests that on average rideshare replaces about 73 percent of the decline in main job earnings.

Event studies for the household balance sheet are shown in Figure 2.6. Panel (a) shows the result for bank balances. Panels (b) and (c) show the result for credit card utilization and credit card balances, respectively. The result in Panel (d) shows net balances (bank balances minus credit card balances). Taken together, these pictures tell a consistent story: households are running down assets and racking up credit card debt prior to starting rideshare. After starting rideshare, these balances stabilize.

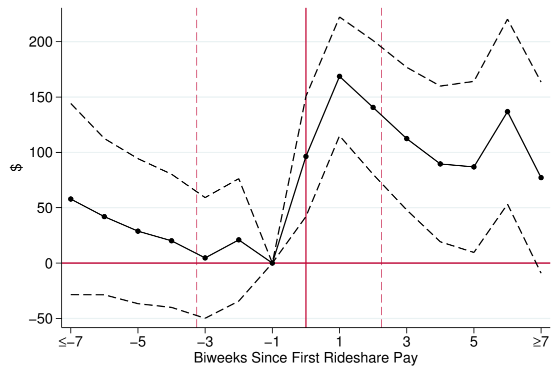
Results for continuously employed, biweekly earners, with matched-coworkers

A key finding from focusing on all ever-rideshare drivers is that rideshare participation follows large, persistent drops in main income, with non-employment increasing by about 9 percentage points. This result makes clear that the full sample faces a mix of transitory and permanent shocks. Focusing on the continuously employed rideshare drivers and on transitory shocks can avoid this complication.

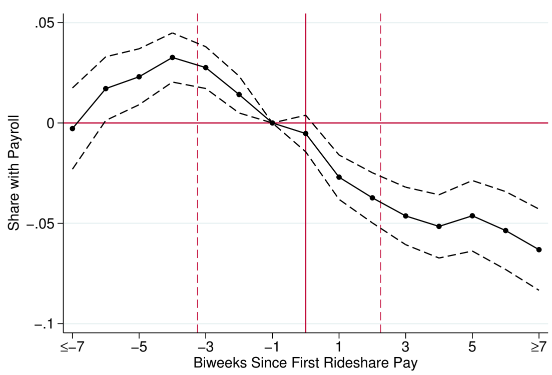
Event study results for continuously employed, biweekly earners are shown in Table 2.4. In these results, coworkers at the same employers are included as a control group. The control group is weighted using inverse-propensity score weights, matching on 2013 income and assets, and accordingly only includes coworkers that were in the data in 2013. Further details of the weighing procedure can be found in Appendix B.4. The counterfactual is that consumption behavior of these rideshare drivers would have evolved similarly in the absence of rideshare income. To compare results across specifications, I bin together coefficients for the pay periods ending in the following windows around first rideshare pay: 90+ days before, 31-90 days before, 1-30 days before, 0-30 days post, 31-90 days post and 90+ days post. In addition, I normalize balances by average daily spending, to address some of the

Figure 2.5: Event Study: Biweekly Income

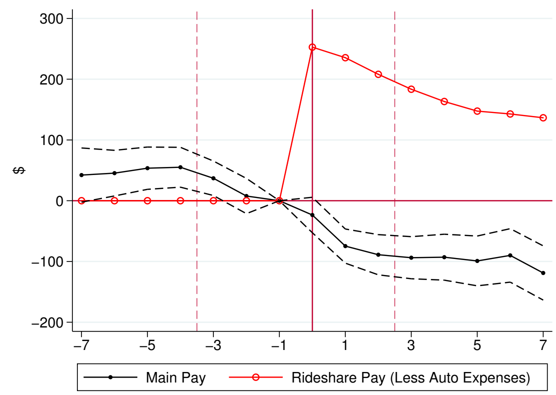
(a) Total Income ($)



(b) Any Payroll Income (Excl. Rideshare)

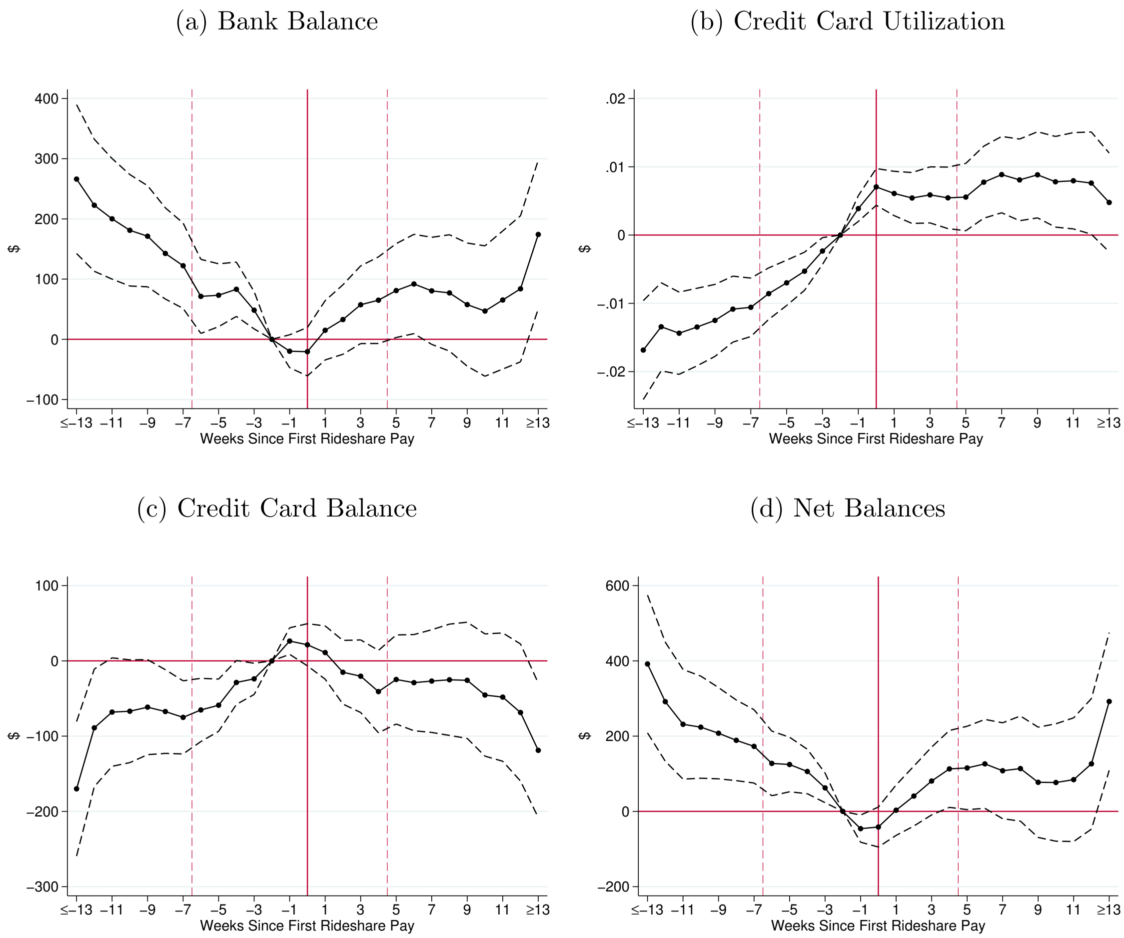


(c) Payroll Income ($) (Excl. Rideshare)



Panel (a) plots the event-study coefficients from Specification 2.12 for total income, including rideshare pay, in dollars. In Panel (b), the dependent variable is an indicator for having payroll income. In Panel (c), the dependent variable is payroll income in dollars. Weekly values are aggregated to the biweek, with biweekly periods numbered sequentially beginning the first week of December 2012. The area between the dashed vertical lines indicates the coefficients are estimated on a balanced sample. 95% confidence intervals are shown in dashed lines around the main estimates. Dependent variables are winsorized at the 1% level.

Figure 2.6: Event Study: Balance Sheet



Panel (a) plots the event-study coefficients from Specification 2.12 for bank balances, in dollars. In Panel (b), the dependent variable is credit card utilization (credit card balance divided by credit limit), for cards with positive balances. In Panel (c), the dependent variable is credit card balances, in dollars. In Panel (d), the dependent variable is net balances (bank balance - credit card balance). The area between the dashed vertical lines indicates the coefficients are estimated on a balanced sample. 95% confidence intervals are shown in dashed lines around the main estimates. Dependent variables are winsorized at the 1% level.

wide dispersion in assets. I omit the period 1-30 days prior to starting rideshare, so results are relative to this period.

Column (1) shows the results for log spending net of auto expenses. There is no evidence of any pretrends for consumption spending, suggesting parallel trends with the control group of coworkers. In the long-run, consumption spending is approximately 2.5 percent higher. Columns (2) and (3) show the results for total income and payroll income, respectively. This group has a slight downward trend in total income in the pre-period, driven by declines in payroll income: 90+ days earlier, total income was about 2.2 percent higher, and payroll income was 2.6 percent higher. Bank and credit card balances (Columns 4-6) also seem to be deteriorating from 90+ days earlier. Because I normalize these balance sheet variables by average daily spending, the interpretation of the coefficients is in terms of days of spending. In payperiods ending 90 days prior to starting ridesharing, the household had about 2 more days of typical consumption in liquid assets. In the immediate post period, the household has about 1.4 days fewer assets. Liquid assets improve as time goes on. Assets in the period 31-90 days after starting rideshare and are not statistically different from the period 31-90 days prior.

Results: Consumption Smoothing

This section focuses on household consumption-smoothing behavior. I begin with a series of results based on Specification (2.14). Column (1) of Table 2.5 includes inverse-propensity score weights discussed in Section B.4, as well as individual fixed effects and year fixed effects. The coefficient on Log Main Pay refers to the elasticity between spending and income in the pre-period for coworkers and is estimated to be 0.3. This estimate is very precise, because it is also identified off of the control group of coworkers. The coefficient on the interaction Log $\text{Pay}_{i,t} \times$ Ever $\text{Rideshare}_{i,t}$ provides a test for whether ever-rideshare drivers have a different sensitivity of spending to earnings. This coefficient suggests ever-rideshare drivers are slightly worse at smoothing, having a total responsiveness of spending to income 3.9 percentage points higher. The key coefficients of interest are on Post $\text{Rideshare}_{i,t} \times$ Log $\text{Pay}_{i,t}$ and Post $\text{Rideshare}_{i,t}$. The coefficient on Post $\text{Rideshare}_{i,t} \times$ Log $\text{Pay}_{i,t}$ implies that spending becomes 6.8 percentage points less sensitive to main income in the period after starting rideshare. The coefficient on Post $\text{Rideshare}_{i,t}$ tells us about the increase in spending if Log $\text{Pay}_{i,t}$ were evaluated at 0 (an out of sample prediction for this group of employed workers). The implied increase is large, 53.8 log points.

Moving to the right are different robustness checks. Column (2) is a more parsimonious specification, dropping the interaction between Log $\text{Pay}_{i,t} \times$ Ever $\text{Rideshare}_{i,t}$. The main interaction on Post $\text{Rideshare}_{i,t} \times$ Log $\text{Pay}_{i,t}$ falls slightly, by about 1 per-

Table 2.4: Event Study Results, Biweekly Earners

	(1) Spending	(2) Total Income	(3) Payroll Income	(4) Bank Balance	(5) CC Balance	(6) Net Balance
Pre, 90+ days	0.00546	0.0220**	0.0257***	1.436**	-0.818	2.205**
	(0.0110)	(0.0103)	(0.00580)	(0.711)	(0.605)	(1.025)
Pre, 31-90 days	0.00749	0.0205**	0.0161***	0.668	-0.0938	0.557
	(0.00962)	(0.00913)	(0.00434)	(0.613)	(0.348)	(0.812)
Post, 0-30 days	0.0395***	0.103***	-0.00247	-0.609	0.409*	-1.434**
	(0.0103)	(0.00970)	(0.00444)	(0.459)	(0.240)	(0.656)
Post, 31-90 days	0.0359***	0.0750***	-0.0172***	-0.111	-0.0575	0.282
	(0.0112)	(0.0107)	(0.00572)	(0.679)	(0.387)	(0.829)
Post, 90+ days	0.0249**	0.0494***	-0.0174***	-0.155	0.140	0.628
	(0.0104)	(0.0101)	(0.00654)	(0.852)	(0.560)	(1.286)
User FE	X	X	X	X	X	X
Paydate FE	X	X	X	X	X	X
NxT	1280232	1277493	1274269	931128	916295	916295
Rideshare N	2217	2217	2217	1140	1140	1140
Control N	64801	64801	64801	43144	43132	43132
R-Sq	0.616	0.620	0.776	0.741	0.856	0.785

The table shows coefficients from an event study specification grouping together pay periods as indicated in the rows. Results are relative to pay periods ending 1-30 days before starting rideshare. Standard errors clustered on user in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

centage point, and the effect on Post Rideshare$_{i,t}$ also falls, suggesting that by not allowing for different pre-period smoothing between rideshare drivers and coworkers will underestimate the benefits of rideshare income. Column (3) drops the weights. This strengthens the effect on Log Pay$_{i,t}$ by about 0.7 percentage points, but the results on the other coefficients are similar to Column (2). Column (4) adds in pay-date fixed effects to the specification in Column (1). The results are very similar to Column (1). Overall, the coefficients are stable as we move across the rows.

Column (5) instruments Log Pay$_{i,t}$ in the regression in Column (3) with average log coworker earnings, Coworker Earnings$_{i,t}$. Log Pay$_{i,t}$ in the interaction is also instrumented. The coefficient on Log Pay$_{i,t}$ rises by 6.2 percentage points, suggesting that households are worse at smoothing firm payroll shocks. The post-period benefits of rideshare increase. The decline in the sensitivity of spending to income is now 9 percentage points. Summing rows (1) and (3) suggests a responsiveness in the post period of 0.2822, slightly higher than the comparable sum of 0.25 in Column (3). Standard errors are clustered on the firm; as a result, the finding is less precise, with a standard error over two times larger than Column (3). While we can reject the pre-period smoothing is the same, the equality of consumption smoothing in the post-period cannot be rejected.

The next set of results shown in Table 2.6 are based on Specification (2.15) in the text. These results test a key prediction of the model, that households will be better able to smooth negative deviations. Post $\widetilde{\text{Rideshare}}_{it} \times$ Log Pay$_{it}$ now refers to the change in response to positive shocks in the post period. The new key coefficient of interest is on Post $\widetilde{\text{Rideshare}}_{it} \times$ Log Pay$_{it} \times$ Neg$_{it}$, which tells us about the household's response to *negative* deviations in income. Column (1) is the result from Column (3) of Table 2.5. Column (2) uses the control group for identification of the pre-period responses, while column (3) interacts all the coefficients shown in the table with an Ever-Rideshare indicator (only the interacted coefficients are shown). In both Columns (2) and (3), the coefficient in the post period becomes insignificant for positive deviations: All the smoothing benefits load on negative deviations. The results are very similar across the two columns, showing a 19 percentage point decrease in the responsiveness of spending to income. This explains part of the reason why the Post Rideshare$_{i,t}$ indicator was so high in Table 2.5—the regression was fitting a single line through a nonlinear relationship. The coefficient on Post $\widetilde{\text{Rideshare}}_{it}$ in this specification shows a 4.78 percent increase in spending in the post period, slightly higher than in the event study.

I return to my more parsimonious specification and explore some additional robustness checks. The next set of results reported in Table 2.7 exploit Uber's staggered geographic entry into different markets. Because identification in this specification comes from differences across cities, the sample is restricted to include households

Table 2.5: Results: Consumption Smoothing - Baseline

	(1) Log Spending	(2) Log Spending	(3) Log Spending	(4) Log Spending	(5) Log Spending
Log Pay	0.304***	0.304***	0.311***	0.292***	0.373***
	(0.00366)	(0.00361)	(0.00341)	(0.00365)	(0.0185)
Log Pay × Ever Rideshare	0.0392**			0.0385*	
	(0.0198)			(0.0199)	
Post Rideshare × Log Pay	-0.0677***	-0.0573***	-0.0593***	-0.0659***	-0.0908***
	(0.0118)	(0.0110)	(0.0110)	(0.0119)	(0.0285)
Post Rideshare	0.538***	0.465***	0.477***	0.526***	0.698***
	(0.0843)	(0.0785)	(0.0784)	(0.0847)	(0.201)
Weights	X	X		X	
User FE	X	X	X	X	X
Year FE	X	X	X		
Paydate FE				X	X
NxT	3059817	3059817	3064697	3059722	3001660
Rideshare N	2217	2217	2217	2217	1932
Control N	64801	64801	64910	64801	64536
OLS/IV	OLS	OLS	OLS	OLS	IV
K/P F-stat					104.22

The table shows coefficients from Specification (2.14) in the text. Weights are discussed in Section (B.4). Standard errors in parentheses. In Columns (1)-(4), standard errors are clustered on user. In Column (5), standard errors are clustered on firm. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 2.6: Results: Consumption Smoothing - Response to Negative Deviations

	(1) Log $\tilde{\text{Spending}}$	(2) Log $\tilde{\text{Spending}}$	(3) Log $\tilde{\text{Spending}}$
Log $\tilde{\text{Pay}}$	0.311***	0.317***	0.351***
	(0.00341)	(0.00437)	(0.0191)
Log $\tilde{\text{Pay}}$ × Neg		-0.0364***	-0.0729***
		(0.00434)	(0.0218)
Post Rideshare × Log $\tilde{\text{Pay}}$	-0.0593***	0.0299	0.0304
	(0.0110)	(0.0416)	(0.0411)
Post Rideshare × Log $\tilde{\text{Pay}}$ × Neg		-0.193***	-0.191***
		(0.0613)	(0.0619)
Post $\tilde{\text{Rideshare}}$	0.477***	0.0475***	0.0479***
	(0.0784)	(0.0106)	(0.0106)
Neg		-0.0138***	-0.0122***
		(0.00133)	(0.00400)
Neg × $\tilde{\text{Post}}$		-0.00728	-0.00901
		(0.0133)	(0.0134)
$\times EverUber == 0$			X
NxT	3064697	3064697	3064697
Rideshare N	2217	2217	2217
Control N	64910	64910	64910

The table shows coefficients from Specification (2.15) in the text. Column (1) is the result from Column (3) of Table 2.5. Column (2) uses the control group for identification of the pre-period responses, while column (3) interacts all the coefficients shown in the table with an Ever-Rideshare indicator (only the interacted coefficients are shown). Bootstrapped standard errors clustered on user in parentheses. See text for more details. * p<0.1, ** p<0.05, *** p<0.01

that were in the data prior to Uber's launch. This restriction cuts the sample size by 2/3. Major cities like San Francisco and New York are dropped from the sample since Uber launched before the data begin. To ensure that the sample is not different in any fundamental ways to the full sample, I first show the OLS estimate for this subsample in Column (1). The coefficients are broadly in line with the main sample. Rideshare drivers appear even worse at smoothing, with an elasticity of spending to income 7.41 percentage points higher than matched coworkers. The benefits of rideshare are slightly larger than the earlier OLS results, showing a 9.6 percentage point decline. In Column (2), Post Rideshare$_{it}$ is instrumented with the indicator for Uber's entry, and Column (3) adds in coworker earnings and all possible interactions to the instrument set. The results in Column (2) shows that the consumption-smoothing benefit gets approximately 2.5 percentage points larger in magnitude, while the other coefficients are unchanged. In Column (3), which includes the full instrument set, the estimated decrease in responsiveness gets very large in magnitude, -0.318. However, the standard errors, which are two-way clustered on firm and city, become 6 times larger than the OLS. The pre-period response of spending to income given in the first row is 0.377, which is similar to the earlier IV result in Table 2.5. Summing the first three rows of Column (3) gives the responsiveness in the post period, 0.377+0.009-0.318=0.068. This result implies consumption insurance from main income increased by 82% after starting rideshare. For completeness, the first stage is shown in Table 2.8. Column (1) shows that Uber Launch × Ever Uber is a strong predictor of rideshare driving—being in the post period increases the probability of driving by 35.8 percentage points. Column (2) shows that coworker earnings are a strong predictor of payroll earnings, with an elasticity of 0.405, very precisely estimated. In Column (4), the interaction has a coefficient of 0.223, a large effect, but imprecisely estimated. Taken together, the full instrument set is sufficiently strong. The last row of Table 2.7 is the Kleibergen-Paap F statistic, which is used to check for weak instruments. For the main regression in column (3) containing the full instrument set, the F-statistic is 11.79, exceeding the "rule-of-thumb" of 10. These standard errors are again two-way clustered on firm and city.

Finally, Table 2.9 examines whether shopping behavior may be driving any of the results. Column (2) adds leads and lags to the baseline specification. The responsiveness of spending to income declines by three percentage points but the increase in consumption smoothing is unchanged. Column (3) aggregates over two biweeks. Now, the responsiveness of spending to income is unchanged, but the increase in consumption smoothing declines by 1 percentage point. Overall, the coefficients are largely stable, suggesting that the regression is picking up something about actual consumption behavior and not changes in shopping or household inventories.

Table 2.7: IV Results: Consumption Smoothing - Instrumenting Uber's Launch

	(1) Log Spending	(2) Log Spending	(3) Log Spending
Log Pay	0.316***	0.316***	0.377***
	(0.00431)	(0.00432)	(0.0172)
Log Pay × Ever Rideshare	0.0741**	0.0795**	0.00879
	(0.0300)	(0.0379)	(0.0791)
StartRideshare × Log Pay	-0.0963***	-0.121*	-0.318**
	(0.0190)	(0.0707)	(0.123)
StartRideshare	0.754***	0.968*	2.368***
	(0.136)	(0.511)	(0.868)
User FE	X	X	X
Year FE	X	X	X
NxT	2257372	2257372	2257372
Rideshare N	620	620	620
Control N	41711	41711	41711
IV	OLS	IV	IV
K-P F-Stat	.	201.99	11.79

The table shows coefficients from a restricted sample in the data before and after Uber's launch in a city. Column (2) is instrumented with Uber's launch into the city. Column (3) is instrumented with both Uber's launch and coworker earnings. Standard errors in parentheses. In Column (1), standard errors are clustered on user. In Column (2), standard errors are clustered on city. In Column (3), standard errors are two-way clustered on city and firm. * p<0.1, ** p<0.05, *** p<0.01

2.6 Model Estimation

The empirical results are qualitatively consistent with key predictions from the model in Section 2.2 for an exogenous decrease in costs of adjusting hours. Few, if any, estimates of the costs of intensive margin frictions on hours exist in the literature. I have argued that rideshare entry provides a credible experiment that can be used to estimate these frictions.

In this section, I proceed with structural estimation of a tractable version of the model in Section 2.2. The experiment I consider is going from a world with frictions of size κ_{pre} to $\kappa_{post} = 0$.[17] The aim of this exercise is to get back-of-the envelope estimates of costs of adjusting hours in traditional jobs, κ_{pre}, and a sense

[17] Rideshare likely has additional costs associated, such as car initial car repairs/ cleaning. As the horizon extends, a one time fixed cost will be a very small share of the total benefits (unless it is very large, like buying a car).

Table 2.8: Instrumenting Uber's Launch - First Stage

	(1) Post	(2) Log Main Pay	(3) Log Main Pay × Ever Rideshare	(4) Log Main Pay × Post
Uber Launch	-0.00129***	-0.0230	0.000203	-0.00914***
	(0.000485)	(0.0362)	(0.000139)	(0.00346)
Uber Launch × Ever Rideshare	0.358**	-0.243	-0.253	1.015
	(0.162)	(0.197)	(0.194)	(1.131)
Coworker Earnings	-0.000558***	0.405***	0.000000117	-0.00396***
	(0.0000764)	(0.0283)	(0.0000194)	(0.000538)
Coworker Earnings × Ever Rideshare	0.0636***	-0.0976	0.309***	0.374**
	(0.0222)	(0.0795)	(0.0962)	(0.155)
Coworker Earnings × Uber Launch	-0.000264***	0.00461	-0.0000180*	-0.00188***
	(0.0000640)	(0.00505)	(0.00000950)	(0.000457)
Coworker Earnings × Uber Launch × Ever Rideshare	0.00144	0.0323	0.0375	0.223
	(0.0229)	(0.0275)	(0.0271)	(0.160)
User FE	X	X	X	X
Year FE	X	X	X	X
NxT	2257372	2257372	2257372	2257372
Rideshare N	620	620	620	620
Control N	41711	41711	41711	41711

The table shows coefficients from Specification (2.19) in the text. Standard errors two-way clustered on city and firm in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 2.9: Robustness to Shopping Behavior

	(1) Log Spending	(2) Log Spending	(3) Log Spending
Log Main Pay	0.311***	0.284***	0.321***
	(0.00341)	(0.00305)	(0.00469)
StartRideshare	0.477***	0.457***	0.413***
	(0.0784)	(0.0795)	(0.107)
StartRideshare × Log Pay	-0.0593***	-0.0581***	-0.0478***
	(0.0110)	(0.0112)	(0.0137)
L1 Log Main Pay		0.0552***	
		(0.00229)	
L1 Post × L1 Log Main Pay		0.0136	
		(0.0139)	
F1 Log Main Pay		0.0338***	
		(0.00243)	
F1 Post × F1 Log Main Pay		-0.0128	
		(0.0139)	
Aggregation	Biweek	Biweek	2 Biweeks
User FE	X	X	X
Paydate FE		X	X
NxT	3064697	2653300	1025638
Rideshare N	2217	2217	1995
Control N	64910	63710	54472

Column (2) includes leads and lags in the main specification, and Column (3) aggregates Specification (2.14) over two biweeks. Standard errors clustered on user in parentheses. * p<0.1, ** p<0.05, *** p<0.01

of the magnitude of the welfare benefits from reducing these costs. The magnitude of these costs is potentially important for a wide range of economic models in labor and macroeconomics. I will structurally estimate two key parameters: the household discount factor, β, and κ_{pre} using Simulated Method of Moments (SMM).

In addition to SMM, there are two other main methods used in the literature to solve models of this class: log-linearizing the Euler equations (Blundell, Pistaferri, and Saporta-Eksten, 2016) and Generalized Method of Moments (GMM). It is difficult to incorporate constraints, like the credit and hours constraints of interest here, in these alternative frameworks. Moreover, GMM performs poorly in the presence of measurement error (Carroll, 2011). My SMM procedure matches well-identified moments from my differences-in-differences research design. While alternative mechanisms like shopping or inventory behavior add measurement error to consumption and are likely to confound some moments, my robustness checks have shown that these alternative activities do appear to be first-order for the responsiveness of spending to biweekly income or the average increase in consumption in the post period. I therefore proceed with my more parsimonious model from Section 2.2, rather than attempting to incorporate shopping/inventory behavior into the model.

Before proceeding with estimation, I begin by delineating additional model assumptions and the calibration of the exogenous variables. I invoke a simple utility function separable in consumption and leisure that is widely used in the literature. Recall my consumption results exclude auto expenses, which are likely to be the largest non-separability for the group considered here. A separable utility function has computational advantages, although my framework can incorporate other classes of utility functions with non-separabilities at additional computational cost. One of the inputs into the model is the biweekly earnings process; I use a first-order Markov process estimated from the app data. After the model is estimated, I consider sensitivity to other parameterizations, before proceeding with welfare calculations.

Additional Model Assumptions

The model presented in Section 2.2 has no closed form solution except in special cases (e.g. quadratic utility). I consider a numerical solution. To make progress, a number of additional assumptions must first be made. Utility is assumed to be the following commonly used, separable utility function:

$$U(c_t, h_t) = \frac{c_t^{1-\rho}}{1-\rho} - \alpha \frac{h_t^{1+1/\eta}}{1+1/\eta}$$

where ρ is the inverse intertemporal elasticity of substitution, η is the labor supply elasticity, and α determines how households weight the disutility of work.[18]

Since interest-rate dynamics are unremarkable over my time period of study, I assume $R_t = R$. An additional assumption that $\beta R < 1$ ("impatient" consumers) ensures that households will not accumulate unlimited assets. Households cannot borrow ($\underline{A}$=0). I assume a coefficient of relative risk aversion of 1.38, taken from (Gourinchas and Parker, 2002), and a labor supply elasticity of 1.35 (the average of the estimated labor supply elasticities for Uber drivers across **?** and Chen, Chevalier,

[18] For example, this same utility function is used to understand lifecycle labor supply in Heathcote, Storesletten, and Violante (2014) and the response of hours to credit constraints in Domeij and Floden (2006).

Rossi, and Oehlsen (2017)). α is chosen so that the household desires no second hours in the steady state at the main job wage.

For the earnings process for second jobs, I assume, $w_t^2 \sim \mathcal{N}(15, 5^2)$, which is the statistical process of Uber earnings reported in Chen, Chevalier, Rossi, and Oehlsen (2017), with an adjustment to mean earnings for expenses.[19] The next subsection explores the main earnings process.

The Biweekly Earnings Process

A first-order Markov process is estimated based on the biweekly income process in the app data. In Figure 2.7, Panel (a), I plot the distribution of residualized earnings (removing year and household fixed effects) of biweekly pay. Earnings changes are centered at zero. The biweekly income process is very volatile: more than 41 percent of biweeks have an earnings deviation greater than 5 percent. Moreover, this distribution has high-kurtosis and "fat-tails," as recently documented in lower frequency earnings data by Guvenen, Karahan, Ozkan, and Song (2016).

Panel (b) shows an impulse response function (IRF) for earnings changes, calculated via a local projection (Jordà, 2005):

$$\text{Log Main Pay}_{i,t+h} = \beta^{(h)}\text{Log Main Pay}_{i,t} + \delta^{(h)}\text{Log Main Pay}_{i,t-1} + \alpha_i^{(h)} + \alpha_t^{(h)} + e_{it}^{(h)} \tag{2.20}$$

where Log Main $\text{Pay}_{i,t+h}$ is log biweekly earnings at time $t+h$, and α_i and α_t are household and paydate fixed effects, respectively. The sample is restricted to earnings received every 14 days. I run this specification for h up to 13 biweeks (one quarter) after time t, and plot the $\beta^{(h)}$ coefficients in Panel (b) of Figure 2.7. The figure can be interpreted as an IRF. The IRF shows that high frequency income shocks appear to follow an ARMA process.

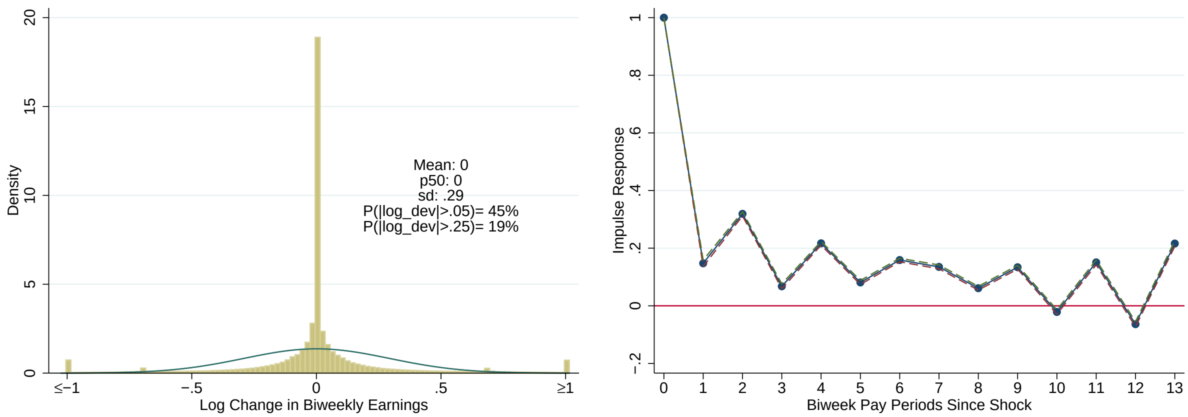
Examining the earnings process alone does not tell us whether earnings changes are coming from hours or wages. Both sources of variation have been explored in the literature. For instance, general equilibrium models with labor adjustment costs imply that wages will need to adjust by more than in a flexible model to induce a labor supply response (see Cogley and Nason, 1995). While adjustment costs may explain quarterly or even monthly variation (e.g. overtime pay), wages tend to be "sticky," particularly in high frequency from paycheck to paycheck.[20] An alternative

[19] Chen, Chevalier, Rossi, and Oehlsen (2017) report mean earnings (before expenses) centered at \$20. I subtract off 25 percent for the Uber fee and other expenses to arrive at a mean wage of \$15.

[20] There are obviously some important exceptions to this simplification, including tips, bonuses, commission, that likely contribute in important ways to the earnings process, and which I abstract

Figure 2.7: Biweekly Earnings Process for Main (Non-Rideshare) Jobs

(a) Earnings Deviations from Median (b) Impulse Response Function



Panel (a) shows the distribution of deviations of log earnings from median biweekly earnings calculated over the current year in the app data. Panel (b) shows the impulse response function for biweekly payroll income in the app data, calculated via the local projection described in the text. 95% confidence intervals based on standard errors clustered at the user level shown by dotted lines (they are very small and may not be visible). In both panels, the sample includes ever-rideshare drivers on strict biweekly paycycles and their matched coworkers. Values winsorized at the 1% level.

framework consistent with the high frequency process is long-term contracts with partial insurance (see, for instance, Lamadon, 2016).

To arrive at a tractable earnings process based on the underlying data, I proceed under the following assumption: hours and overtime pay are the primary source of volatility in the main job. I discretize the biweekly earnings changes in the app data and calculate an empirical Markov transition matrix, which is reported in Appendix Table B.2. I translate this into hours by assuming that at a deviation of 0, households work 80 hours (40/hours week). Above 80 hours, households are assumed to receive overtime pay at 1.5 times the wage. I estimate the average wage as the average of Median Earnings/80 in the app data.

Table 2.10 summarizes these assumptions and the calibrated parameters. The household discount rate, β, and the fixed costs of work, κ, remain to be estimated.

from. Appendix B.2 explores the hours process in household survey data, showing a substantial amount of variation in earnings comes from hours.

Table 2.10: Exogenous Model Parameters

Parameter	Value	Notes
α	$= 80^{-1/\eta}\overline{w_1 c}^{-\rho}$	$\overline{h} = 80$
ρ	1.38	(Gourinchas and Parker, 2002)
η	1.35	(Chen, Chevalier, Rossi, and Oehlsen, 2017; Angrist, Caldwell, and Hall, 2017)
$R_t = R$	$(1.001)^{1/26}$	Biweekly Calibration
w_1	\$15.53	App Data: Average($p50_i$(Earnings)/80)
$F(w_2)$	$\sim \mathcal{N}(15,3)$	Chen, Chevalier, Rossi, and Oehlsen (2017). - estimated costs
$F(h_1)$	Simulated Markov chain	App Data

Solution Method

Given a parameterization of the model, the solution is characterized by the optimal policy functions for consumption and hours. There are many ways to solve for the policy functions via numerical methods, e.g. value function iteration, Euler equation iteration, and endogenous grid methods. I proceed by backwards induction from time T using standard endogenous grid methods. As the horizon recedes, this will converge to the steady state solution under the assumptions (Deaton, 1991). The difference from the standard case without hours is that now cash on hand is a function of hours as well. The cases given by Equation (2.11) are used to solve implicitly for the hours policy.

Once I have the solved policy functions, I consider what happens when we go from a world with frictions of size κ_{pre} to $\kappa_{post} = 0$. I simulate the model for 1,000 agents, and match to a simple regression run on the simulated data using the variation from when rideshare turns on in a hypothetical city—the same variation used in my IV regression.[21] The SMM estimating equation is given by:

$$\min_{\theta} m(\theta)' V^{-1} m(\theta) \tag{2.21}$$

[21] As soon as a household gets access to the new policy function, consumption will immediately jump. This is inconsistent with my empirical results, which showed a jump in consumption only upon receipt of first income. In the real world, adoption lags after Uber enters a city. This may be due to information frictions—households will take time to learn about about ridesharing, perhaps through someone in their network who also drives or after seeing an advertisement.

where $m(\theta) = b_k - \beta_k(\theta)$. b_k are the reduced-form estimates that I target. I will match the model to my OLS estimates from Column 1 of Table 2.5 and three main IV estimates: instrumenting with leave-out mean firm earnings (Column 5 of Table 2.5), instrumenting with Uber's launch (Column 2 of Table 2.7), and the full instrument set (Column 3 of Table 2.7). $\beta_k(\theta)$ are the corresponding model moments, conditional on a parameter vector, $\theta = [\beta, \kappa]$. The (inverse) estimated variance-covariance matrix from the reduced-form regressions is used as the weighting matrix.

Welfare

The welfare gains from reducing the fixed cost of work will generally be less than κ due to non-participation. In addition, the consumption gains are also not sufficient for welfare analysis, because increased labor supply in the post period decreases welfare. To answer the question, "How much would the household be 'willing to pay' in the pre-period to eliminate the fixed costs of a second job," I consider a measure of consumption equivalence variation, defined as the value of ω that solves:

$$V((1+\omega)c^{pre*}, h^{pre*}) = V(c^{post*}, h^{post*}) \tag{2.22}$$

where $c^{pre*}, h^{pre*}, c^{post*}, h^{post*}$ are the converged (infinite-horizon) consumption and hours rules in the pre and post periods, respectively. The consumption and hours policies will depend on the distribution of constrained hour draws in the main job, the distribution of second wage draws, and the distribution of assets. Average willingness-to-pay is calculated by aggregating over these distributions:

$$\mathbb{E}[WTP] = \int\int\int \omega(A_{pre}; e_1; w_2) c^{pre*} f(e_1) f(w_2) f(A_{pre}) de_1 \; dw_2 \; dA_{pre} \tag{2.23}$$

I assume that the distribution of hour draws for the main job has not changed over the pre and post period. While rideshare jobs were not available in the "pre-period," I will assume, as in the main estimation, that a second job with the same wages was available, after paying the fixed cost of work. The distribution of assets will change in the pre and post-periods (the household will not need to acquire as many assets since they can now work if they receive a very low income draw). I use the pre-period asset distribution in this exercise.

Estimation Results

Estimation results are shown in Table 2.11. Columns (1), (3), (5) and (7) reproduce the targeted moments from earlier regressions. Column (2) shows the first set of results, matching to my OLS estimates of Column (1) of Table 2.5 (pre-period smoothing for ever-rideshare drivers is the summation of rows 1 and 2 of Table 2.5). Structural estimation can nearly perfectly match these coefficients with just the two parameters. κ_{pre} is estimated at \$391 per biweek and the discount factor β is estimated to be 0.965. Willingness-to-pay for removing these large fixed costs of work, calculated as described in Section 2.6, are reported in Row 8 of Table 2.11. Willingness to pay is estimated to be 59.90 per biweek. Column (4) shows the results from matching to the IV specification instrumenting with coworker earnings from Column (5) of Table 2.5. Now, κ_{pre} is estimated at \$536 per biweek and the discount factor β is estimated to be 0.96 and willingness to pay is estimated to be 70.90 per biweek.

Note that this is a biweekly β, so the annualized discount factor is $0.96^{26} = 35$ percent. While this is a considerable degree of impatience, the estimate is not far out of line with other empirical calibrations in the literature using high-frequency data. Ganong and Noel (2017) find that fitting data on UI benefit exhaustion requires 30 percent of agents be hand-to-mouth. Laibson, Maxted, Repetto, and Tobacman (2017) use credit card data and estimate a period-ahead discount factor of 0.504 and a long-term annualized discount factor of 0.987 (so-called β-δ discounting).

Column (6) reproduces the estimated coefficients from Column 2 of Table 2.7, instrumenting for Uber's launch, and Column (5) reports the corresponding estimation results. These coefficients are matched with a fixed cost of \$1,269, 2 times larger, with a more reasonable discount factor, 0.992, or 81.1 percent at an annualized rate. Column (8) reproduces the estimated coefficients from Column 3 of Table 2.7, instrumenting with both Uber's launch and coworkers' income, and Column (7) reports these estimation results. Weighting by the uncertainty of the estimates matches the consumption response in the pre-period, but underestimates the increased ability to smooth consumption. The estimation sets the fixed costs of work in the pre-period very high, at over \$4,000 per pay period, which has the effect of shutting off second hours in the pre-period.[22] Given the uncertainty in my IV estimates, I cannot reject equality across the IV specifications. For this reason, I proceed treating the estimates in Column (4)—the most conservative results based on IV estimates—as the baseline.

To give an idea as to the stability of these structural estimates to different parameterizations, Table 2.12 shows the results for a range of ρ (the inverse elasticity of substitution) and η (labor supply elasticity) values that are typically seen in the literature. The estimates for κ range between \$240 and \$1,520 dollars. Willingness to pay is much less dispersed: between \$60 and \$90 per biweek. Again, this is be-

[22] The full IV-estimates can be easily matched if we assume a structural break in the distributions of main job earnings or second job earnings, or a household preference shock.

Table 2.11: Model Estimation Results

	(1) Target	(2) Model	(3) Target	(4) Model	(5) Target	(6) Model	(7) Target	(8) Model
LogPay	0.3432	0.3431	0.3730	0.3730	0.3162	0.3162	0.3769	0.3929
Log Pay × Post Rideshare	-0.0677	-0.0704	-0.0908	-0.0920	-0.1211	-0.1231	-0.3183	-0.2061
Post Rideshare	0.5380	0.5380	0.6980	0.6981	0.9681	0.9338	2.3677	1.5329
Fixed Cost of Work, κ_{pre}	-	390.9	-	536.0	-	1,2698.9	-	4,554.8
(S.E.)	-	(7.59)	-	(3.1)	-	(1.1)	-	(38.3)
Discount Factor, β	-	0.965	-	0.9604	-	0.9923	-	0.9935
(S.E.)	-	(0.0002)	-	(0.0005)	-	(0.0005)	-	(0.0084)
WTP	-	59.91	-	70.90	-	147.31	-	188.80
V^{-1} Weighting Matrix	-	X	-	X	-	X	-	X

Columns (1), (3), (5) and (7) are targeted moments, and Columns (2), (4), (6) and (8) are estimation results. Standard errors in parentheses based on numerical Jacobians.

Table 2.12: Estimation Results for Range of Alternative Parameter Values

			ρ		
			1	1.5	2
η	0.5	κ_{pre}	243.61	340.06	368.67
		β	0.9740	0.9531	0.9336
		WTP	87.37	79.85	72.30
	1	κ_{pre}	310.97	449.67	537.97
		β	0.9750	0.9545	0.936
		WTP	84.03	69.66	62.6
	1.5	κ_{pre}	481.35	730.59	945.40
		β	0.9760	0.956	0.9380
		WTP	88.52	74.22	68.16
	2	κ_{pre}	650.86	1,151.68	1,524.47
		β	0.9762	0.9564	0.9400
		WTP	90.48	85.19	80.02

The table shows the structural estimates matching to the coefficients from the IV specification reproduced in Column (4) of Table 2.11, for the given η and ρ indicated in the table.

cause high values of the fixed cost do not necessarily map to welfare; the household can simply choose not to participate. Examining the grid of results, a number of interesting patterns emerge. First, the estimated costs are strictly increasing in η and ρ. Second, estimated β's are decreasing in ρ, but are relatively stable across $\eta' s$. Swanson (2012) shows that relative risk aversion for these preferences is given by: $1/(\rho^{-1} + \eta)$. Risk aversion is thus increasing in ρ and decreasing in η. If a household wants to smooth consumption using assets (high ρ) or labor supply (high η), then it must be the case that the fixed costs are large to generate big increases in consumption smoothing. The willingness to pay to eliminate the costs are decreasing for $\eta < 1$, but then increase as the costs rise and the household wants to smooth more.

Policy Functions

I plot the policy functions for consumption and hours implied by the baseline parameter estimates in Figure 2.8. Panel (a) shows consumption as a function of cash on hand (which is itself a function of hours) for the pre- (red) and post-periods (green). I consider two cases: a 20 percent cut in main hours (lines indexed by "o") and steady state hours (lines indexed by "+"). The consumption policy looks similar in the steady state, but is now considerably higher for a negative deviation in total hours. In addition, the probability of being at a particular value of cash on hand will be different because households can now control hours. The distribution of cash on hand in each period is shown below the consumption policy functions. In the post period, the household is less likely to be on the hand-to-mouth portion of the consumption function where consumption is equal to income.

Panel (b) shows the hours policy function. The hours decision is very different in the pre and post periods. In the pre period, households do not participate in the second labor market for shocks in the main job (up to 20 hours) due to the high costs of taking a second job. After reducing the cost of κ, households participate in the second job when they face a drop in hours in their main job. In the low asset state (green line indexed by "o"), households will work even more for the same hours deviation.

Impulse Response Functions

To illustrate model dynamics, I next consider an experiment where the household receives a one-time 20-hour cut in hours in their main job from the steady state. The results of this experiment are shown in Figure 2.9.

The impulse response functions for the regime before flexible jobs are indexed by "o". We see that consumption falls with assets. Given the costs of participating in the second job, it is not beneficial for the household to participate. The final panel shows the Lagrange multiplier on the borrowing constraint, interpretable as the gain in marginal utility from relaxing the borrowing constraint.

The world where labor supply can be increased frictionlessly in second jobs is indexed by "+". By comparison, consumption is far less volatile. We see that nearly all the loss in hours in the main job is compensated with the second job. The welfare losses from the borrowing constraint are roughly one-third the size for this shock to main hours.

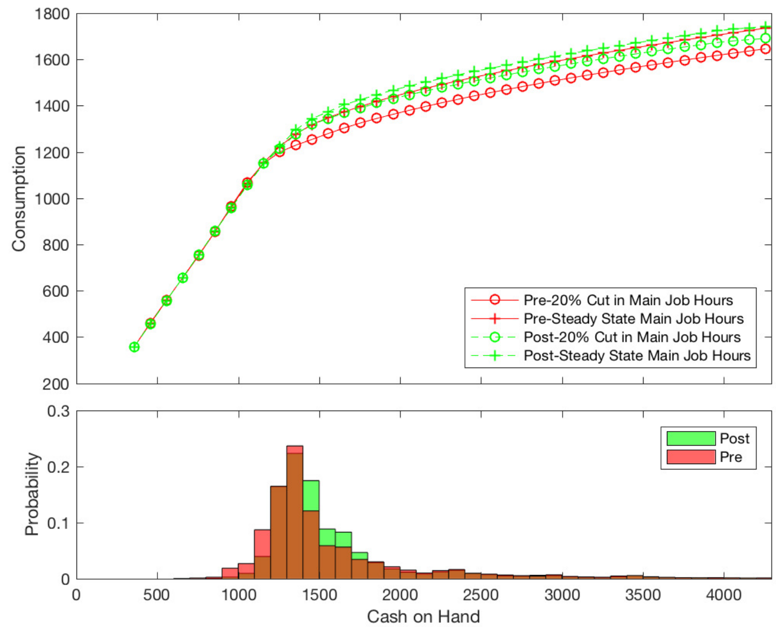
Counterfactuals

Finally, I use the model to explore a number of counterfactuals that place the willingness to pay estimates in context. First, I calculate the willingess to pay to eliminate all negative shocks in the main job.[23] Results are reported in the first row of Table

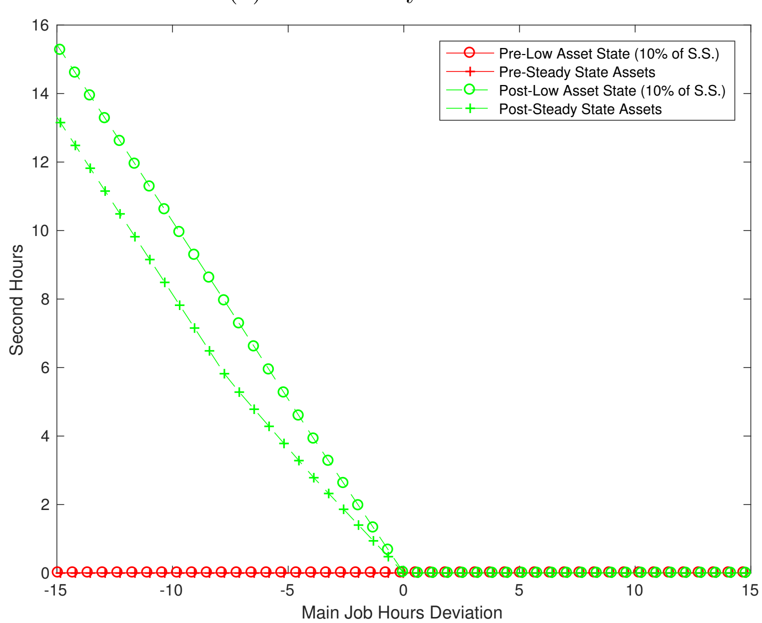
[23] The model also includes positive shocks (overemployment). For small positive shocks, the household is actually better off because of the overtime premium. Large positive shocks, although rare, generate welfare losses. On net, the household would be worse off by about 2 percent if positive

Figure 2.8: Policy Functions

(a) Consumption Policy Function

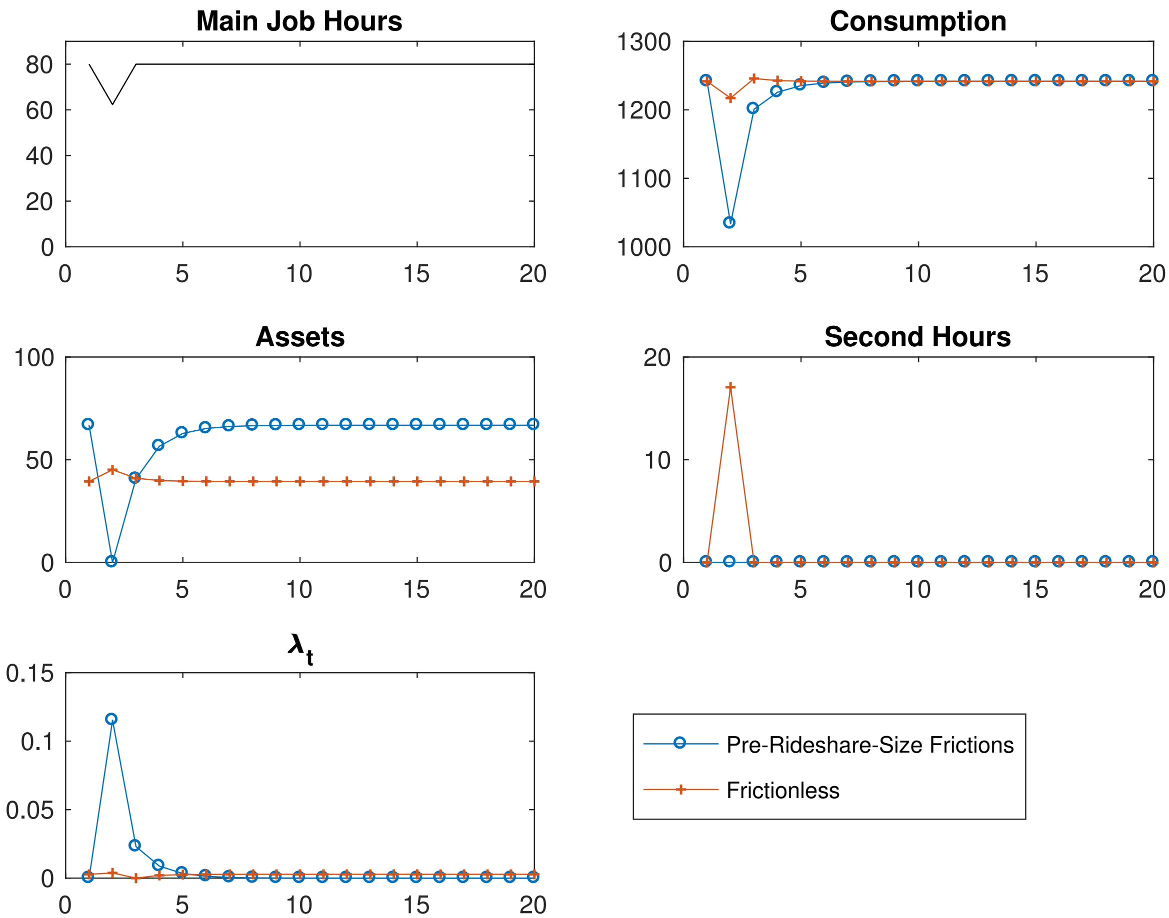


(b) Hours Policy Function



Policy functions use the estimated parameterization from Column (4) of Table 2.11.

Figure 2.9: Impulse Response Functions



Impulse response functions for a 20 hour shock to main earnings. The model is parameterized with the estimates from Column (3) of Table 2.11.

Table 2.13: Counterfactuals

Counterfactual	WTP ($)
Eliminating All Negative Shocks	143.51
Eliminating Borrowing Constraint	105.10
Eliminating Fixed Costs of Work	**70.90**
Eliminating Extreme Negative Shocks (Bottom 20%)	68.60
Increasing borrowing constraint by $100	21.80

Counterfactuals are for the estimated parameterization from Column (4) of Table 2.11.

2.13. The household would be willing to pay 143.51 per pay period, on average, to eliminate all negative shocks, or $3,700 per year. Next, I calculate how much the household would be willing to pay to eliminate the borrowing constraint. The inability to borrow generates a welfare loss of $105.10 (Row 2). Households would be willing to pay $21.80 to increase the borrowing constraint by $100, approximately 8 percent of after-tax biweekly earnings (Row 5). This large willingess-to-pay to reduce the borrowing constraint can provide insight into why some households might undertake costly credit card debt or payday loans. Recall that eliminating the fixed cost of work was valued at $70.90 per biweek in my baseline estimates, which is 67 percent of the gains from completely eliminating the borrowing constraint. As one more point of reference, I calculate the welfare gains just from removing the bottom 25 percent of shocks. This is valued at $68.60, approximately the same as gaining access to costless second jobs.

2.7 Conclusion

The typical worker that selects into ridesharing appears to be using a flexible job to mitigate volatility in a main job. In the period after starting rideshare, rideshare income replaces 73 percent of income losses from main payroll jobs. The link between spending and main income, which is around 1/3 in the population, declines by 82 percent in my specification exploiting the staggered geographic entry of rideshare and income movements common to all workers at the firm. When these moments are matched to a structural model with labor supply frictions, biweekly fixed costs of work are estimated at over $500. In my preferred specification, households would

shocks were eliminated.

be willing to pay $70.90 per week or $1,800 per year, to eliminate these costs. Even though I focus on ridesharing employment, the benefits of flexibility should extend to any second job with hours flexibility and limited search and transaction costs.

This study has two important implications. The first is for the welfare costs of income fluctuations. The large fixed costs of second work estimated in this paper provide an alternative explanation for why households may not seek to increase hours when faced with "small" shocks. One interpretation has been that households are fully insured with assets/savings and highly value their leisure. This paper qualifies this statement to be that leisure is more valuable than the costs of finding additional employment.

Second, this paper provides insight on *when* flexible work is valuable. A recent study by Mas and Pallais (forthcoming) finds that workers prefer stability over flexibility in their job arrangements. In reality, workers often do not have complete control over their hours from week to week. When faced with volatile incomes in main jobs and credit market imperfections, flexible jobs can be valuable. When credit constrained, using labor supply instead of assets to smooth transitory shocks is a "second-best" way to smooth because of the disutility of work. However, the availability of flexible labor supply can provide substantial benefits in the presence of credit market imperfections. I estimate that the welfare gains from eliminating costly second jobs is about 2/3 of the gains from completely eliminating the borrowing constraint.

One relevant question for welfare is whether the benefits estimated in this paper are simply a transfer from the incumbent sector, taxi drivers. Appendix B.1 examines wages, hours and earnings in the taxi industry, finding no apparent impact on taxi drivers (although the value of taxi medallions has notably fallen).[24] A second policy concern is that work in ridesharing and related industries operates largely outside of the existing legal framework governing employment. New policies have been proposed to extend certain existing employment provisions to non-traditional employment relationships,[25] and the legal definition of an "employee" is currently being debated in the courts.[26] To the extent that these policies limit flexibility, they could end up hurting workers using flexible work as a consumption smoothing mechanism; at the same time, work practices that gain an edge from operating outside a regulatory framework could also put workers and others at risk.

[24] Hu, Winnie, "Taxi Medallions, Once a Safe Investment, Now Drag Owners Into Debt," *The New York Times*. 9/10/2017.

[25] See, for instance, (Harris and Krueger, 2015)

[26] For a discussion of ongoing litigation, see Isaac, Mike and Noam Scheiber. April 21, 2016."Uber Settles Cases With Concessions, but Drivers Stay Freelancers," *The New York Times* Link

The welfare estimates from this paper depend on a variety of assumptions, one of which is the earnings process in the second job. If wages fall, because monopsonistic platforms set lower wages, or because more workers enter the sector, driving wages down, then obviously so will the benefits for workers. For the latter reason, ridesharing is likely a better smoothing mechanism for idiosyncratic, rather than aggregate shocks. As the rideshare industry and related sectors continue to grow, the policy concerns and general equilibrium implications just highlighted will likely become more relevant.

Chapter 3

The Response of Consumer Spending to Changes in Gasoline Prices[1]

[1]This work is joint with Michael Gelman, Yuriy Gorodnichenko, Shachar Kariv, Mathew Shapiro, Dan Silverman, and Steve Tadelis. An online appendix is available at http://www-personal.umich.edu/ shapiro/papers/gasprices_appendix.pdf

3.1 Introduction

Few macroeconomic variables grab headlines as often and dramatically as do oil prices. In 2014, policymakers, professional forecasters, consumers and businesses all wondered how the decline of oil prices from over $100 per barrel in mid-2014 to less than $50 per barrel in January 2015 would influence disposable incomes, employment, and inflation. A key component for understanding macroeconomic implications of this shock is consumers' spending from the considerable resources freed up by lower gasoline prices (the average saving was more than $1,000, or approximately 2 percent of total spending per household).[2] Estimating the quantitative impact of such changes is central to policy decisions. Yet, because of data limitations, a definitive estimate has proved elusive. Recently, big data has opened unprecedented opportunities to shed new light on the matter. This paper uses detailed transaction-level data provided by a personal financial management service to assess the spending response of consumers to changes in gasoline prices over the 2013-2016 period.

Specifically, we use this information to construct high-frequency measures of spending on gasoline and on non-gasoline items for a panel of more than half a million U.S. consumers. We use cross-consumer variation in the intensity of spending on gasoline interacted with the large, exogenous, and permanent decline in gasoline prices to identify and estimate the partial equilibrium marginal propensity to consume (MPC) out of savings generated by reduced gasoline prices. Given the low elasticity of demand for gasoline and the nature of the oil price shock, one can think of this MPC as measuring the response of spending to a permanent, unanticipated income shock. Our baseline estimate of the MPC is approximately one. That is, consumers on average spend all of their gasoline savings on non-gasoline items. There are lags in adjustment, so the strength of the response builds over a period of weeks and months.

Our results are useful and informative in several dimensions. First, our estimate of the MPC is largely consistent with the permanent income hypothesis (PIH), a theoretical framework that became a workhorse for analyses of consumption, and that has been challenged in previous studies. Second, our findings suggest that, ceteris paribus, falling oil prices can give a considerable boost to the U.S. economy via increased consumer spending (although other factors can offset output growth). Third, and also consistent with the PIH, we show that consumers' liquidity was not important for the strength of the consumer spending response to gasoline price shocks. Fourth, our analysis highlights the importance of having high-frequency

[2]According to the U.S. Consumer Expenditure Survey, average total household spending in 2014 was $53,495 total, while the average household spending on gasoline was $2,468.

transaction data at the household level for estimating consumer reactions to income and price shocks.

This paper is related to several strands of research. The first strand, surveyed in Jappelli and Pistaferri (2010), is focused on estimating consumption responses to income changes. Typically, studies in this area examine if and how consumers react to anticipated, transitory income shocks and, like our approach, provide the partial equilibrium response of household spending to a shock, not the general equilibrium outcome for aggregates. A common finding in this strand of research is that, in contrast to predictions of the PIH, consumers often spend only upon the realization of an income shock, rather than upon its announcement, although the size of this excess sensitivity depends on household characteristics. Baker (forthcoming) and Kueng (2015) document this pattern using data similar to what we study here and Gelman, Kariv, Shapiro, Silverman, and Tadelis (2015) report it for the same data source that we use.

At the same time, estimating spending responses to unanticipated, highly persistent income shocks has been challenging, because identifying such shocks is particularly difficult.[3] Indeed, we are not aware of an estimate of the MPC from this kind of shock. Thus, in sharp contrast to existing literature, we possibly provide the first estimate of MPC for an unanticipated, permanent shock to income. To this end, we exploit a particularly clear-cut source of variation in household budgets (spending on gasoline) with a number of desirable properties. Specifically, we use a large, salient, unanticipated, permanent (or perceived to be permanent) shock. We examine spending responses at the weekly frequency while, due to data limitations, the vast majority of previous studies estimate responses at much lower frequencies. As we discuss below, the high-frequency dimension allows us to obtain crisp estimates of the MPC and thus provide a more informative input for policy making.

The second strand to which we contribute studies the effects of oil prices on the economy. In surveys of this literature, Hamilton (2008) and Kilian (2008) emphasize that oil price shocks can influence aggregate outcomes via multiple channels (e.g., consumer spending, changes in expectations) but disruption of consumers' (and

[3]Previous studies examined responses of consumption to highly persistent income shocks due to job displacements (e.g. Stephens, 2001) or health (e.g. Gertler and Gruber, 2002). However, these income shocks are likely combined with other changes in the lives of affected consumers which makes identification of MPC challenging. An alternative strategy is to use statistical decompositions in spirit of Jappelli and Pistaferri (2006) but these estimates of MPC may depend on the assumptions of statistical models. Changes is taxes may provide a useful source of variation (see e.g. Neri, Rondinelli, and Scoccianti, 2017) but it is often hard to identify the timing of these shocks (tax changes are typically announced well before the changes are implemented) and the persistence of shocks (tax changes could be reversed with a change in government).

firms') spending on goods other than energy is likely to be a key mechanism for amplification and propagation of the shocks. Indeed, given the low elasticity of demand for gasoline, changes in gasoline prices can materially affect non-gasoline spending budgets for a broad array of consumers. As a result, a decrease in gasoline prices can generate considerable savings for consumers which could be put aside (e.g., to pay down debt or save) or used to spend on items such as food, clothing, furniture, etc.

Despite the importance of the MPC out of gasoline savings, research on the sensitivity of consumer non-gasoline spending to changes in the gasoline price has been scarce. One reason for the scarcity of research on the matter has been data limitations. Available household consumption data tend to be low frequency, whereas consumer spending, gasoline prices, and consumer expectations can change rapidly. For example, the interview segment of the U.S. Consumer Expenditure Survey (CEX) asks households to recall their spending over the previous month. These data likely suffer from recall bias and other measurement errors that could attenuate estimates of households' sensitivity to changes in gasoline prices (see Committee on National Statistics, 2013). The diary segment of the CEX has less recall error, but the panel dimension of the segment is short (14 days), making it difficult to estimate the consumer response to a change in prices. Because the CEX is widely used to study consumption, we do a detailed comparison of our approach using the app data with what can be learned from using the CEX. We find that analysis of the CEX produces much noisier estimates.

Grocery store barcode data, such as from AC Nielsen, have become a popular source to measure higher-frequency spending. These data, however, cover only a limited category of goods. For example, gasoline spending by households is not collected in AC Nielsen, making it impossible to exploit heterogeneity in gasoline consumption across households. As a result, most estimates of MPC tend to be based on time series variation in aggregate series (see e.g. Edelstein and Kilian, 2009).

There are a few notable exceptions. Using loyalty cards, Hastings and Shapiro (2013) are able to match grocery barcode data to gasoline sold at a large grocery store retailer with gasoline stations on site. We show that households typically visit multiple gasoline station retailers in a month, suggesting limitations to focusing on consumer purchases at just one retailer. There is also some recent work using household data to identify a direct channel between gasoline prices and non-gasoline spending. Gicheva, Hastings, and Villas-Boas (2010) use weekly grocery store data to examine the substitution to sale items as well as the response of total spending. They find that households are more likely to substitute towards sale items when gasoline prices are higher, but they must focus only on a subset of goods bought in grocery stores (cereal, yogurt, chicken and orange juice), making it difficult to extrapolate.

Perhaps the closest work to ours is a policy report produced by the Farrell and Greig (2015), which also uses "big data" to examine the response of consumers to the 2014 fall in gasoline prices, and finds an average MPC of approximately 0.6. This report differs from our study in both its research design and its data. Most importantly, our data include a comprehensive view of spending, across many credit cards and banks. In contrast, the Chase report covers a vast number of consumers, but information on their spending is from Chase accounts only. If, for example, consumers use a non-Chase credit card or checking account, any spending on that account would be missed in the J.P. Morgan Chase Institute analysis, and measurement of household responses may therefore be incomplete. In this paper, we confirm this by showing that an analysis based on accounts in one financial institution leads to a significantly attenuated estimate of the response of spending to changes in gasoline prices.

This paper proceeds as follows. Section II describes trends in gasoline prices, putting the recent experience into historical context. In Section III, we discuss the data, Section IV describes our empirical strategy, and Section V presents our results. Specifically, we report baseline estimates of the MPC and the elasticity of demand for gasoline. We contrast these estimates with the comparable estimates one can obtain from alternative data. In Section V we also explore robustness of the baseline estimates and potential heterogeneity of responses across consumers. Section VI concludes.

3.2 Recent Changes in Gasoline Prices: Unanticipated, Permanent and Exogenous

In this section, we briefly review recent dynamics in the prices of oil and gasoline and corresponding expectations of future prices. We document that the collapse of oil and gasoline prices in 2014-2015 was highly persistent, unanticipated, and exogenous to demand conditions in the United States. These properties of the shock are important components of our identification strategy.

Unanticipated and Permanent

In Panel A of Figure 3.1, the solid black line shows the spot price of gasoline at New York Harbor, an important import and export destination for gasoline. The New York Harbor price is on average 70 cents lower than average retail prices, although the two series track each other very closely. The dashed line shows the one-year-ahead futures price for that date. The futures price tracks the spot price closely, suggesting the market largely treats gasoline price as a random walki.e., the best prediction for one-year-ahead price is simply the current price.

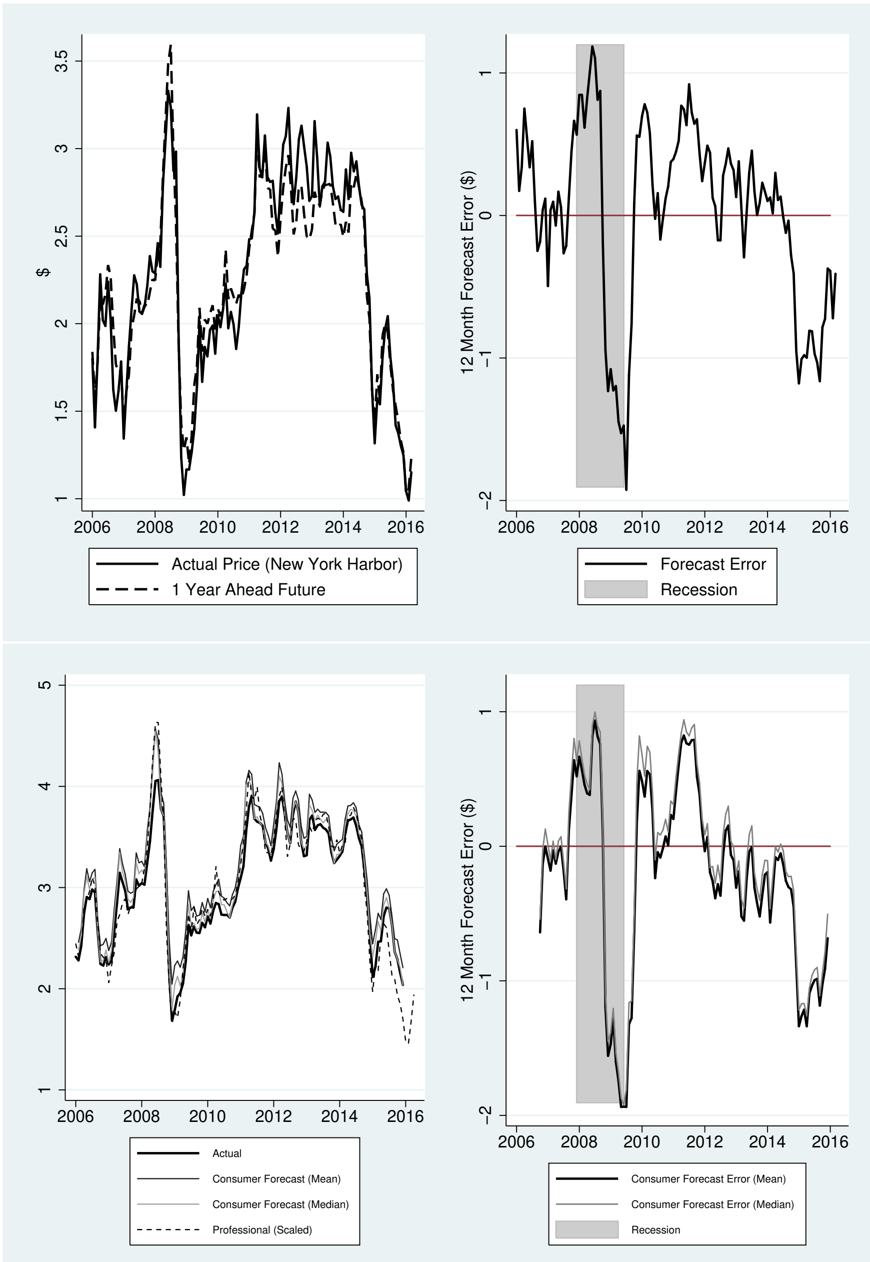
Panel B shows the difference between the realized and predicted spot price. The behavior of one-year-ahead forecast errors indicates that financial markets anticipated neither the run-up nor the collapse of gasoline prices in 2007-2009. Likewise, the dramatic decline in gasoline prices in 2014-2015 was not anticipated. The Michigan Survey of Consumers has asked households about their expectations for changes in gasoline prices over the next one-year and five-year horizons. Panel C of Figure 1 plots the mean and median consumer expectations along with the actual price and the mean one-year-ahead prediction in the Survey of Professional Forecasters. While consumers expect a slightly higher price relative to the present price than professional forecasters, the basic pattern is the same as in Panel A: the current price appears to be a good summary of expected future prices. Consistent with this observation, Anderson, Kellogg and Sallee (2012) fail to reject the null of a random walk in consumer expectations for gasoline prices. Thus, consumers perceive changes in gasoline prices as permanent. Also similar to the financial markets, consumers were not anticipating large price changes in 2007-2009 or 2014-2015 (Panel D).

Figure 3.1 shows large movements in prices during the Great Recession (shaded). Unlike the recent episode that is the subject of this paper, we would not use it to identify the MPC because this fluctuation in commodity prices in the Great Recession surely represents an endogenous response to aggregate economic conditions.

When put into historical context, the recent volatility in gasoline prices is large. Table 3.1 ranks the largest one-month percent changes in oil prices since 1947. When available, the change in gasoline prices over the same period is also shown.[4] The price drops in 2014-2015 are some of the largest changes in oil and gasoline prices in the last 60 years. Note that in 1986, gasoline prices and oil prices actually moved in opposite directions, indicating that the process generating gasoline prices can sometimes differ from oil.

[4]Oil spot prices exist back to 1947, while the BLS maintains a gasoline price series for urban areas back to 1976. In our analysis, we use AAA daily gasoline prices retrieved from Bloomberg (3AGSREG). The series comes from a daily survey of 120,000 gasoline stations. These data almost perfectly track another series from the EIA which are point in-time estimates from a survey of 900 retail outlets as of 8am Monday.

Figure 3.1: Gasoline prices and expectations



Notes: Panel A shows the New York Harbor spot price, as the 1 year ahead future price. Panel B shows the 1 year ahead forecast error, defined as the difference between the realization of the spot price and the forecast 1 year earlier. Panel C shows the gasoline price, and the weighted mean and median expectations from the Michigan Surveys of Consumers. See https://data.sca.isr.umich.edu/sda-public/cgi-bin/hsda?harcsda+sca. In the survey, households are asked, About how many cents per gallon do you think gasoline prices will (increase/decrease) during the next twelve months compared to now? We add the household response to this question to the current gasoline price. We also plot the futures used in panel A, scaled by the average difference between the spot price and retail price over the period (dotted line). Panel D shows retail gasoline prices and the consumer forecast made 12 months earlier. Shaded area is the Great Recession.

Table 3.1: Largest monthly changes in oil and gasoline prices

Largest Decreases					
Date	Percent Change		Date	Percent Change	
	Oil	Gas		Oil	Gas
1986:2	-33	-6	1974:1	135	
2008:12	-28	-21	1990:8	47	10
2008:10	-26	-14	1986:8	30	-5
2008:11	-25	-32	1948:1	24	
2014:12	-22	-11	1990:9	23	9
2015:1	-20	-18	2009:3	23	1

Table shows the month-to-month percent change in West Texas Intermediate spot oil prices (FRED series OILPRICE and MCOILWTICO) and the corresponding change in average monthly regular gasoline prices, when available, from January 1946 February 2016. For gasoline prices, the table use the BLS U.S. city average (BLS series APU000074714), since it is available further back in time than other available gasoline price data.

Exogenous

Why did prices of oil and oil products such as gasoline fall so much in 2014-2015? While many factors could have contributed to the dramatic decline in the prices, the consensus view, summarized in Baffes, Kose, Ohnsorge, and Stocker (2015), attributes a bulk of the decline to supply-side factors. Specifically, this view emphasizes that key forces behind the decline were, first, OPEC's decision to abandon price support and, second, rapid expansion of oil supply from alternative sources (shale oil in the U.S., Canadian oil sands, etc.). Consistent with this view, other commodity prices had modest declines during this period, which would not have happened if the decline in oil prices was driven by global demand factors. Observers note that the collapse of oil prices in 2014-2015 is similar in many ways to the collapse in 1985-1986, when more non-OPEC oil supply came from Mexico, the North Sea and other sources, and OPEC also decided to abandon price support. In short, available evidence suggests that the 2014-2015 decline in oil prices is a shock that was supply-driven and exogenous to U.S. demand conditions. In contrast, Hamilton (2009) and others observe that the run up in oil and gasoline prices around 2007-2009 can be largely attributed to booming demand, stagnant production, and speculators, and the consequent decline of the prices during this period, to collapsed global demand (e.g. the Great Recession and Global Financial Crisis).

3.3 Data

Our analysis uses high-frequency data on spending from a financial aggregation and bill-paying computer and smartphone application (henceforth, the app).[5] The app had approximately 1.4 million active users in the U.S. in 2013.[6] Users can link almost any financial account to the app, including bank accounts, credit card accounts, utility bills, and more. Each day, the app logs into the web portals for these accounts and obtains central elements of the user's financial data including balances, transaction records and descriptions, the price of credit and the fraction of available credit used. Using data for a similar service, (Baker, forthcoming) documents that over 90 percent of users link all their checking, savings, credit card, and mortgage accounts. Given the non-intrusive automatic data collection, attrition rates are moderate (approximately five percent per quarter).

We draw on the entire de-identified population of active users and data derived from their records from January 2013 until February 2016. The app does not collect demographic information directly and, thus, we ar unable to study heterogeneity in responses across demographic groups or to use weights or similar methods to correct possible imbalances in the population of the app's users. However, for a subsample of users, the app employed a third-party that gathers both public and private sources of demographics, anonymizes them, and matches them back to the de-identified dataset. Table 1 in Gelman, Kariv, Shapiro, Silverman, and Tadelis (2014) (replicated in Online Appendix Table C2) compares the gender, age, education, and geographic distributions in a subset of the sample to the distributions in the U.S. Census American Community Survey (ACS), representative of the U.S. population in 2012. The app's user population is heterogeneous (including large numbers of users of different ages, education levels, and geographic location) and, along some demographic dimensions, contains proportions similar to those found in the US population. Consistent with this pattern, Baker (forthcoming) observes that, as the online industry had matured, the differences between the population of a similar app's users and the U.S. population became small by 2013.

[5]These data have previously been used to study the high-frequency responses of households to shocks such as the government shutdown (Gelman, Kariv, Shapiro, Silverman, and Tadelis, 2015) and anticipated income, stratified by spending, income and liquidity (Gelman, Kariv, Shapiro, Silverman, and Tadelis, 2014).

[6]All data are de-identified prior to being made available to the project researchers. Analysis is carried out on data aggregated and normalized at the individual level. Only aggregated results are reported.

Identifying Spending Transactions

Not every transaction reported by the app is spending. For example, a transfer of funds from one account to another is not. To avoid double counting, we exclude transfers across accounts, as well as credit card payments from checking accounts that are linked within the app. If an account is not linked, but we still observe a payment, we count this as spending when the payment is made. We identify transfers in several ways. First, we search if a payment from one account is matched to a receipt in another account within several days. Second, we examine transaction description strings to identify common flags like “transfer”, “tfr”, etc. To reduce the chance of double counting, we exclude the largest single transaction that exceeds $1,000 in a given week, as this kind of transaction is very heavily populated by transfers, credit card payments, and other non-spending payments (e.g., payments to the U.S. Internal Revenue Service). We include cash withdrawals from the counter and ATM in our measure of spending. To ensure that accounts in the app data are reasonably linked and active, we keep all users who were in the data for at least 8 weeks in 2013 and who did not have breaks in their transactions for more than two weeks. More details are provided in Online Appendix A.

Using Machine Learning to Classify Type of Spending

Our analysis requires classification of spending by type of goods. To do so, we address several challenges in using transactional data from bank accounts and credit cards. First, transactional data are at the level of a purchase at an outlet. For many purchases, a transaction will include many different goods. In the case of gasoline, purchases are carried out mainly at outlets that exclusively or mainly sell gasoline. Hence, gasoline purchases are relatively easy to identify in transactional data. Second, for the bulk of transactions in our data, we must classify the outlet from the text of the transaction description, rather than classifications provided by financial institutions. We therefore use a machine learning (ML) algorithm to classify spending based on transaction descriptions. In this section, we provide an outline of the classification routine, and compare our ML predictions in the data provided by the app with external data. As economic analysis increasingly uses naturally-occurring transactional data to replace designed survey data, applications of ML like the one we use will be increasingly important.

The ML algorithm constructs a set of rules for classifying the data as gasoline or non-gasoline. This requires a training data set to build a classification model, and a testing data set not used in the training step to validate the model predictions. Two of the account providers in the data classify spending directly in the transaction description strings, using merchant category codes (MCCs). MCCs are four digit codes used by credit card companies to classify spending and are also recognized by the U.S. Internal Revenue Service for tax reporting purposes. Our main MCC of interest is 5541, Automated Fuel Dispensers. Purchases of gasoline could also fall into MCC code 5542, Service Stations, which in practice covers gasoline stations with convenience stores.[7] We group transactions with these two codes together because distinguishing transactions as 5542 or 5541 without the MCC is nearly impossible with only the transaction descriptions.[8]

A downside of this approach is that transactions at a Service Station may either be for gasoline, for food or other items, or both. According to the National Association of Convenience Stores (NACS), which covers gasoline stations, purchases of non-gasoline items at gasoline stations with convenience stores (i.e. Service Stations) account for about 30 percent of sales at "Service Stations. Although the app data do not permit us to differentiate gasoline and non-gasoline items at Service Stations, we can use transaction data from "Automated Fuel Dispensers (which do not have an associated convenience store), as well as external survey evidence to separate purchases of non-gasoline items from purchases of gasoline. Specifically, according to the 2015 NACS Retail Fuels Report (NACS 2015), 35 percent of gasoline purchases are associated with going inside a gasoline station's store. Conditional on going inside the store, the most popular activities are to "pay for gasoline at the register" (42%), "buy a drink (36%), "buy a snack" (33%), buy cigarettes (24%), and buy lottery tickets" (22%). The last four items are likely to be associated with relatively small amounts of spending. This conjecture is consistent with the distribution of transactions for "Service Stations" and "Automated Fuel Dispensers" in the data we study. In particular, approximately 60 percent of transactions at "Service Stations" are less than $10 while the corresponding share for "Automated Fuel Dispensers" is less than 10 percent. As we discuss below, the infrequent incidence of gasoline purchases totaling less than $10 is also consistent with other data sources. Thus, we exclude Service Stations transactions less than $10 to filter out purchases of non-gasoline items. Using one of the two providers with MCC information (the one with more data), we train a Random Forest ML model to create binary classifications of transactions into those made at a gasoline station/service station and those that were made elsewhere. Figure 2 shows an example of decision trees used to classify transactions into gasoline and non-gasoline spending. A tree is a series of rules that train the model to classify a purchase as gasoline or not. The rules minimize

[7] "Service Stations" do not include services such as auto repairs, motor oil change, etc.

[8] E.g., a transaction string with word "Chevron" or "Exxon" could be classified as either MCC 5541 or MCC 5542.

the decrease in accuracy when a particular model "feature," in our case transaction values and words in the transaction strings, is removed. In the Figure 3.2 example, the most important single word is "oil." If a transaction string contains the word oil, the classification rule is to move to the right, otherwise the rule is to move to the left. If the string does not contain the word oil, the next most important single word is "exxonmobil." Figure 3.2 also demonstrates how the decision tree combines transaction string keywords with transaction amounts. For example, oil is a very strong predictor of gasoline purchase but it can be further refined by the transaction amount. The tree continues until all the data are classified.

We then use the second provider to validate the quality of our ML model.[9] The ML model is able to classify spending with approximately 90% accuracy in the testing data set, which is a high level of precision. Both Type I and Type II error rates are low. See online Online Appendix Table B.1. More details on the procedure can be found in Online Appendix B.

We can also use the app data to investigate which gasoline stations consumers typically visit. The top ten chains of gasoline stations in the app data account for most of gasoline spending. On average, the app data suggest that the typical consumer does 66 percent of his or her gasoline spending in one chain and the rest of gasoline spending is spread over other chains. Thus, while for a given consumer there is a certain degree of concentration of gasoline purchases within a chain, an analysis focusing on only one gasoline retailer, such as in Gicheva, Hastings, and Villas-Boas (2010) or Hastings and Shapiro (2013), particularly one not in the top ten chains, would miss a substantial amount of gasoline spending.

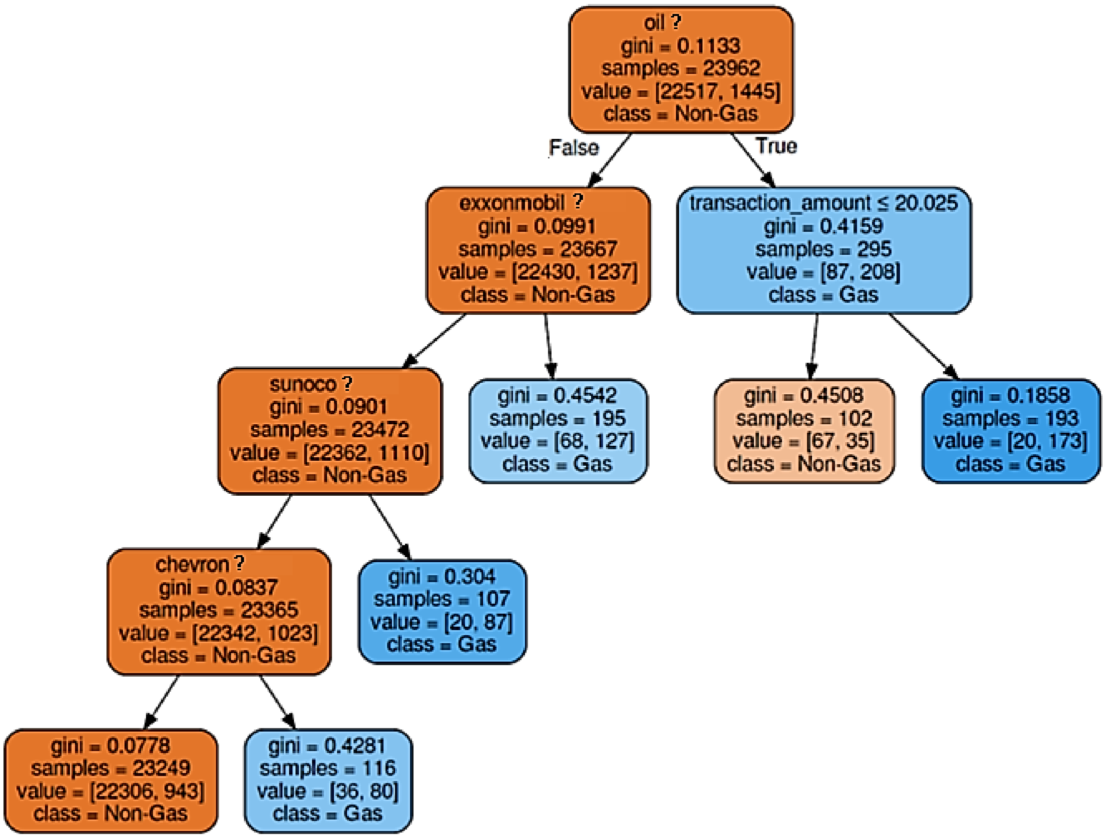
Comparison with the Consumer Expenditure Survey

We compare our measures of gasoline and non-gasoline spending with similar measures from the Consumer Expenditure Survey (CEX).[10] We use both the CEX Diary Survey and Interview Survey. In the diary survey, households record all spending in written diaries for 14 days. Therefore, this survey provides an estimate of daily

[9] Card providers use slightly different transaction strings, and one may be concerned that training the model on a random subsample of data from both card providers, and testing it on another random subsample, can provide a distorted sense of how our ML model performs on data from other card providers. Thus, using a card from one account provider to train, and testing on an entirely different account provider, helps to assure that the ML model is valid outside of the estimation sample. Classification of transactions based on ML applied to both card providers yields very similar results.

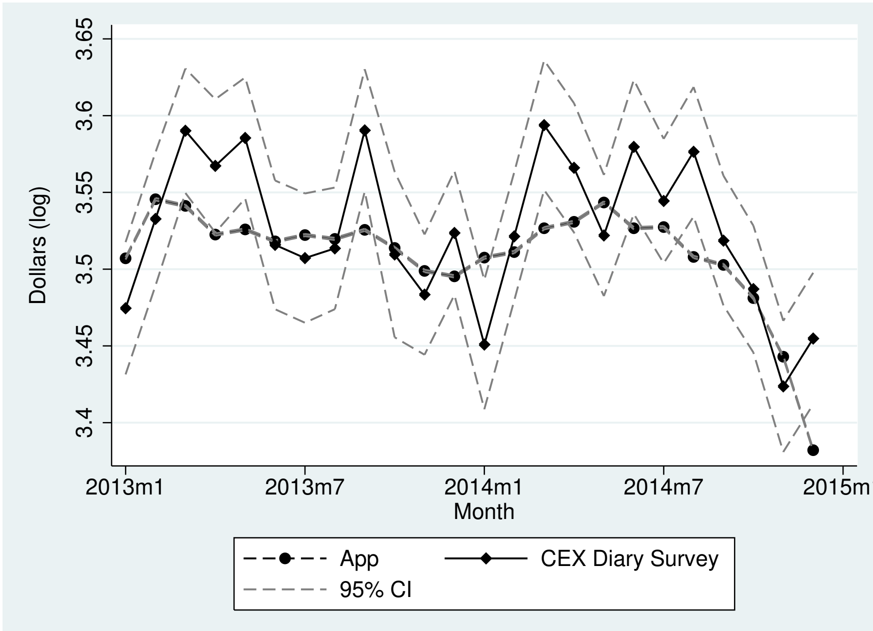
[10] While the definition of the spending unit is different in the CEX (household) and the app (user), Baker (2016) shows for a similar dataset that linked accounts generally cover the whole household.

Figure 3.2: An example machine learning decision tree



Notes: the figure shows an example of decision trees estimated on a training dataset used to classify transactions into gasoline and non-gasoline spending. Blue boxes represent classification into gasoline (class=Gas) purchases and orange boxes represent classification into non-gasoline purchases (class=Non-Gas). The shades indicate how strong of a predictor that feature is (darker shades mean stronger predictors). The first line inside the box refers to the "feature"either a particular word in the "bag of words," or a transaction amount cutoff, used as predictors in the model. The second line gives the gini value, which is a measure of impurity that the classification algorithm minimizes at every node with its choice of feature. "Oil" is the most important feature, based on the gini criteria, and so is chosen first. The "sample" line gives the remaining number of observations to be classified at the node (we start training with a dataset with 23,962 observations in this example). The "value" tells you how many of the samples fall into each category ([gas, not gas]) if you were to classify them based on the decision rules that have led you to the node. Once a branch reaches an end, or "leaf," the classification rule made is the classification with the maximum value. See Online Appendix B for more details.

Figure 3.3: Distribution of log gasoline spending: CEX Diary versus App



Notes: the figure shows the distribution of daily log spending on gasoline in the Diary segment of the Consumer Expenditure Survey (CEX) and in the app data. Gasoline spending in the app data is identified using machine learning (ML). App includes all transactions that ML identifies as purchases of gasoline. App¿$10 includes transactions that ML identifies as purchases of gasoline and that are greater than $10. See text for further details.

gasoline spending that should be comparable to the daily totals we observe in the app. In Figure 3.3, we compare the distribution of spending in our data (solid lines) and in the diary survey (dashed line). We find that the distributions are very similar, with one notable exception: the distribution of gasoline purchases in the app data has more mass below $10 (solid gray line) than the CEX Diary data. As we discussed above, this difference is likely to be due to our inability to differentiate gasoline purchases and non-gasoline purchases at "Service Stations." In what follows, we restrict our ML predictions to be greater than $10 (solid black line).

The CEX Diary Survey provides a limited snapshot of households' gasoline and other spending. In particular, since a household on average only makes 1 gasoline purchase per week in the diary, we expect only to observe 2 gasoline purchases per household, which can be a noisy estimate of gasoline spending at the household level. Idiosyncratic factors in gasoline consumption that might push or pull a purchase from one week to the next could influence the measure of a household's gasoline purchases by 50% or more. In addition, because the survey period in the diary is so short, household fixed effects cannot be used to control for time-invariant household heterogeneity. Hence, while a diary survey could be a substitute for the app data in principle, the short sample of the CEX diary makes it a poor substitute in practice.[11]

The CEX Interview Survey provides a more complete measure of total spending, as well as a longer panel (4 quarters), from which we can make a comparison with estimates based on spending reported by the app at longer horizons. Panel A of Figure 4 reports the histogram (bin size is set to $1 intervals) of monthly spending on gasoline in the CEX Interview data for 2013-2014.[12] The distribution has clear spikes at multiples of $50 and $100 with the largest spikes at $0 and $200. In contrast, the distribution of gasoline purchases in the app data has a spike at $0 but the rest of the distribution exhibits considerably less bunching, particularly at large values like $200 or $400 that correspond with reporting $50 or $100 per week, respectively. In addition, the distribution of gasoline spending has a larger mass at smaller amounts in the app data than in the CEX Interview data. These differences are consistent with recall bias in the CEX Interview Survey data. As argued by Binder (2017), rounding in household surveys can reflect a natural uncertainty of households about how much they spent in this category.

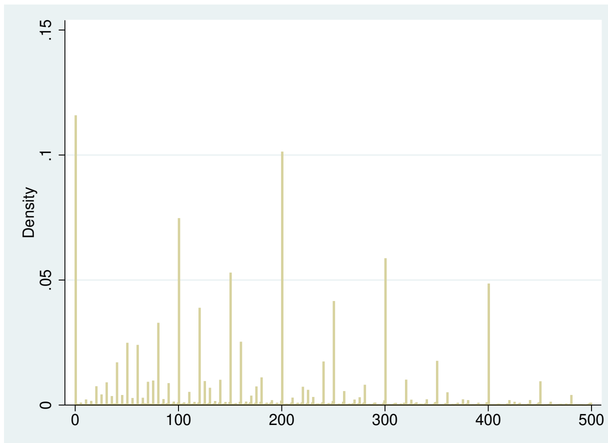
Table 3.2 compares moments for gasoline and non-gasoline spending across the CEX and the app data. We find that the means are similar across data sources. For example, mean (median) biweekly gasoline spending in the CEX Diary Survey is $84.72 ($65.00), while the app counterpart is $87.83 ($58.03). Similarly, mean (median) non-gasoline spending is $1,283 ($790.56) in the CEX Diary Survey and $1,561.38 ($1,084.38) in the app data. The standard deviation (interquartile range) tends to be a bit larger in the app data than in the CEX, which reflects a thicker right tail of spending in the app data. This pattern is consistent with top-coding and under-representation of higher-income households in the CEX, a well-documented phenomenon (Sabelhaus et al. 2015). The moments in the CEX Interview Survey (quarterly frequency) are even closer to the moments in the app data. For example, mean (median) spending on gasoline is 647(540) in the CEX Interview Survey data and $628 ($475) in the app data, while the standard deviations (interquartile ranges) are $531 ($630) and $588 ($660) respectively. In each panel of Table 2, we also

[11]We have done a comparison of the CEX diary spending for January 2013 through December 2014. In a regression of log daily spending for days with positive spending on month time effects and day of week dummies, the month effects estimated in the CEX and app have a correlation of 0.77. (Finer than monthly comparison of the app and CEX is not possible because the CEX provides only the month and day of week, but not the date, of the diary entry.

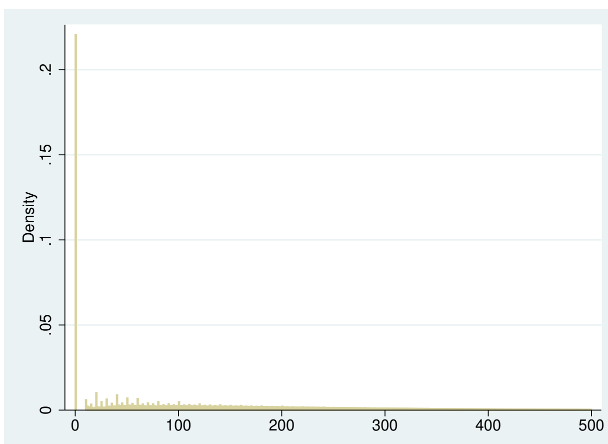
[12]The CEX Interview Survey question asks households to report their "Average monthly expense for gasoline."

Figure 3.4: Reported gasoline spending (monthly)

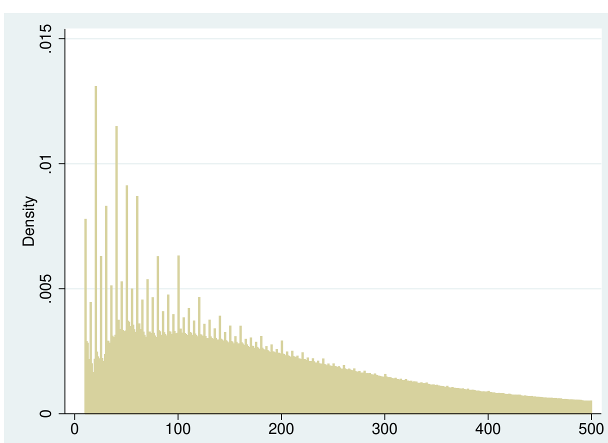
a. CEX Interview Survey



b. App data



c. App data excluding zero spending



Notes: the figure reports monthly spending on gasoline in the Interview segment of the Consumer Expenditure Survey (CEX) and in the app data. The horizontal axis is in dollars. The size of the bin in is set to $1 in all panels.

compare the distribution of the ratio of gasoline spending to non-gasoline spending, a central ingredient in our analysis. The moments for the ratio in the CEX and the app data are similar. For instance, the mean ratio is 0.08 for the CEX Interview Survey and 0.07 for the app data, while the standard deviation of the ratio is 0.07 for both the CEX Interview Survey and the app data.[13]

In summary, spending in the app data is similar to spending in the CEX data. Thus, although participation in the app is voluntary, app users have spending patterns similar to the population. In addition to reflecting survey recall bias and top-coding, some of the differences could reflect consumers buying gasoline on cards that are not linked to the app (such as credit cards specific to gasoline station chains), the ML procedure missing some gasoline stations, or gasoline spending done in cash that we could not identify. We will address these potential issues in our robustness tests.

3.4 Empirical Strategy

The discourse on potential macroeconomic effects of a fall in gasoline prices centers on the question of how savings from the fall in gasoline prices are used by consumers. Specifically, policymakers and academics are interested in the marginal propensity to consume (MPC) from savings generated by reduced gasoline prices.[14] Define MPC as

$$dC_{it} = -MPC * d(P_t Q_{it}) \tag{3.1}$$

where i and t index consumers and time, C is spending of non-gasoline items, P is the price of gasoline, and Q is the quantity of consumed gasoline. Note that we define the MPC as an increase in spending (measured in dollars) in response to a dollar *decrease* in spending on gasoline after the price of gasoline declines.[15]

Equation (3.1) is a definition, not a behavioral relationship. Of course, Q_{it}, the quantity of gasoline purchased, and overall non-gasoline spending, C_{it}, are simultaneously determined, with simultaneity being an issue at the individual as well as aggregate level. In this section, we develop an econometric relationship that yields

[13]Online Appendix Figure C1 shows the density of the gasoline to non-gasoline spending ratio for the CEX and app data.

[14]For example, Janet Yellen (Dec 2014) compared the fall in gasoline prices to a tax cut: "[The decline in oil prices] is something that is certainly good for families, for households, it's putting more money in their pockets, having to spend less on gas and energy, so in that sense it's like a tax cut that boosts their spending power."

[15]The MPC is likely different across groups of people, but our notation and estimation refers to the average MPC.

Table 3.2: Comparison of spending in the CEX and app data, 2013

Frequency and type of spending	Moment			
	Mean	St. Dev.	Median	Inter-quartile range
Panel A. Biweekly				
Spending on gas, $	84.72	83.42	65.00	101.44
CEX Diary Survey	87.83	98.79	58.03	130.87
App				
Spending on non-gasoline items, $	1,283.36	1,470.93	790.56	1,380.66
CEX Diary Survey	1,561.38	1,784.67	1,084.38	1,519.75
App				
Gasoline to non-gasoline spending	0.15	0.25	0.06	0.14
CEX Diary Survey	0.09	0.15	0.04	0.11
App				
Panel B. Quarterly				
Spending on gasoline, $				
CEX Interview Survey	646.63	530.87	540.00	630.00
App	627.94	588.24	475.33	660.18
Spending on non-gasoline items, $				
CEX Interview Survey	10,143.78	8,141.67	7,728.70	7,406.49
App	11,264.85	11,391.42	8,392.24	8,605.46
Gasoline to non-gasoline spending				
CEX Interview Survey	0.08	0.07	0.06	0.08
App	0.07	0.07	0.05	0.07

Notes: Means and standard deviation are from the distribution winsorized at the 1% level. The variables from the CEX use population sample weights. For Panel A, the ratio for a consumer/household is calculated as average value of the sum all gasoline spending during a biweekly period in 2013 divided by total non-gasoline spending in the corresponding biweekly period in 2013. For the app data, we mimic the design of the CEX Diary Survey by randomly drawing a two-week period for each user and discarding data for other weeks. For Panel B, the ratio for a consumer/household is calculated as the sum of all gasoline spending in a quarter, divided by total non-gasoline spending in that quarter.

identification of the MPC based on the specific sources of variation of gasoline prices discussed in the previous sections.

At the aggregate level, one important determinant of gasoline spending is aggregate economic conditions. As discussed in Section II, the 2007-2008 collapse in gasoline prices has been linked to the collapse in global demand due to the financial crisisdemand for gasoline fell driving down the price at the same time that demand was falling for other goods. Individual-level shocks are another important source of simultaneity bias and threat to identification. Consider a family going on a road trip to Disneyland; this family will have higher gasoline spending (long road trip) and higher total consumption in that week due to spending at the park. Yet another example is a person who suffers an unemployment spell; this worker will have lower gasoline spending (not driving to work) and lower other spending (a large negative income shock).

This discussion highlights that gasoline purchases and non-gasoline spending are affected by a variety of shocks. Explicitly modelling all possible shocks, some of which are expected in advance by households (unobservable to the econometrician), would be impossible. Fortunately, this is not required to properly identify the policy-relevant parameterthe sensitivity of non-gasoline spending to changes in gasoline spending induced by exogenous changes in the price of gasoline. This parameter may be interpreted as a partial derivative of non-gasoline spending with respect to the price of gasoline and thus could be mapped to a coefficient estimated in a regression. For this, we only need to satisfy a weaker set of conditions. First, we need exogenous, unanticipated shocks to gasoline prices. These shocks should be unrelated to the regression residual absorbing determinants of non-gasoline consumption unrelated to changes in gasoline prices. Second, we need to link non-gasoline spending to the price of gasoline (i.e., P_t), rather than purchases of gasoline (P_tQ_{it}).

As we established in Section II, shocks to gasoline prices in the period of our analysis were unanticipated, exogenous, and permanent so that we have an exogenous source of variation. To link the partial derivative of interest to a regression coefficient and to link it with cross-sectional variation in pre-determined propensity to spend on gasoline, we manipulate equation (1) as follows:

$$\begin{aligned}
\frac{dC_{it}}{\overline{C}_i} &= d\log C_{it} = -MPC \times \frac{d(P_tQ_{it})}{\overline{C}_i} = -MPC \times \frac{d(P_tQ_{it})}{(\overline{PQ})_i} \times \frac{(\overline{PQ})_i}{\overline{C}_i} \\
&= -MPC \times \frac{d(P_tQ_{it})}{(\overline{PQ})_i} \times s_i \\
&= -MPC \times \frac{\overline{Q}_i dP_t + \overline{P} dQ_{it}}{(\overline{PQ})_i} \times s_i \\
&- (MPC \times s_i \times \frac{dP_t}{\overline{P}} + MPC \times s_i \times \frac{dQ_{it}}{\overline{Q}_i}) \\
&= -MPC \times s_i \times d\log P_t - MPC \times s_i \times (\frac{dQ_{it}}{\overline{Q}_i}\frac{\overline{P}}{dP_t}) \times \frac{dP_t}{\overline{P}} \\
&= -MPC \times s_i \times d\log P_t - MPC \times s_i \times \epsilon \times d\log P_t \\
&= -MPC \times (1+\epsilon) \times s_i \times d\log P_t
\end{aligned} \tag{3.2}$$

where bars denote steady-state values, $s_i \equiv \frac{PQ_{it}}{\overline{C}_i}$ is the ratio of gasoline spending to non-gasoline spending,[16] and ϵ is the price elasticity of demand for gasoline (a negative number). Now the only source of time variation in the right-hand side of the equation is the price of gasoline. The identifying variation in equation (3.2) comes from time-series fluctuations in the price of gasoline interacted with the predetermined cross-sectional share of spending on gasoline.[17] The cross-section variation is essential for this paper since there is a single large episode of gasoline price movements in the sample period. One can also derive the specification from a utility maximization problem and link the MPC to structural parameters (see Online Appendix D). Thus, regressing log non-gasoline spending on the log of gasoline price multiplied by the ratio of gasoline spending to non-gasoline spending yields an estimate of $-\text{MPC}(1+\epsilon)$. Note that we have an estimate of -MPC scaled by $1+\epsilon$, but the scaling should be small if demand is inelastic. As discussed below, there is some variation in the literature on ϵ's estimated using household versus aggregate data. To ensure that a measure of ϵ is appropriate for our sample, we note:

[16]We calculate s_i as the ratio of consumer i's annual spending on gasoline to his/her annual spending on non-gasoline items in 2013. Using annual frequency in this instance helps to address seasonal variation in gasoline spending as well as considerable high frequency variation in the intensity of gasoline spending (e.g., trips to gasoline stations, spending per trip). Additionally, the use of 2013 data to calculate the share makes it pre-determined with respect to the shock to gasoline prices in the estimation period. In short, by using s_i for 2013, we approximate the response around the point where gasoline prices were high.

[17]Edelstein and Kilian (2009) consider a similar specification at the aggregate level.

$$
\begin{aligned}
d\log P_t Q_{it} &= d\log P_t + d\log Q_{it} \\
&= d\log P_t + d\log P_t \frac{d\log Q_{it}}{d\log P_t} = \big(1 + \frac{d\log Q_{it}}{d\log P_t}\big) \times d\log P_t \\
&= (1+\epsilon) \times d\log P_t
\end{aligned} \tag{3.3}
$$

Similar to equation (2), the only source of time variation in the right-hand side of equation (3) is the price of gasoline. Thus, a regression of $dlogP_tQ_{it}$ on $dlogP_t$ yields an estimate of elasticity $(1+\eta)$, which is the partial derivative of gasoline spending with respect to the price of gasoline, and the residual in this regression absorbs determinants of gasoline purchases unrelated to the changes in the price of gasoline.[18] The estimated $(1+\epsilon)$ and $-MPC(1+\epsilon)$ can be combined to obtain the MPC.

In the derivation of equations (3.2) and (3.3) we deliberately did not specify the time horizon over which sensitivities are computed, as these may vary with the horizon. For example, with lower prices, individuals may use their existing cars more intensively or may purchase less fuel-efficient cars. There may be delays in adjustment to changes in prices (e.g., search for a product). It might take time to notice the price change (Coibion and Gorodnichenko, 2015). The very-short-run effects may also depend on whether a driver's tank is full or empty when the shock hits. To obtain behavioral responses over different horizons, we build on the basic derivation above and estimate a multi-period long-differences model, where both the MPC and the price elasticity are allowed to vary with the horizon. Additionally, we introduce aggregate and idiosyncratic shocks to overall spending, and idiosyncratic shocks to gasoline spending. Hence,

$$\Delta_k \log C_{it} = \beta_k \times s_i \times \Delta_k \log P_t + \psi_t + \epsilon_{it} \tag{3.4}$$

$$\Delta_k \log P_t Q_{it} = \delta_k \Delta_k \log P_t + u_{it} \tag{3.5}$$

where $\beta = -MPC(1+\eta)$, $\delta = (1+\eta)$, $\Delta_k x_t = x_t - x_{(t-k)}$ is a k-period-difference operator, ψ_t is the time fixed effect, and ϵ_{it} and u_{it} are individual-level shocks to spending.[19] By varying k, we can recover the average impulse response over k-periods so that we can remain agnostic about how quickly consumers respond to a

[18] Because the dependent variable is spending on gasoline rather than volume of gasoline, elasticity ϵ estimated by this approach also includes substitution across types of gasoline (Hastings and Shapiro, 2013).

[19] Note that there are time effects only in equation (3.4). Since we have argued that changes in gasoline prices are exogenous over the time period, time effects are not needed for consistency of

change in gasoline prices.[20] Given that our specification is in differences, we control for consumer time-invariant characteristics (gender, education, location, etc.) as well as for the level effect of s_i on non-gasoline spending. To minimize adverse effects of extreme observations, we winsorize dependent variables $\Delta_k logC_{it}$ and $\Delta_k \log P_t Q_i t$ as well as s_i at the bottom and top one percent.

Because we are interested in the first-order effects of the fall in gasoline prices on consumer spending, we include the time fixed effects in specification (3.4). As a result, we obtain our estimate after controlling for common macroeconomic shocks and general equilibrium effects (e.g., changes in wages, labor supply, investment). Thus, consistent with the literature estimating MPC for income shocks (e.g. Shapiro and Slemrod, 2003; Johnson, Parker, and Souleles, 2006b; Parker, Souleles, Johnson, and McClelland, 2013; Jappelli and Pistaferri, 2010), we estimate a partial equilibrium MPC.

We assume a common price of gasoline across consumers in this derivation. In fact the comovement of gasoline prices is very strong (see Online Appendix Figure C2) and thus little is lost by using aggregate gasoline prices. Furthermore, when computing s_i we use gasoline spending rather than gasoline prices and thus our measure of s_i takes into account geographical differences in levels of gasoline prices. We find nearly identical results when we use local gasoline prices.

Note that gasoline and oil prices are approximately random walks and thus $\Delta_k logP_t$ can be treated as an unanticipated, permanent shock. To the extent oil prices and, hence, gasoline prices are largely determined by global factors or domestic supply shocks, rather than domestic demandwhich is our maintained assumption for our sample periodOLS yields consistent estimates of MPC and ϵ. Formally, we assume that the idiosyncratic shocks to spending are orthogonal to these movements in gasoline prices. Given the properties of the shock to gasoline prices in 2014-2015, the PIH model predicts that the response of spending from the resulting change in resources should be approximately equal to the change in resources ($MPC \approx 1$) and take place quickly.

The approach taken in specifications (4) and (5) has several additional advantages

estimation of either (3.4) or (3.5). In (3.4), they may improve efficiency by absorbing aggregate shocks to overall spending. We cannot include time effects in (3.5) because they would completely absorb the variation in gasoline prices. But again note that the presence of an aggregate component in u does not make the estimates of δ biased under our maintained assumption that gasoline prices are exogenous to the U.S. economy in the estimation period. (The standard errors account for residual aggregate shocks.)

[20]For example, if $\log C_{it} = \sum_(s=0)^{\infty} \psi_s shock_{(t-s)} + u_t$ and u_t summarizes variation orthogonal to the shock series of interest, then the impulse response is $\{\psi_s\}_{s=0}^{\infty}$ and the long-difference regression recovers $\beta_k = k^{-1} \sum_{s=0}^{k-1} \psi_s$.

econometrically. First, as discussed in Griliches and Hausman (1986), using "long differences" helps to enhance signal-to-noise ratio in panel data settings. Second, specifications (4) and (5) allow straightforward statistical inference. Because our shock $\Delta_k log P_t$ is effectively national and we expect serial within-user correlation in spending, we cluster standard errors on two dimensions: time and person. This approach to constructing standard errors is much more conservative than the common practice of clustering standard errors only by a consumer, employer, or location (e.g., Johnson et al. 2006, Levin et al. 2017). To make our results comparable to previous studies, we also report standard errors clustered on user only. Third, although the variables are expressed in logs, equation (2) shows that we estimate an MPC rather than an elasticity and thus there is no need for additional manipulation of the estimate. This aspect is important in practice because the distribution of spending is highly skewed (in our data, the coefficient of skewness for weekly spending is approximately four) and specifications estimating MPC on levels of spending (rather than logs) are likely sensitive to what happens in the right tail of the spending distribution. Finally, because oil and gasoline prices change every day and the decline in the price of oil (and gasoline) was spread over time, there is no regular placebo test on a "no change" period or before-after comparison. However, these limitations are naturally addressed using regression analysis.

To summarize, our econometric framework identifies the MPC from changes in gasoline prices by interacting two sources of variation: a large, exogenous, and permanent change in gasoline prices, with the pre-determined share of spending on gasoline. The econometric specification also accounts for the response of spending on gasoline to lower prices by allowing a non-zero elasticity of demand for gasoline and allowing for lagged adjustment of gasoline spending to changes in gasoline prices.

3.5 Results

In this section, we report estimates of MPC and ϵ for different horizons, frequencies, and populations. We also compare estimates based on our app data to the estimates based on spending data from the CEX.

Sensitivity of Expenditure to Gasoline Prices

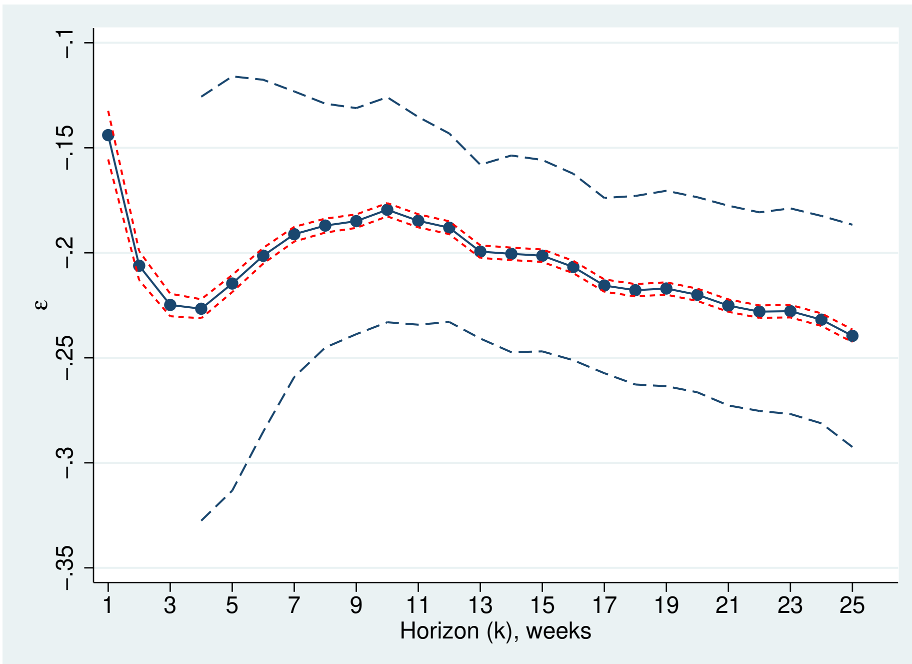
We start our analysis with the estimates of MPC and ϵ at weekly frequency for different response horizons. Panel A of Figure 5 shows $\hat{\epsilon}$ and 95 percent confidence bands, for k=0,...,26 weeks. Table 3, Row 1, gives the point estimates for selected horizons. The point estimates indicate that the elasticity of demand for gasoline is increasing in the horizon (i.e., over time, consumers have greater elasticity of demand): estimated elasticity changes from -0.20 at the horizon of 15 weeks to -0.24 at the horizon of 25 weeks. When we use our preferred standard errors clustered by time and user, confidence intervals are very wide at short horizons; estimates become quite precise at horizons of 12 weeks and longer. In contrast, the conventional practice of clustering standard errors by user yields tight confidence bands but these likely understate sampling uncertainty in our estimates because there is considerable cross-panel dependence in the data.

This estimate is broadly in line with previously reported estimates. Using aggregate data, the results in Hughes, Knittel, and Sperling (2008) suggest that U.S. gasoline demand is significantly more inelastic today compared with the 1970s. Regressing monthly data on aggregate per capita consumption of gasoline on changes in gasoline prices, they estimate a short-run (monthly) price elasticity of -0.034 to -0.077 for the 2001 to 2006 period, compared with -0.21 to -0.34 for the 1975-1980 period. The Environmental Energy Administration EIA (2014) also points to an elasticity close to zero, and also argues this elasticity has been trending downward over time.[21] In contrast to Hughes, Knittel, and Sperling (2008) our findings suggest that gasoline spending could still be quite responsive to gasoline price changes. In general, our results lie in between the Hughes, Knittel, and Sperling (2008) estimates and previous estimates using household expenditure data to measure gasoline price elasticities. Puller and Greening (1999) and Nicol (2003) both use the CEX interview survey waves from the 1980s to the early 1990s to estimate the elasticity of demand. The approaches taken across these papers are very different. Nicol (2003)'s approach is to estimate a structural demand system. Puller and Greening (1999), on the other hand, take advantage of the CEX modules about miles traveled that were only available in the 1980s, as well as vehicle information. Both of these papers find higher price elasticities of demand at the quarterly level, with estimates in Nicol (2003) ranging from -0.185 for a married couple with a mortgage and 1 child, to -0.85 for a renter with two children, suggesting substantial heterogeneity across households. Puller and Greening (1999)'s baseline estimates are -0.34 and -0.47, depending on the specification. A more recent paper by Levin et al. (2017) uses city level price data and city level expenditure data obtained from Visa credit card expenditures. They estimate the elasticity of demand for gasoline to be closer to ours, but still higher, ranging from 0.27 to 0.35. Their data are less aggregate than the other studies, but more aggregate than ours because we observe individual level data. Also, we observe

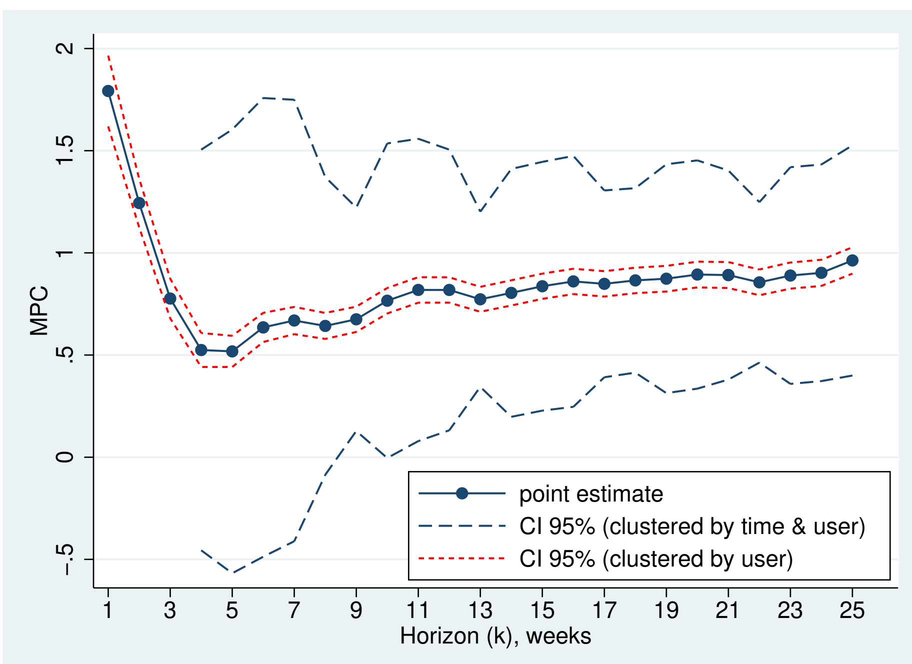
[21]EIA (2014) reports, "The price elasticity of motor gasoline is currently estimated to be in the range of -0.02 to -0.04 in the short term, meaning it takes a 25% to 50% decrease in the price of gasoline to raise automobile travel 1%. In the mid 1990's, the price elasticity for gasoline was higher, around -0.08."

Figure 3.5: Dynamic response to a change in gasoline price

Panel A. Elasticity of demand for gasoline, ϵ



Panel B. MPC



Notes: the figure reports estimates of elasticity of demand for gasoline ϵ (Panel A) and marginal propensity to consume (MPC) (Panel B) based on specifications (3.4) and (3.5). Standard errors clustered at both the week and user level. First three standard errors clustered by time and user are omitted for readibility of the graph. Dashed lines show 95 percent confidence interval. See text for further details.

Table 3.3: Estimated elasticity of demand and MPC: Baseline and estimates for single financial providers

Accounts	**Sample**		**Elasticity of demand for gasoline, ϵ**			**MPC**		
			Horizon (weeks)			Horizon (weeks)		
			5	15	25	5	15	25
		Row	(4)	(5)	(6)	(7)	(8)	(9)
Baseline	All	1	-0.215	-0.201	-0.240	0.518	0.837	0.963
			(0.050)	(0.023)	(0.027)	(0.554)	(0.311)	(0.287)
			[0.002]	[0.002]	[0.002]	[0.040]	[0.031]	[0.033]
Large Provider #1	Any Account	2	-0.256	-0.265	-0.314	0.125	0.586	0.515
			(0.048)	(0.024)	(0.029)	(0.835)	(0.482)	(0.387)
			[0.005]	[0.003]	[0.003]	[0.074]	[0.057]	[0.058]
	Core	3	-0.290	-0.261	-0.293	0.199	0.666	0.661
			(0.053)	(0.025)	(0.027)	(0.928)	(0.526)	(0.427)
			[0.005]	[0.003]	[0.004]	[0.066]	[0.050]	[0.058]
Large Provider #2	Any Account	4	-0.242	-0.260	-0.367	0.064	0.321	0.642
			(0.069)	(0.038)	(0.054)	(0.425)	(0.218)	(0.266)
			[0.005]	[0.003]	[0.004]	[0.066]	[0.050]	[0.058]
	Core	5	-0.223	-0.237	-0.326	-0.057	0.429	0.616
			(0.065)	(0.034)	(0.051)	(0.487)	(0.306)	(0.321)
			[0.008]	[0.006]	[0.006]	[0.120]	[0.092]	[0.102]

Large Provider #3	Any Account	6	-0.183	-0.179	-0.246	0.281	0.561	0.639
			(0.051)	(0.034)	(0.045)	(0.736)	(0.431)	(0.371)
			[0.005]	[0.004]	[0.004]	[0.086]	[0.068]	[0.073]
	Core	7	-0.161	-0.170	-0.241	0.204	0.575	0.734
			(0.052)	(0.035)	(0.046)	(0.595)	(0.366)	(0.349)
			[0.009]	[0.007]	[0.007]	[0.144]	[0.113]	[0.125]

Notes: The table reports estimates of elasticity of demand for gasoline ϵ and marginal propensity to consume (MPC) based on specifications (3.4) and (3.5) for horizons 5, 15, and 25 weeks. Row 1 presents the baseline estimates based on the full sample. In the rest of the table, the sample is restricted to a single provider indicated in the left column. In other words, we restrict the sample to accounts only at a specific provider so that we can mimic the data observed by a single provider. In rows (2), (4) and (6), the table report estimates for the case when we use any account of a provider. In rows (3), (5) and (7), the table report estimates based on core accounts; that is, to be part of the estimation sample, a user has to have at least one checking and one credit-card account with a given provider and have at least one transaction per month on each account. In all specifications, robust standard errors reported in parentheses are clustered at both the consumer and week level. Standard errors reported in squared brackets are clustered at the consumer level. See text for further details.

expenditures from all linked credit and debit cards and are not restricted only to Visa.

Panel B of Figure 3.5 shows the dynamics of $\hat{MPC}$ and 95 percent confidence bands over the same horizons with point estimates at selected horizons in the first row of Table 3.3. At short time horizons (contemporaneous and up to 3 weeks), the estimates vary considerably from nearly 2 to 0.5 but the estimates are very imprecise when we use standard errors clustered by time and user. Starting with the four-week horizon, we observe that $\hat{MPC}$ steadily rises over time and becomes increasingly precise. After approximately 12 weeks, $\hat{MPC}$ stabilizes between 0.8 and 1.0 with a standard error of 0.3. The estimates suggest that, over longer horizons, consumers spend nearly all their gasoline savings on non-gasoline items. The standard errors are somewhat smaller at monthly horizons (4-5 weeks) since the shock. While this pattern is not surprising given that β and δ in equations (4) and (5) at long horizons are effectively averages over many periods, we suspect this is also because the residual variance in consumption tends to be lower at monthly frequency due to factors like frequency of shopping, recurring spending, and bills paid, while in other weeks, the consumption process has considerably more randomness (see Coibion, Gorodnichenko, and Koustas, 2017). Similar to the case of ϵ, confidence bands are much tighter when we use standard errors clustered only by user.

There are not many estimates of the MPC derived from changes in gasoline prices. Farrell and Greig (2015) report examines the same time period that we do using similar data. It finds an MPC of 0.6, lower than our estimate. This finding likely arises from the use of data from a single financial institution rather than our more comprehensive data. This is an important advantage of the app data because many consumers have multiple accounts across financial institutions. The app's users have accounts on average in 2.6 different account providers (the median is 2). As a result, we have a more complete record of consumer spending. To illustrate the importance of this point, we rerun our specification focusing on a subgroup of consumers with accounts at the top three largest providers.[22] Specifically, we restrict the sample to accounts only at a specific provider so that we can mimic the data observed by a single provider. In rows (2), (4) and (6) of Table 3 we report estimates of ϵ and the MPC at horizons 5, 15 and 25 weeks for the case when we use any account at the provider. The MPC estimates based on data observed by a single provider are lower and have larger standard errors than the baseline, full-data MPC estimates reported in row (1). For example, the $\hat{MPC}$ for Provider 1 (row 2) at the 25-week horizon is 0.515, which is approximately half of the baseline $\hat{MPC}$ at 0.963. The standard error clustered by time and user for the former estimate is 0.387, so that we cannot

[22]These providers cover 49.6 percent of accounts in the data and 55.0 percent of total spending.

reject equality of the estimates as well as equality of the former estimate to zero. However, with the conventional practice of clustering standard errors only by user, one can reject equality of the estimates.

One may be concerned that having only one account with a provider may signal incomplete information because the user did not link all accounts with the app. To address this concern, we restrict the sample further to consider users that have at least one checking and one credit-card account with a given provider. In this case, one may hope that the provider is servicing "core" activities of the user. In rows (3), (5) and (7), we re-estimate our baseline specification with this restriction. We find estimates largely similar to the case of any account, that is, the estimated sensitivity to changes in gasoline prices is attenuated and more imprecise relative to the baseline where we have accounts linked across multiple providers.

These results for the single-provider data are consistent with the view that consumers can specialize their card use. For example, one card (account) may be used for gasoline purchases while another card (account) may be used for other purchases. In these cases, because single-provider information systematically misses spending on other accounts, MPCs estimated on single-provider data could be attenuated severely. We conjecture that using loyalty cards of a single gasoline retailer may also lead to understated estimates of MPC since loyalty cards are used only by 18 percent of consumers (National Association of Convenience Stores, 2015).

Robustness

While our specification has important advantages, there are nevertheless several potential concerns. First, if s_i in specification (3.4) is systematically underestimated because a part of gasoline spending is missing from our data, for instance, due to gasoline retailer cards that are not linked to the app, then our estimate of the MPC will be mechanically higher. Second, suppose instead that we are misclassifying some spending, or that consumers buy a large portion of their gasoline in cash, so that this spending shows up in our dependent variable. Misclassifying gasoline spending as non-gasoline spending will generate a positive correlation between non-gasoline spending and the gasoline price. Third, while a random walk may be a good approximation for the dynamics of gasoline prices, one may be concerned that gasoline prices have a predictable component, so that estimated reaction mixes up responses to unanticipated and predictable elements of gasoline prices. Indeed, some changes in gasoline prices are anticipated due to seasonal factors.[23]

[23]In the summer, many states require a summer blend of gasoline which is more expensive than a winter blend.

A practical implication of the first concern (i.e., cases where consumers use gasoline retailer cards that are not linked to the app) is that consumers with poorly linked accounts should have zero spending on gasoline. To evaluate if these cases could be quantitatively important for our estimates of MPC and ϵ, we estimate specifications (3.4) and (3.5) on the sample that excludes households with zero gasoline spending in 2013 (recall that the app data have a larger spike at zero than the counterpart in the CEX Interview Survey). Row (2) of Table 3.4 reports MPC estimates for this restricted sample at horizons k=5,15,25. We find that these estimates are very close to the baseline reported in row (1).

To address the second concern about cash spending, we note that, according to NACS (2015), less than a quarter of consumers typically pay for gasoline in cash and approximately 80 percent of consumers use credit and debit cards for purchases of gasoline. Furthermore, cash spending only shows up in the dependent variable, generating a positive correlation that will cause us to underestimate the MPC. In a robustness check, we exclude ATM and other cash withdrawals from the dependent variable. We find (row 3) that both the MPC and elasticity of demand estimated on these modified data are nearly identical to the baseline estimates. This finding is consistent with the intensity of using cash as means of payment being similar for gasoline and non-gasoline spending.

For the third concern relating to expected changes in gasoline prices, we turn to data from the futures market. In particular, we use changes in one-month-ahead futures for spot prices at New York Harbor (relative to last week's prediction for the month ahead) instead of the change in gasoline prices since last week. Specifically, let F_t^h denote the futures price at time t for month $t+h$. Then, in lieu of $\Delta_k \log P_t$ in our baseline specification (4), we instead use $\Delta_k \log F_t \equiv \log F_t^1 - \log F_(t-k)^1$ for $k \in 1,,25$. While the focus on one-month change is arguably justified given approximate random walk in gasoline prices, we also try the average change in the yield curves for gasoline prices over longer horizons (two years) to have a measure of changes in gasoline prices that are perceived as persistent: $\widetilde{\Delta_k log F_t} \equiv 1/24 \sum_(h = 1)^2 4(\log F_t^h - \log F_{(t-k)}^h)$. In either one-month change (row 4 of Table 4) or average change over two years (row 5), the results are very similar to our baseline.

Comparison with MPC using CEX

To appreciate the significance of using high-quality transaction-level data for estimating the sensitivity of consumers to income and price shocks, we estimated the sensitivity using conventional, survey-based data sources such as the Consumer Expenditure Survey (CEX). This survey provides comprehensive estimates of household

Table 3.4: Robustness of MPC estimate

Sample	Row	Elasticity of demand for gasoline, ϵ			MPC		
		Horizon (weeks)			Horizon (weeks)		
		5	15	25	5	15	25
		(1)	(2)	(3)	(4)	(5)	(6)
Baseline	1	-0.215	-0.201	-0.240	0.518	0.837	0.963
		(0.050)	(0.023)	(0.027)	(0.554)	(0.311)	(0.287)
		[0.002]	[0.002]	[0.002]	[0.040]	[0.031]	[0.033]
Exclude zero gasoline spending in 2013	2	-0.214	-0.201	-0.239	0.541	0.865	1.021
		(0.050)	(0.023)	(0.027)	(0.563)	(0.312)	(0.293)
		[0.002]	[0.002]	[0.002]	[0.039]	[0.032]	[0.033]
Exclude ATM withdrawals	3	-0.213	-0.198	-0.235	0.401	0.870	1.062
		(0.050)	(0.023)	(0.027)	(0.754)	(0.443)	(0.391)
		(0.023)	[0.001]	[0.002]	[0.034]	[0.027]	[0.028]
Change in one-month-ahead gasoline futures	4	-	-	-	0.204	0.855	0.907
					(0.440)	(0.234)	(0.216)
					[0.030]	[0.022]	[0.025]
Average change in the yield curve of gasoline futures	5	-	-	-	0.388	1.033	1.110
					(0.531)	(0.281)	(0.259)
					[0.038]	[0.025]	[0.030]

Notes: the table reports estimates of elasticity of demand for gasoline ϵ and marginal propensity to consume (MPC) based on specifications (3.4) and (3.5) for horizons 5, 15, and 25 weeks. Row 1 presents the baseline estimates based on the full sample. The estimation sample in row 2 excludes consumers with zero spending on gasoline in 2013. In row 3, we exclude ATM withdrawals and other cash withdrawals in calculation of the growth rate of non-gasoline spending. In rows 4 and 5, we replace actual changes in gasoline prices with changes in futures prices of gasoline in specification (3.4); specification (3.5) is estimated as in the baseline, so ϵ is the same as in row 1. Specifically, let F_t^h denote the futures price made at time t for period $t+h$. Then, in lieu of $\Delta_k \log P_t$ in our baseline specifications (3.4), we instead use $\Delta_k \log F_t \equiv \log F_t^1 - \log F_{(t-k)}^1$ for $k \in \{1, , 25\}$ in row 4 and the average change in the yield curves for gasoline prices over longer horizons (two years) $\Delta_k \widetilde{\log F_t} \equiv 1/24 \sum_{h=1}^{24} \left(\log F_t^h - \log F_{t-k}^h \right)$ in row 5. In all specifications, robust standard errors in parentheses are clustered at both the consumer and week level. Standard errors reported in squared brackets are clustered at the consumer level. See text for further details.

consumption across all goods in the household's consumption basket and is the most commonly used household consumption survey. In this exercise, we focus on the interview component of the survey which allows us to mimic the econometric analysis of the app data.

In this survey, households are interviewed for 5 consecutive quarters and asked about their spending over the previous quarter. Note that the quarters are not calendar quarters; instead, households enter the survey in different months and are asked about their spending over the previous three months. The BLS only makes available the data from the last 4 interviews; therefore, we have a one-year panel of consumption data for a household. Given the panel design of the CEX Interview Survey, we can replicate aspects of our research design described above. Specifically, we calculate the ratio of gasoline spending to non-gasoline spending in the first interview. We then estimate the MPC in a similar regression over the next three quarters for households in the panel. For this specification, we use BLS urban gasoline prices which provide a consistent series over this time period (see note for Table 3.1).

In the first row of Table 3.5, we estimate our baseline specification for the app data at the quarterly frequency: the estimates are slightly different from the estimates based on the weekly frequency, though much less precise. The standard errors clustered by user and time are so large that we cannot reject the null of equality of the estimates over time or across frequencies.

Note that in estimates from the app in row 1 we continue to use complete histories of consumer spending over 2013-2016 while the CEX tracks households only for four quarters. To assess the importance of having a long spending series at the consumer level, we modify the app data to bring it even closer to the CEX data. Specifically, for every month of our sample, we randomly draw a cohort of app users and track this cohort for only four consecutive quarters, thus mimicking the data structure of the CEX. Then, for a given cohort, we use the first quarter of the data to calculate s_i and use the remainder of the data to estimate ϵ and MPC. Results are reported in row 2 of Table 5. Generally, patterns observed in row 1 are amplified in row 2. In particular, the elasticity of demand for gasoline is even lower at shorter horizons and even greater at the longer horizons. In a similar spirit, the estimated MPC increases more strongly in the horizon when we track consumers for only four quarters relative to the complete 2013-2016 coverage. Also note that by tracking users only for four quarters, the difference between standard errors clustered by time and user and standard errors clustered by user is much smaller.

Panel B of Table 5 presents estimates based on the CEX. To maximize the precision of CEX estimates, we apply our approach to the CEX data covering 1980-2015. The point estimates (row 3) indicate that non-gasoline spending declines in response to decreased gasoline prices. Standard errors are so large that we cannot reject the

Table 3.5: Elasticity of demand for gasoline and MPC: Consumer Expenditure Survey (CEX) versus App

Sample	Row	**Elasticity of demand for gasoline, ϵ** Horizon (weeks) 5	15	25	**MPC** Horizon (weeks) 5	15	25
		(1)	(2)	(3)	(4)	(5)	(6)
Panel A: App data (quarterly)							
Baseline	1	-0.084	-0.126	-0.383	1.000	1.050	2.320
		(0.118)	(0.110)	(0.157)	(0.478)	(0.627)	(0.797)
		[0.003]	[0.003]	[0.004]	[0.030]	[0.030]	[0.050]
CEX Sample design	2	0.005	-0.111	-0.432	0.732	1.574	4.942
		(0.073)	(0.066)	(0.057)	(0.548)	(0.287)	(1.017)
		[0.006]	[0.007]	[0.011]	[0.093]	[0.162]	[0.445]
Panel B: CEX							
1980-2014	3	-0.429	-0.330	-0.352	-0.917	-0.526	-1.929
		(0.041)	(0.027)	(0.027)	(0.659)	(0.638)	(1.265)
		[0.010]	[0.009]	[0.012]	[0.227]	[0.178]	[0.266]
1985-1987	4	-0.478	-0.396	-0.449	9.624	5.074	4.116
		(0.168)	(0.145)	(0.090)	(5.959)	(4.057)	(4.031)
		[0.067]	[0.062]	[0.064]	[2.140]	[1.380]	[1.533]
1990-1992	5	-0.636	-0.562	-0.512	-6.420	-3.005	-4.917
		(0.136)	(0.142)	(0.106)	(4.775)	(4.516)	(6.450)
		[0.055]	[0.051]	[0.067]	[2.210]	[1.790]	[2.450]

2014-2015	6	-0.431	-0.408	-0.447	1.384	2.020	6.895
		(0.084)	(0.043)	(0.074)	(1.407)	(1.568)	(2.306)
		[0.046]	[0.042]	[0.063]	[1.008]	[0.908]	[1.487]

Notes: the table reports estimates of elasticity of demand for gasoline ϵ and marginal propensity to consume (MPC) based on specifications (4) and (5) for horizons 1, 2, and 3 quarters. The CEX estimates use the ratio of gasoline spending to non-gasoline spending calculated in the first interview, and exclude this period from estimation. For the baseline estimates in Row 1, we use the same 2013 ratio of gasoline spending to non-gasoline spending as in the baseline estimates, and aggregate the spending and gasoline prices to the quarterly level. In row 2, we replicate the CEX sampling scheme, randomly selecting a start month for a user and keeping only the data for the 12 month period that follows it (if a full 12 months of data follow). We similarly use the non-gasoline consumption calculated in the first quarter, and exclude this period from the estimation. In all specifications, robust standard errors in parentheses are clustered at both the consumer and week level. Standard errors reported in squared brackets are clustered at the consumer level. See text for further details.

null of no response. The estimated elasticity of demand for gasoline is approximately -0.4, which is a double of the estimates based on the app data and is similar to some of the previous CEX-based estimates (e.g. Nicol, 2003).

One should be concerned that the underlying variation of gasoline prices is potentially different across datasets. The dramatic decline in gasoline prices in 2014-2015 was largely determined by supply-side and foreign-demand factors, but it is less clear that one may be equally confident about the dominance of this source of variation over a longer sample period. Indeed, Barsky and Kilian (2004) and others argue that oil prices have often been demand-driven in the past. In this case, one may find a wrong-signed or a non-existent relationship between gasoline prices and non-gasoline spending. To address this identification challenge, we focus on instances when changes in oil prices were arguably determined by supply-side factors.

Specifically, we follow Hamilton (2009, 2011) and consider several episodes with large declines in oil prices: (i) the 1986 decline in oil prices (1985-1987 period); (ii) the 1990-1991 rise and fall in oil prices (1989-1992 period); (iii) the 2014-2015 decline on oil prices. Estimated MPCs and elasticities for each episode are reported in rows (4)-(6). The 1986 episode generates positive MPCs but the standard errors continue to be too high to reject the null of no response. The 2014-2015 episode generates similar, implausible large estimates of MPC, although the estimates are more precise. The 1990-1992 episode yields negative MPCs with large standard errors.

In summary, the CEX-based point estimates are volatile and imprecise. The data are inherently noisy. Moreover, when limited to sample periods that have credibly exogenous variation in gasoline prices, the sample sizes are far too small to make precise, robust inferences. Furthermore, these estimates do not appear to be particularly robust. These results are consistent with a variety of limitations of the CEX data such as small sample size, recall bias, and under-representation of high-income households. These results also illustrate advantages of using high-frequency (weekly) data relative to low-frequency (quarterly) data for estimating sensitivity of consumer spending to gasoline price shocks. The app's comprehensive, high frequency data, combined with a natural experimentthe collapse of oil and gasoline prices in 2014help us resolve these issues and obtain precise, stable estimates of MPC and elasticity of demand for gasoline.

Heterogeneity in Responses

Macroeconomic theory predicts that the responses of consumers to changes in income (or prices) could be heterogeneous with important implications for macroeconomic dynamics and policy. For example, Kaplan and Violante (2014) present a theoretical framework where "hand-to-mouth" (HtM) consumers with liquidity constraints should exhibit a larger MPC to transitory, anticipated income shocks than non-HtM consumers for whom these constraints are not binding. Kaplan and Violante (2014) document empirical evidence consistent with these predictions and quantify the contribution of consumer heterogeneity in terms of liquidity holdings for the 2001 Bush tax rebate. In a similar spirit, Mian and Sufi (2014), McKay, Nakamura and Steinsson (2016), and many others document that consumers' liquidity and balance sheets can play a key role for aggregate outcomes.

The conventional focus in this literature is the consumption response to transitory, anticipated income shocks because the behavior of HtM and non-HtM consumers should be particularly different in this case. First, HtM consumers spend an income shock when it is realized rather than when it is announced, while non-HtM consumers respond to the announcement and exhibit no change in spending at the time the shock is realized. Second, the MPC of non-HtM consumers should be small (this group smooths consumption by saving a big fraction of the income shock), while the MPC of HtM consumers should be large (the income shock relaxes a spending constraint for these consumers).

This sharp difference in the responses hinges on the temporary, anticipated nature of the shock. For other shocks, the responses may be alike across HtM and non-HtM consumers. For example, when the shock is permanent and unanticipated, HtM and non-HtM consumers should behave in the same way (Mankiw and Shapiro, 1985): both groups should have MPC=1 at the time of the shock. Intuitively, non-HtM consumers have MPC=1 because their lifetime resources change permanently and, accordingly, these consumers adjust their consumption by the size of the shock when the shock happens. HtM consumers have MPC=1 because they are in a corner solution and would like to spend away every dollar they receive in additional income the moment they receive it. Thus, macroeconomic theory predicts that, in this case, the MPC should be similar across HtM and non-HtM consumers and that the MPC should be close to one. We focus this section on testing these two predictions.

For these tests one needs to identify HtM and non-HtM consumers. This seemingly straightforward exercise has proved to be a challenge in applied work due to a number of data limitations, which have made researchers use proxies for liquidity constraints. As a result, estimated MPCs should be interpreted with caution and important caveats. For example, Kaplan, Violante, and Weidner (2014) argue that identification of HtM consumers requires information on consumers' liquidity holdings just before they receive pay checks. Because the Survey of Consumer Finances (SCF), the dataset used in Kaplan, Violante, and Weidner (2014), reports average balances for a household as well as average monthly income, Kaplan and Violante are forced to make assumptions about payroll frequency (also not reported in the SCF) and behavior of account balances (e.g., constant flow of spending). Given heterogeneity in payment cycles (i.e., weekly, biweekly, monthly) and spending patterns across consumers, this procedure can mix HtM and non-HtM consumers and, thus, yield an attenuated estimate of MPC.

In contrast, the app data allow us to take Kaplan, Violante, and Weidner (2014)'s definition literally. We identify the exact day of a consumer's payroll income (if any), and examine bank account and credit card balances of the consumer the day before this payment arrives. If a consumer has several pay checks per month, we treat these as separate events. A consumer is classified as HtM in a given month if, for any pay check events in the previous month, the consumer has virtually no liquid assets (less than \$100 in the consumer's checking or savings accounts net of credit card debt), or the consumer is in debt (the sum of the consumers' liquid assets and available balance on credit cards is negative) and is within \$100 of the consumer's credit card limits. Denote the dummy variable identifying hand-to-mouth consumers at this frequency with D^*_{it}. We find that, in the app data, roughly 20% of consumers are HtM, which is similar to the estimate reported in Kaplan, Violante, and Weidner (2014) for a nationally representative sample of U.S. households in the Survey of Consumer Finances.

To allow for heterogeneity in the MPC by liquidity, we add interaction terms to the baseline specifications (3.4) and (3.5):

$$\begin{aligned} \Delta_k \log C_{it} &= \beta_1 \times s_i \times \Delta_k \log P_t + \beta_2 \times s_i^{gas} \times \Delta_k \log P_t \times D_{it} \\ &= +\mu_0 \times D_{it} + \mu_1 \times s_i \times D_{it} + \psi_t + \omega_t \times D_{it} + \epsilon_{it} \end{aligned} \tag{3.6}$$

$$\Delta_k \log PQ_{it} = \delta_1 \times \Delta_k \log P_t + \delta_2 \times \Delta_k \log P_t \times D_{it} + \xi \times D_{it} + u_{it} \tag{3.7}$$

where D_{it} is a variable measuring the presence/intensity of liquidity constraints identifying HtM consumers, and $\omega_t \times D_{it}$ is the time fixed effect specific to HtM consumers. We have several options for D_{it}. One could use a dummy variable equal to one if a consumer is liquidity constrained in period $t-k-1$ (recall that Δ_k operator calculates the growth rate between periods $t-k$ and t). We denote this "lagged" measure of HtM with $D_{it} \equiv D^*_{i,t-k-1}$ where D^*_{it} is a dummy variable equal to one if consumer i at time t satisfies the Kaplan-Violante HtM criteria and zero otherwise. Alternatively, because liquidity constraints may be short-lived, one may want to use measures that are calculated over a longer horizon to identify serial HtM consumers. To this end, we construct three measures on the 2013 sample which are not used in the estimation of MPC and ϵ. Specifically, for each month of data available for consumer i in 2013, we use three metrics to classify consumers as HtM or not. We consider the average value of D^*_{it} (this continuous variable provides a sense of frequency of liquidity constraints; we denote this measure with $\overline{D}_{i,2013}$, the modal value of $D^*_{i,t}$ (most frequent value; we denote this measure with $D_{i,2013}$, or the minimum value of $D^*_{i,2013}$ during the 2013 part of the sample. The latter measure, which we denote with $D_{i,2013}$, is equal to one only if a consumer is identified as HtM in every month in 2013.

Irrespective of which measure we use, we find in results reported in Table 6 that estimated MPCs are very similar for HtM and non-HtM consumers. Although the point estimates for HtM consumers tend to be larger at short horizons (e.g., 5 weeks), we generally cannot reject the null of equal MPCs across the groups or the null that estimated MPCs are equal to one, which is consistent with the PIH predictions.

3.6 Conclusion

How consumers respond to changes in gasoline prices is a central question for policymakers and researchers. We use big data from a personal financial management service to examine the dynamics of consumer spending during the 2014-2015 period when gasoline prices plummeted by 50 percent. Given the low elasticity of demand for gasoline, this major price reduction generated a large windfall for consumers equal to approximately 2 percent of total consumer spending.

We document that the marginal propensity to consume (MPC) out of these savings is approximately one. Since the change in gasoline prices was unexpected and permanent, this estimate can be interpreted as capturing MPC out of permanent income, an object that has been most difficult to estimate with previously available data. We argue that our results are consistent with the predictions of the permanent income hypothesis, including the prediction that both hand-to-mouth, and non-hand-to-mouth, consumers should respond similarly to an effectively permanent change.

While estimating the macroeconomic effects of the change in oil prices is beyond the scope of this paper, this partial equilibrium estimate provides a first-step input for quantifying the effects on the aggregate economy, which depend on a number of factors. The aggregate effects of changes in gasoline prices potentially depend on general equilibrium effects and redistribution of resources in the economy. The aggregate response to a gasoline price shock may be a function of the sensitivity of, for example, sectoral wages and employment to energy price shocks (see Online Appendix D for a model). Depending on specific assumptions about utility and production functions, general equilibrium effects can amplify or attenuate the immediate effects that we estimate. Moreover, there are income effects arising from the ownership of energy resources both domestically and abroad that will have macroeconomic effects. Nevertheless, any offsetting macroeconomic effects, e.g., from changes in oil

Table 3.6: MPC by liquidity status

	Elasticity of demand for gasoline, ϵ			**MPC**		
	Horizon (weeks)			Horizon (weeks)		
Measure of Hand-to-mouth consumers (HtM)	5	15	25	5	15	25
	(1)	(2)	(3)	(4)	(5)	(6)
A. Lagged HtM						
Non-HtM	-0.191	-0.155	-0.179	0.569	0.894	0.660
	(0.057)	(0.025)	(0.026)	(0.749)	(0.438)	(0.356)
	[0.004]	[0.003]	[0.003]	[0.087]	[0.060]	[0.058]
HtM	-0.239	-0.240	-0.293	0.547	1.044	0.896
	(0.058)	(0.030)	(0.031)	(0.829)	(0.472)	(0.471)
	[0.016]	[0.010]	[0.009]	[0.297]	[0.186]	[0.181]
	0.052	0.000	0.000	0.915	0.883	0.658
P-value	[0.003]	[0.000]	[0.000]	[0.857]	[0.779]	[0.408]
B. Average HtM						
Non-HtM	-0.186	-0.164	-0.199	0.618	0.776	0.783
	(0.054)	(0.024)	(0.027)	(0.619)	(0.311)	(0.299)
	[0.003]	[0.002]	[0.002]	[0.053]	[0.043]	[0.045]
HtM	-0.274	-0.317	-0.373	0.742	1.336	1.383
	(0.055)	(0.030)	(0.035)	(0.838)	(0.537)	(0.428)
	[0.007]	[0.005]	[0.005]	[0.098]	[0.088]	[0.098]
	0.007	0.000	0.000	0.932	0.073	0.077
P-value	[0.000]	[0.000]	[0.000]	[0.653]	[0.000]	[0.000]
C. Modal HtM in 2013						
Non-HtM	-0.191	-0.172	-0.209	0.597	0.899	0.893
	(0.053)	(0.024)	(0.027)	(0.651)	(0.358)	(0.320)
	[0.003]	[0.002]	[0.002]	[0.052]	[0.042]	[0.044]
HtM	-0.255	-0.287	-0.340	0.896	0.851	0.965
	(0.053)	(0.028)	(0.032)	(0.546)	(0.270)	(0.291)
	[0.006]	[0.004]	[0.004]	[0.106]	[0.092]	[0.099]
	0.012	0.000	0.000	0.579	0.410	0.641
P-value	[0.000]	[0.000]	[0.000]	[0.007]	[0.095]	[0.382]

D. Extreme HtM						
Non-HtM	-0.195	-0.182	-0.221	0.613	0.924	0.933
	(0.053)	(0.024)	(0.028)	(0.639)	(0.355)	(0.319)
	[0.003]	[0.002]	[0.002]	[0.049]	[0.040]	[0.042]
HtM	-0.284	-0.318	-0.365	1.029	0.912	1.064
	(0.056)	(0.029)	(0.033)	(0.571)	(0.297)	(0.310)
	[0.009]	[0.006]	[0.006]	[0.156]	[0.136]	[0.146]
	0.002	0.000	0.000	0.470	0.472	0.753
P-value	[0.000]	[0.000]	[0.000]	[0.001]	[0.230]	[0.626]

Notes: the table reports estimates of MPC and ϵ based on equations (6)-(7) over k periods, where k is shown in the top row of the table. s_i^{gas} is the ratio of gasoline spending to non-gasoline spending for 2013 for consumer i. The title of each panel indicates how the presence/intensity of liquidity constraints is measured. Denote the dummy variable identifying hand-to-mouth consumers for a given month with D_{it}^*. Panel A uses a dummy variable equal to one if a consumer is liquidity constrained in period t-k-1 (recall that Δ_k operator calculates the growth rate between periods t-k and t), i.e. $D_{it} \equiv D_{i,t-k-1}^*$. For other panels, we construct three measures on the 2013 sample which is not used in the estimation of MPC and ϵ: the average value of D_{it}^* (this continuous variable provides a sense of frequency of liquidity constraints; we denote this measure with $D_{,2013}$, the modal value of D_{it}^* (most frequent value; we denote this measure with $D_{i,2013}$, or the minimum value of $D_{i,2013}^*$ during the 2013 part of the sample. The latter measure, which we denote with $D_{i,2013}$ and refer to as "extreme," is equal to one only if a consumer is identified as hand-to-mouth in every month in 2013. Robust standard errors in parentheses are clustered by week and consumer. Standard errors reported in squared brackets are clustered at the consumer level. P-value is the p-value for the test of HtM and non-HtM responses being equal. See text for further details.

field production or from exports to foreign, oil-rich countries, do not obviate the interest in estimates of response of U.S. consumers to a very significant shock to their budget sets coming from gasoline prices.

We also show why previous attempts to estimate the MPC out of gasoline savings led to lower and/or more imprecise estimates due to data limitations (e.g., low frequency of data, incomplete coverage of consumer spending, short panel) in earlier studies. Our analysis highlights the substantial potential of administrative big data for enhancing national economic statistics, as well as estimates of key, policy-relevant macroeconomic parameters.

Bibliography

ABOWD, J. M., AND D. CARD (1987): "Intertemporal Labor Supply and Long-Term Employment Contract," *The American Economic Review*, 77(1), 50–68.

ABRAHAM, K. G., K. SANDUSKY, J. C. HALTIWANGER, AND J. R. SPLETZER (2017): "Measuring the Gig Economy: Current Knowledge and Open Issues," .

AGUIAR, M., AND E. HURST (2005): "Consumption versus Expenditure," *Journal of Political Economy*, 113(5), pp. 919–948.

——— (2007): "Life-Cycle Prices and Production," *American Economic Review*, 97(5), 1533–1559.

——— (2013): "Deconstructing Life Cycle Expenditure," *Journal of Political Economy*, 121(3), pp. 437–492.

AIYAGARI, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *The Quarterly Journal of Economics*, 109(3), 659–684.

ALAN, S., O. ATTANASIO, AND M. BROWNING (2009): "Estimating Euler equations with noisy data: Two exact GMM estimators," *Journal of Applied Econometrics*, 24(2), 309–324.

ALTONJI, J. (1993): "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," *Journal of Political Economy*, 3(2), S176–S215.

ALTONJI, J., AND C. H. PAXSON (1992): "Labor Supply, Hours Constraints, and Job Mobility," *Journal of Human Resources*, 27(2), 256–278.

ANGRIST, J. D., S. CALDWELL, AND J. V. HALL (2017): "Uber vs. Taxi: A Driver's Eye View," *NBER Working Paper No. 23891.*

ATTANASIO, O., E. BATTISTIN, AND H. ICHIMURA (2007): "What Really Happened to Consumption Inequality in the United States?," vol. Hard-to-Measure

Goods and Services: Essays in Honor of Zvi Griliches of *NBER Book Series Studies in Income and Wealth*, pp. 515–543. National Bureau of Economic Research, University of Chicago Press.

Attanasio, O., and L. Pistaferri (2014): "Consumption inequality over the last half century: Some evidence using the new PSID consumption measure," *American Economic Review*, 104(5), 122–126.

Baffes, J., M. A. Kose, F. Ohnsorge, and M. Stocker (2015): "The Great Plunge in Oil Prices: Causes, Consequences, and Policy Responses," *World Bank Group Policy Research Note*, 01.

Baker, S. R. (forthcoming): "Debt and the Response to Household Income Shocks: Validation and Application of Linked Financial Account Data," *Journal of Political Economy*.

Barsky, R. B., and L. Kilian (2004): "Oil and the Macroeconomy Since the 1970s," *Journal of Economic Perspectives*, 18(4), 115–134.

Beaudry, P., and J. DiNardo (1991): "The Effect of Implicit Contracts on the Movement of Wages Over the Business Cycle: Evidence from Micro Data," *Journal of Political Economy*, 99(4), 665–688.

Berger, T., C. Chen, and C. B. Frey (2017): "Drivers of Disruption? Estimating the Uber Effect," *Oxford Martin School Working Paper*.

Binder, C. (2017): "Measuring Uncertainty Based on Rounding: New Method and Application to Inflation Expectations," *Journal of Monetary Economics*, 90, 1–12.

Blundell, R., C. Meghir, and P. Neves (1993): "Labour supply and intertemporal substitution," *Joural of Econometrics*, 59, 137–160.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98(5), 1887–1921.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2016): "Consumption Inequality and Family Labor Supply," *American Economic Review*, 106(2), 387–435.

Board of Governors of the Federal Reserve System (2017): "Report on the Economic Well-Being of U.S. Households in 2016," .

BODIE, Z., R. C. MERTON, AND W. F. SAMUELSON (1992): "Labor supply flexibility and portfolio choice in a life cycle model," *Journal of Economic Dynamics and Control*, 16, 427–449.

BRAZIL, N., AND D. S. KIRK (2016): "Uber and Metropolitan Traffic Fatalities in the United States," *American Journal of Epidemiology*, 184(3), 192–198.

CARD, D. (1990): "Labor Supply with a Minimum Hours Threshold," *Carnegie-Rochester Conference Series on Public Policy*, 33, 137–168.

——— (1994): "Intertemporal Labor Supply: An Assessment," in *Advances in Econometrics, Sixth World Congress*, ed. by C. Sims, vol. II. Cambridge University Press.

CARD, D., R. CHETTY, AND A. WEBER (2007): "Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market," *Quarterly Journal of Economics*, 122(4), 1511–1560.

CARROLL, C. (2011): "Death to the Log-Linearized Consumption Euler Equation! (And Very Poor Health to the Second-Order Approximation)," *Advances in Macroeconomics*.

CARROLL, C. D. (1997): "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis," *Quarterly Journal of Economics*, CXII(1), 1–56.

CHEN, M. K., J. A. CHEVALIER, P. E. ROSSI, AND E. OEHLSEN (2017): "The Value of Flexible Work: Evidence from Uber Drivers," *NBER Working Paper No. 23296*.

CHEN, M. K., AND M. SHELDON (2011): "Dynamic Pricing in a Labor Market: Surve Pricing and Flexible Work on the Uber Platform," .

CHETTY, R., J. FRIEDMAN, S. LETH-PETERSEN, T. H. NIELSEN, AND T. OLSEN (2014): "Active vs. Passive Decisions and Crowd-out in Retirement Savings Accounts: Evidence from Denmark," *Quarterly Journal of Economics*, 129(3), 1141–1219.

CHETTY, R., AND A. SZEIDL (2007): "Consumption Commitments and Risk Preferences," *Quarterly Journal of Economics*, 122(2), 831–877.

COGAN, J. F. (1981): "Fixed Costs and Labor Supply," *Econometrica*, 49(4), 945–963.

Cogley, T., and J. M. Nason (1995): "Output Dynamics in Real-Business-Cycle Models," *The American Economic Review*, 85(3), 492–511.

Cohen, P., R. Hahn, J. Hall, S. Levitt, and R. Metcalfe (2016): "Using Big Data to Estimate Consumer Surplus: The Case of Uber," *NBER Working Paper No. 22627*.

Coibion, O., and Y. Gorodnichenko (2015): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, 195, 2644–2678.

Coibion, O., Y. Gorodnichenko, and G. H. Hong (2015): "The Cyclicality of Sales, Regular and Effective Prices: Business Cycle and Policy Implications," *American Economic Review*, 105(3), 993–1029.

Coibion, O., Y. Gorodnichenko, and D. Koustas (2017): "Consumption Inequality and the Frequency of Purchases," NBER Working Paper No. 23357.

Committee on National Statistics (2013): *Measuring What We Spend: Toward a New Consumer Expenditure Survey*. National Academies Press, Washington, D.C.

Crawley, E. (2018): *Time Aggregation in Panel Data on Income and Consumption.*

Deaton, A. (1991): "Saving and Liquidity Constraints," *Econometrica*, 59(5), 1221–1248.

Dills, A. K., and S. E. Mulholland (2017): "Ride-Sharing, Fatal Crashes, and Crime," .

Domeij, D., and M. Floden (2006): "The labor-supply elasticity and borrowing constraints: Why estimates are biased," *Review of Economic Dynamics*, 9(2), 242 – 262.

Edelstein, P., and L. Kilian (2009): "How sensitive are consumer expenditures to retail energy prices?," *Journal of Monetary Economics*, 56(6), 766–779.

Farrell, D., and F. Greig (2015): "How Falling Gas Prices Fuel the Consumer: Evidence from 25 Million People," *J.P. Morgan Chase Institute.*

——— (2016a): "The Online Platform Economy: Has Growth Peaked?," *JP Morgan Chase & Co. Institute.*

——— (2016b): "Paychecks, Paydays, and the Online Platform Economy: Big Data on Income Volatility," *JP Morgan Chase & Co. Institute*.

FRIEDMAN, M. (1957): *A Theory of the Consumption Function*. Princeton University Press.

GANONG, P., AND P. NOEL (2017): "Consumer Spending During Unemployment: Positive and Normative Implications," .

GELMAN, M. (2016): "What Drives Heterogeneity in the Marginal Propensity to Consume? Temporary Shocks vs Persistent Characteristics," .

GELMAN, M., Y. GORODNICHENKO, S. KARIV, D. KOUSTAS, M. D. SHAPIRO, D. SILVERMAN, AND S. TADELIS (2016): "The Response of Consumer Spending to Changes in Gasoline Prices," .

GELMAN, M., S. KARIV, M. D. SHAPIRO, D. SILVERMAN, AND S. TADELIS (2014): "Harnessing naturally occurring data to measure the response of spending to income," *Science*, 345(6193), 212–215.

——— (2015): "How Individuals Smooth Spending: Evidence from the 2013 Government Shutdown Using Account Data," *NBER Working Paper No. 21025*.

GERTLER, P., AND J. GRUBER (2002): "Insuring Consumption Against Illness," *American Economic Review*, 92(1), 51–70.

GICHEVA, D., J. HASTINGS, AND S. VILLAS-BOAS (2010): "Investigating Income Effects in Scanner Data: Do Gasoline Prices Affect Grocery Purchases?," *American Economic Review: Papers and Proceedings*, 100(480-484).

GOURINCHAS, P.-O., AND J. A. PARKER (2002): "Consumption over the Life Cycle," *Econometrica*, 70(1), pp. 47–89.

GRILICHES, Z., AND J. A. HAUSMAN (1986): "Errors in variables in panel data," *Journal of Econometrics*, 31(1), 93–118.

GROSS, D. B., AND N. S. SOULELES (2002): "Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data," *The Quarterly Journal of Economics*, 117(1), 149–185.

GRUBER, J. (1997): "The Consumption Smoothing Benefits of Unemployment Insurance," *The American Economic Review*, 87(1), pp. 192–205.

GUVENEN, F., F. KARAHAN, S. OZKAN, AND J. SONG (2016): "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Dynamics?," .

HAGEDORN, M., AND I. MANOVSKII (2008): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited," *American Economic Review*, 98(4), 1692–1706.

HALL, J., J. HORTON, AND D. KNOEPFLE (2017): "Labor Market Equilibration: Evidence from Uber," .

HALL, J. D., C. PALSSON, AND J. PRICE (2016): "Is Uber a substitute or complement for public transit?," .

HALL, J. V., AND A. B. KRUEGER (????): "An Analysis of the Labor Market for Uber's Driver-Partners in the United States," .

——— (2016): "An Analysis of the Labor Market for Uber's Driver-Partners in the United States," *NBER Working Paper 22843*.

HAM, J. C. (1986): "Testing Whether Unemployment Represents Intertemporal Labour Supply Behaviour," *The Review of Economic Studies*, 53(4), pp. 559–578.

HAMILTON, J. (2008): *New Palgrave Dictionary of Economics*chap. Oil and the Macroeconomy. Palgrave McMillan Ltd, 2nd edn.

HAMILTON, J. D. (2009): "Causes and Consequences of the Oil Shock of 2007-08," *Brookings Papers on Economic Activity*, pp. 215–283.

——— (2011): *Routledge Handbook of Major Events in Economic History*chap. Historical Oil Shocks, pp. 239–265. Routledge Taylor and Francis Group, New York.

HARRIS, S. D., AND A. B. KRUEGER (2015): "A Proposal for Modernizing Labor Laws for Twenty-First-Century Work: The "Independent Worker"," *Hamilton Project Discusion Paper 2015-10*.

HASTINGS, J. S., AND J. M. SHAPIRO (2013): "Fungibility and Consumer Choice: Evidence from Commodity Price Shocks," *The Quarterly Journal of Economics*, 128(4), 1449–1498.

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2014): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*, 107(7), 2075–2126.

Hughes, J. E., C. R. Knittel, and D. Sperling (2008): "Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand," *Energy Journal*, 29(1), 93–114.

Jackson, E., A. Looney, and S. Ramnath (2017): "The Rise of Alternative Work Arrangements: Evidence and Implications for Tax Filing and Benefit Coverage," *Office of Tax Analysis Working Paper 114*.

Jappelli, T., and L. Pistaferri (2006): "Intertemporal Choice and Consumption Mobility," *Journal of the European Economic Association*, 4(1), 75–115.

——— (2010): "The Consumption Response to Income Changes," *Annual Review of Economics*, 2(1), 479–506.

Johnson, D. S., J. A. Parker, and N. S. Souleles (2006a): "Household Expenditure and the Income Tax Rebates of 2001," *American Economic Review*, 96(5), 1589–1610.

——— (2006b): "Household Expenditure and the Income Tax Rebates of 2001," *American Economic Review*, 96(5), 1589–1610.

Jordà, Ò. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95(1), 161–182.

Juster, F., J. Smith, and F. Stafford (1999): "The measurement and structure of household wealth," *Labour Economics*, 6(2), 253–275.

Kaplan, G., G. Violante, and J. Weidner (2014): "The wealthy hand-to-mouth," *Brookings Papers on Economic Activity*, Spring 2014, 77–138.

Kaplan, G., and G. L. Violante (2014): "A Model of the Consumption Response to Fiscal Stimulus Payments," *Econometrica*, 82(4), 1199–1239.

Katz, L. F., and A. B. Krueger (2016): "The Rise and Nature of Alternative Work Arrangements in the United States 1995-2015," .

Kilian, L. (2008): "The Economic Effects of Energy Price Shocks," *Journal of Economic Literature*, 46(4), 871–909.

Koustas, D. (2018): "Consumption Insurance and Multiple Jobs: Evidence from Rideshare Drivers," .

Krishnan, P. (1990): "The Economics of Moonlighting: A Double Self-Selection Model," *The Review of Economics and Statistics*, 72(2), 361–367.

KUENG, L. (2015): "Revisiting the Response of Household Spending to the Alaska Permanent Fund Dividend Using CE Data," .

——— (2016): "Explaining Consumption Excess Sensitivity with Near-Rationality: Explaining Consumption Excess Sensitivity with Near-Rationality: Evidence from Large Predetermined Payments," .

LAIBSON, D., P. MAXTED, A. REPETTO, AND J. TOBACMAN (2017): "Estimation Discount Functions with Consumption Choices over the Lifecycle," .

LAMADON, T. (2016): "Productivity Shocks, Long-Term Contracts and Earnings Dynamics," .

LOW, H. W. (2005): "Self-insurance in a life-cycle model of labour supply and savings," *Review of Economic Dynamics*, 8, 945–975.

LUCAS, R. E. (2003): "Macroeconomic Priorities," *American Economic Review*, 93(1), 1–14.

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

MANKART, J., AND R. OIKONOMOU (2017): "Household Search and the Aggregate Labour Market," *Review of Economic Studies*, 84(1), 1735–1788.

MANKIW, N. G., AND M. D. SHAPIRO (1985): "Trends, Random Walks, and Tests of the Permanent Income Hypothesis," *Journal of Monetary Economics*, 16, 165–174.

MAS, A., AND A. PALLAIS (forthcoming): "Valuing Alternative Work Arrangements," *American Economic Review*, NBER Working Paper 22708.

MINCER, J. (1962): *Aspects of Labor Economics*chap. Labor Force Participation of Married Women: A Study of Labor Supply, pp. 63–105. Princeton University Press.

MORDUCH, J., AND R. SCHNEIDER (2017): *The Financial Diaries: How American Families Cope in a World of Uncertainty*. Princeton University Press.

NATIONAL ASSOCIATION OF CONVENIENCE STORES (2015): "Retail Fuels Report," Discussion paper.

NERI, A., C. RONDINELLI, AND F. SCOCCIANTI (2017): "Household spending out of a tax rebate: Italian "€80 tax bonus"," *Working Paper Series 2099, European Central Bank.*

NICOL, C. (2003): "Elasticities of demand for gasoline in Canada and the United States," *Energy Economics*, 25, 201–214.

OLAFSSON, A., AND M. PAGEL (2016): "The Liquid Hand-to-Mouth: Evidence from Personal Finance Management Software," .

PARKER, J. A., N. S. SOULELES, D. S. JOHNSON, AND R. MCCLELLAND (2013): "Consumer Spending and the Economic Stimulus Payments of 2008," *American Economic Review*, 103(6), 2530–2553.

PAXSON, C. H., AND N. SICHERMAN (1996): "The Dynamics of Dual Job Holding and Job Mobility," *Journal of Labor Economics*, 14(3), 357–393.

PISCHKE, J.-S. (1995): "Measurement error and earnings dynamics: Some estimates from the psid validation study," *Journal of Business and Economic Statistics*, 13(3), 305–314.

PISTAFERRI, L. (2003): "Anticipated and Unanticipated Wage Changes, Wage Risk, and Intertemporal Labor Supply," *Journal of Labor Economics*, 21(3), 729–754.

——— (2015): "Household Consumption: Research Questions, Measurement Issues, and Data Collection Strategies," *Journal of Economic and Social Measurement.*

PULLER, S., AND L. A. GREENING (1999): "Household adjustment to gasoline price change: an analysis using 9 years of US survey data," *Energy Economics*, 21, 37–52.

REICH, M., AND J. PARROTT (2018): "An Earnings Standard for New York City App Based Drivers," *Center for New York City Affairs at the New School and The Center on Wage and Employment Dynamics at the University of California, Berkeley.*

RENNA, F., AND R. OAXACA (2006): "The Economics of Dual Job Holding: A Job Portfolio Model of Labor Supply," *IZA DIscussion Paper No. 1915.*

SAPORTA-EKSTEN, I. (2014): "Job Loss, Consumption and Unemployment Insurance," Ph.D. thesis, Stanford University.

Shapiro, M. D., and J. Slemrod (2003): "Consumer Response to Tax Rebates," *American Economic Review*, 93(1), 381–396.

Shishko, R., and B. Rostker (1976): "The Economics of Multiple Job Holding," *The American Economic Review*, 66(3), 298–308.

Skinner, J. (1987): "A superior measure of consumption from the panel study of income dynamics," *Economics Letters*, 23(2), 213–216.

Stephens, M. J. (2001): "The Long-Run Consumption Effects of Earnings Shocks," *The Review of Economics and Statistics*, 83(1), 28–36.

Swanson, E. T. (2012): "Risk Aversion and the Labor Margin in Dynamic Equilibrium Models," *American Economic Review*, 102(4), 1663–1691.

——— (2014): "Implications of Labor Market Frictions for Risk Aversion and Risk Premia," .

——— (2015): "Risk Aversion, Risk Premia, and the Labor Margin with Generalized Recursive Preferences," .

Working, H. (1960): "Note on the Correlation of First Differences of Averages in a Random Chain," *Econometrica*, 28(4), 916–918.

Zeldes, S. P. (1989): "Consumption and Liquidity Constraints: An Empirical Investigation," *Journal of Political Economy*, 97(2), pp. 305–346.

Zhao, N. L. (2015): "Search and Multiple Jobholding," *Job Market Paper*.

Appendix A

Data Appendix

This appendix describes how gig workers are identified, how income is classified, and how consumption and components of the household balance sheets are measured.

A.1 Identifying Gig Economy Users

In order to identify gig economy income, I first need a definition of the gig economy. Because the time lag between first wanting to work and actually earning income is important for this study, I restrict to gig economy companies for which households could reasonably expect to earn income soon after signing up. I therefore exclude Airbnb or Etsy, for instance, where demand is more uncertain. I also focus on companies that have low barriers to entry, so this excludes the so-called "expert" economy, like Fiverr.

Gig income is defined as credits that appear in a checking or savings account. I also check that there are no corresponding debits from the same checking account that would indicate that the credits are a return credit. Gig economy employers typically pay weekly or higher frequency, whereas traditional employers pay at biweekly frequency. To ensure I am not capturing any traditional employees at gig firms, I drop any individuals for whom the typical gap between gig paychecks is 13, 14 or 15 days.

Event Dates

I determine a starting gig based on when I first observe income greater than \$1.[1]

[1] This will exclude most account verifications, which could also be a useful event date to determine when a household first considered working in a gig economy job. However, analysis in Koustas

The second important restriction I make here is that the account where gig income is first received must be observed for a preceding period before starting the gig job (the "pre period"). A household can sign up for the personal finance app after starting the gig job, which would mean they would only be observed in a "post period." Since gig work can sometimes have gaps between weeks worked, I choose 4 weeks of lead time as my cutoff. I also restrict to households that remain in the data at least 4 weeks after first receiving gig income.

Final Gig Economy Sample Restrictions

Some households sign up for a gig company causally to "try it out" or "see what it is about." My final gig sample includes all households who have at least 3 weeks of gig income (excluding any small transactions associated with account verification) at any single gig company (not necessarily the first one they first sign up for). I also drop a small number of observations that make extremely large amounts of money on the platforms (maximum income in any week greater than $3000) since these are likely professionals operating on the platform.

A.2 Data Definitions

Total Spending and Assets

My definition of total spending is net of identifiable payments and transfers. I calculate total spending and average balances of bank accounts and credit cards over the calendar week (Sunday-Saturday). The app data have a large right tail for consumption and asset values. To ensure outliers are not driving results, all values are winsorized at the 1 percent level.

(2018) found that this event date contains little signal.

Appendix B

Appendix for Chapter 2

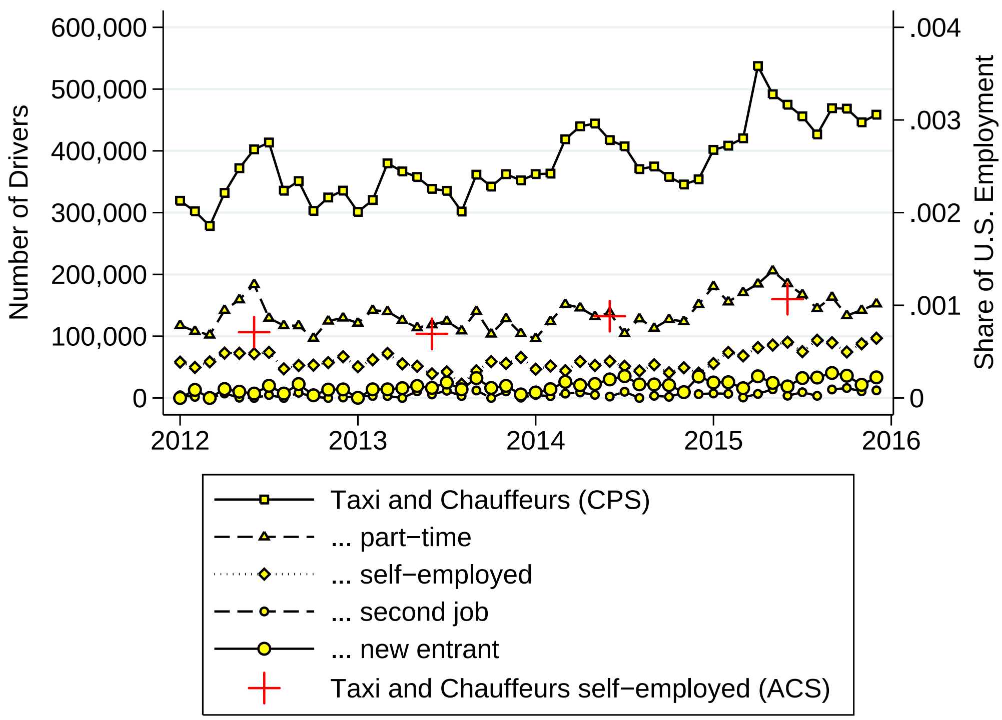
B.1 Supporting Evidence from Household Surveys

Evidence on the Taxi and Chauffeur Industry

Previous work measuring non-traditional work arrangements has highlighted that contingent workers are not well-captured in the major U.S. survey datasets (Katz and Krueger, 2016), although this work has not focused on ride-share drivers specifically. Figure 2.1 shows the number of workers reporting "Taxi and Chauffeur" as their main occupation in the CPS. In 2016, these workers represented 0.3 percent of the workforce, up by approximately 100,000 workers since 2012. While this increase is not large enough to be consistent with the growth from ridesharing, it is possible that it does capture at least part of the rise in ridesharing. However, digging closer into the composition of this rise suggests otherwise. Figure B.1 shows that none of the rise comes from self-employed or dual job holders, which is where we would expect rideshare jobs to show up.

Figure 2.1 showed that there are equal number of rideshare workers as there are traditional taxi workers by 2016. On an hours-adjusted basis, however, traditional taxis still have the edge since many are working full-time. Nevertheless, this is a large supply increase in the transportation sector. Whether this has had an impact on incumbent workers is an important consideration for policy as well as economic modeling. While it appears to be the case that traditional surveys like the CPS do not capture gig economy income, the traditional taxi sector should be well-represented. I next use the CPS to examine total earnings, wages and hours in the "Taxi and Chauffeur" sector. Time series for these variables are plotted in Figure B.2.

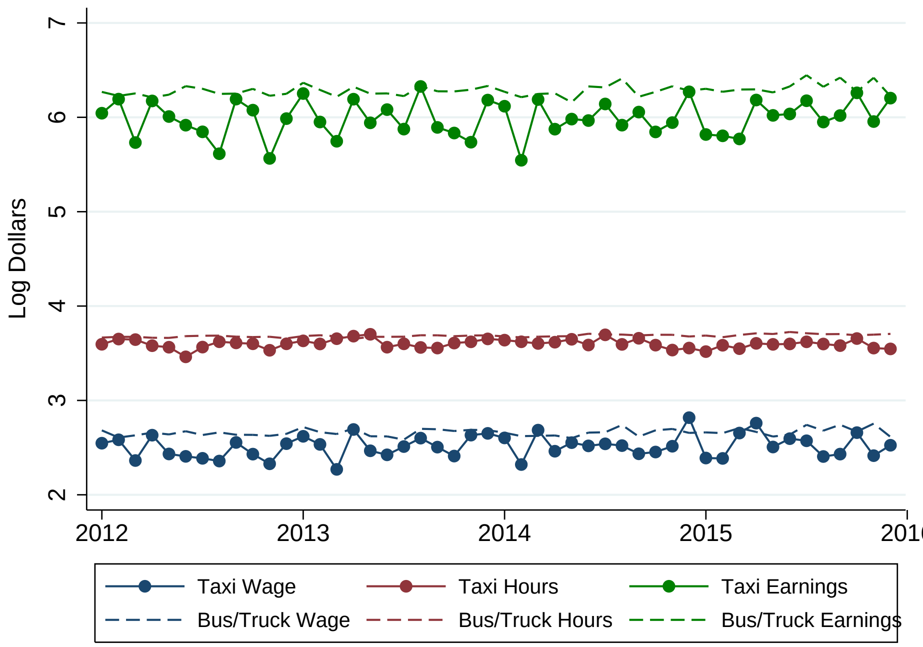
Figure B.1: Taxi Drivers in the CPS: Employment Composition



"Taxi and Chauffeurs (CPS)" is the weighted count of currently employed individuals with the occupation code "Taxi and Chauffeurs" (occupation code [*peio1ocd*] 9140) in the Current Population Survey Basic Monthly Files. "Taxi and Chauffeurs (ACS)" is the comparable statistic from the American Community Survey (occupation code *peio1ocd* 9140). The ACS occurs throughout the year, and so I assign ACS estimates to mid-year.

As shown in the solid lines in the figure, average wages, hours and earnings are completely flat in the taxi sector. In a related analysis, Berger, Chen, and Frey (2017) examine employment and annual earnings of "Taxi and Chauffeurs" in ACS data using a triple-difference research design comparing outcomes pre- and post-Uber's launch, relative to outcomes for bus and truck drivers. They find no effects in their most robust specifications controlling for time trends. Their paper uses different event dates—the launch of Uber Black, not Uber X—and of course the ACS has only low frequency (annual) variation. Unfortunately, sample sizes in the CPS monthly files are too small to exploit cross-sectional variation (the median CBSA has just 3 taxi drivers per quarter in a MORG sample). Figure B.2 shows nominal values for

Figure B.2: Taxi Drivers vs. Bus/Truckers in the CPS: Hours and Wages



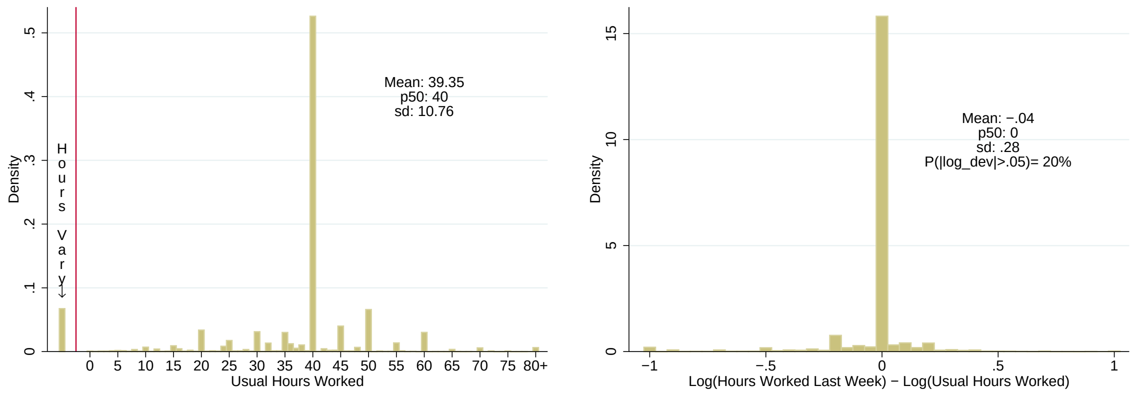
Source: Current Population Survey Basic Monthly Files. Hours worked (*pehrusl1*) and hourly wages (*prernhly*) for households in the Merged Outgoing Rotation Groups. "Taxi" refers to "Taxi and Chauffeurs" (occupation code [*peio1ocd*] 9140) while "Bus/Truck" refers to "Bus drivers" (9120) and "Driver/sales workers and truck drivers" (9130).

wages and earnings; we might be concerned that the counterfactual values would have increased over this period. I follow Berger, Chen, and Frey (2017) and compare taxi drivers to bus and truck drivers. The time-series for these groups are shown in dashed lines, and look very similar to taxi drivers. This suggests that the increase in supply has largely been accommodated by increased demand for transportation services.[1]

[1] The value of taxi medallions has declined. Since this decline does not show up in wages for taxi drivers, one interpretation is that taxi drivers themselves are not the residual claimants on the rents.

Figure B.3: Hours and Hours Deviations in the CPS

(a) Usual Hours Per Week (b) Log(Hours Last Week) - Log(Usual Hours)



Panel A reports usual hours worked as reported by households in the CPS. Panel B shows log deviations between hours households report working last week, compared to their usual weekly hours. Source: CPS Basic Monthly Files, 2013-2016.

B.2 Evidence on the Hours Process

The CPS provides evidence on hours at the monthly level. The CPS has a number of limitations: for instance, it only records rounded hours and there is likely significant recall error. Even with these limitations, there is a surprising amount of volatility in hours in any week. In Figure B.3, I plot usual weekly hours, and the deviation of hours worked last week from usual weekly hours. 6 percent of households in the sample report hours usually vary, and so this deviation cannot be calculated. In addition, 20 percent of households on any given month report not being within 5 percent of usual hours, a substantial amount of hours volatility.

B.3 Construction of Main Sample

Sources of income (and spending) are not pre-categorized or organized in the app data in the same manner as traditional survey or other datasets. Instead, we see only rows with raw transaction strings and amounts. This appendix describes how I identify rideshare drivers, how I measure consumption spending and household balance sheets, and how I construct a "control group" of coworkers.

Classification of Transactions Using Machine Learning

The app data are in the form of raw transaction strings. This is an important difference compared with consumption survey data like the CEX, which come pre-categorized into universal classification codes (UCC), or AC Nielsen data, where UPCs which can be easily aggregated into categories of goods. In this section, I describe how I use information in the transaction strings to categorize spending into different categories of goods. Gelman, Gorodnichenko, Kariv, Koustas, Shapiro, Silverman, and Tadelis (2016) use a binary machine learning (ML) model to categorize spending into "gasoline" and "not gasoline." I extend the binary ML model to multiple classes of goods.

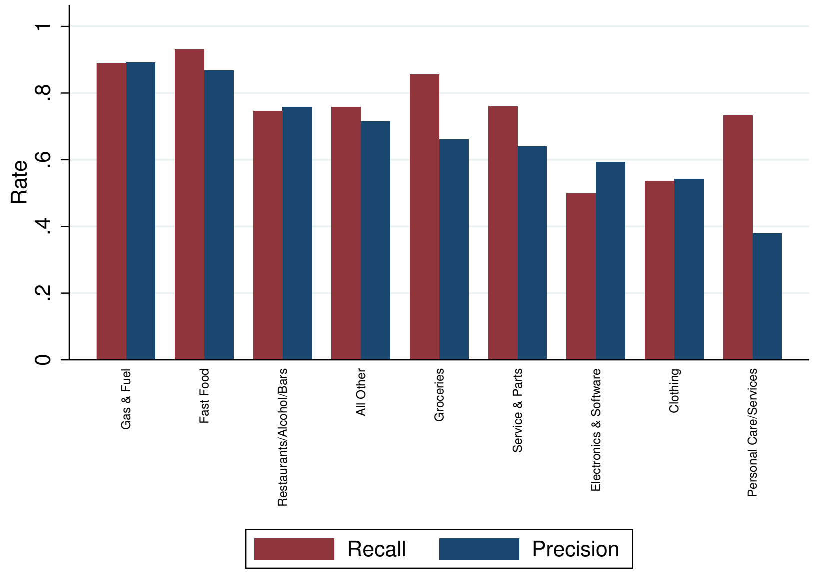
The ML procedure requires both a "training" data set—data actually used to fit a classification model—and a "testing" data set to evaluate the out of sample performance of the model. Two account providers in the app data report merchant category codes (MCCs) in their transaction strings. MCCs are four digit codes used by credit card companies to classify spending and are also recognized by the U.S. Internal Revenue Service for tax reporting purposes. I manually classify the many MCC codes into the following 10 categories: gas & fuel, fast food, restaurants/alcohol/bars, groceries, auto service & parts, electronics & software, clothing, personal care/services, travel spending, and all other spending. These two accounts with MCC codes represent about 3% of all app transactions.

I use the larger of the two account providers with MCC codes as the training data set, and test the performance of the model on the smaller account. I explicitly set aside the second account provider as the training data set because transaction strings, which we will feed into the model to classify the data, can differ across account providers. Therefore, if we train on data from the two accounts, we may fit our two cards extremely well, but we may have a poor "out of sample" fit of our model.

The model I use is a random forest classifier, which fits a number of separate decision trees to bootstrapped samples of the data; the final decision rule is the majority rule over the models. A decision tree is a series of classification rules that ultimately lead to a classification of a purchase. The rules, determined by the algorithm, minimize the decrease in accuracy when a particular model "feature" is removed. The features used to train the model are the transaction values (rounded to the nearest 50 cents) and a "bag of words"—individual words that appear in the transaction strings.

Two summary statistics commonly used to asses the fit of a multiclass model are

Figure B.4: Recall and Precision by Category



"recall" and "precision," which are defined as follows:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negatives}} \tag{B.1}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positives}} \tag{B.2}$$

These two summary statistics of model fit for each category of spending are shown in Figure B.4.

My ML model is able to predict gasoline spending and fast food spending particularly well, with around 90 percent recall and precision. Restaurant spending, all other groceries, and service and parts, have recall of around 75 percent and precision of around 70 percent. Electronics, clothing and services, in particular, are less well-predicted by the model, likely because of the larger variety of transaction strings associated with spending in these categories.

Rideshare Income

I search transaction strings in the app data associated with rideshare income. Rideshare income is paid weekly and on the same day, which is one way to distinguish it from traditional employees at a firm, who are typically paid on a biweekly schedule. In practice, since many rideshare drivers do not work every week, my sample keeps households with a modal gap between rideshare paychecks of 7 days.[2] Approximately 18,000 ever-rideshare drivers satisfy this criteria in the app data.

Knowing the timing of signing up for Uber is slightly complicated. Before 2016, Uber would first verify a new account by making a \$0.01 deposit. This indicates that the user has signed up for Uber. Interestingly, there can be large gaps between account verification and first rideshare earnings. Appendix figure B.6 shows the distribution of this gap. I consider a household as starting Uber the week *before* at least \$1 of earnings are observed (since income is lagged 1 week from first working). Figure B.5 compares the number of ever-rideshare drivers in the app data to data from Uber reported in Hall and Krueger (2016). Comparing the two series, the app contains approximately 1-2 percent of ever-drivers and entry follows a similar growth trend.

Consumption and Assets

For the baseline sample, I calculate total spending and average balances of bank accounts and credit cards over the calendar week (Sunday-Saturday). The app data have a large right tail for consumption and asset values. To ensure outliers are not driving results, all values are winsorized at the 1 percent level.

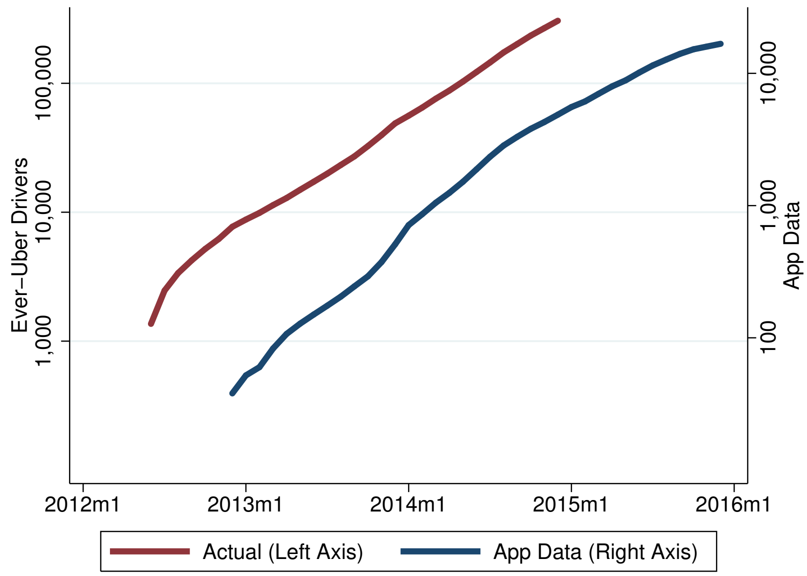
I construct a second subsample of rideshare drivers and matched coworkers with regular biweekly earnings. One biweekly paycheck is defined as being paid 14 days ago and being paid 14 days from now (therefore, three paychecks in a row must be observed).[3] Biweekly earnings are the most common type of earnings process in the United States. While I could also aggregate the data to monthly level to capture weekly and monthly earners, this would cut the number of time-series observations in half; given that the time-series is already rather short (whereas the number of observations is relatively large) this is undesirable. Moreover, if most workers are paid biweekly, this is the closest to the actual decision-making time frame of the household. Households need to make consumption and asset decisions to get them through until the next paycheck.

In some cases, I observe payrolls from multiple employers in an account. This could be because the household has dual earners or from multijob holding; unfortunately, I am unable to separately differentiate these in the app data. If there are

[2] As we move later in the sample, some drivers switch to "instant pay," I modify the restriction so that the modal gap is for pay received on paydays with everyone else, which could be satisfied if the driver switched from a regular payment scheme earlier in the sample.

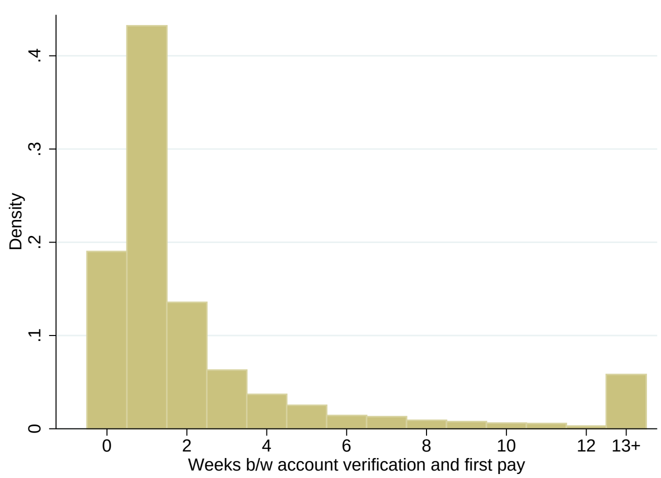
[3] In implementation, I look for income over the window 13-16 days after last income.

Figure B.5: Cumulative Count of Uber Drivers in the App v. Total Uber Drivers



"Actual" shows the cumulative sum of "Number of New Driver-Partners Starting Each Month in the United States," from Figure 3 of Hall and Krueger (????). "App Data" is the one-week lag of the number users in the app who have their first observed earnings in the indicated week.

Figure B.6: Gaps Between Account Verification and First Rideshare Pay



multiple employers, I first sort each employer by average log payroll amount. The biweekly pay series with the highest average log pay is considered the main employer from its first to its last observed receipt. Figure B.7 provides a hypothetical example of an individual with multiple observed income streams. In this example, my algorithm treats the series highlighted in green as the primary employer from January 26, 2017 through March 9, 2017. There is a break between the start of a new job on March 30, 2017, and this new job then becomes the primary employer. The period from March 16, 2017 through March 23, 2017 will be dropped. In addition, this hypothetical household has weekly income from another source, highlighted in orange. "Total" payroll income over the period January 26, 2017 through February 2, 2017 will be \$3,000 (\$2,000+\$500+\$500).

The timing of consumption and assets I construct is illustrated in Figure B.8. I measure spending over the 14 days following paycheck receipt, assuming that this consumption decision is made following the receipt of income and starting period assets. In the typical U.S. paycycle income received at date t reflects hours worked in the previous two weeks, usually with a lag of one week in between. Therefore, since the paycycle ended a week earlier, income should largely be known with certainty at the close of the pacycle. Nevertheless, I find spending is most responsive in the two weeks after receipt of the income. This suggests credit constraints, inattention, or complexity limit the full understanding of arriving income. It is also possible that people time their consumption with their income for behavioral reasons.

For the subsample of biweekly earners, I calculate measures of the household balance sheet the day before payroll income is received, which is consistent with the theoretical literature on consumption, as in Kaplan and Violante (2014).

Control Group

I next identify a group of non-rideshare drivers so that I can isolate common shocks and trends. While we do not observe demographics in the app data, I do observe sources of non-rideshare income. I use this information to construct a control group of other households in the data who receive income from the same payroll employer. In particular, I restrict to a common set of employers shared with rideshare drivers in the six months predating their starting rideshare.

As in Baker (forthcoming), Ganong and Noel (2017), and Gelman, Kariv, Shapiro, Silverman, and Tadelis (2014), I identify payroll income as income containing transaction strings like "payroll" or "salary." These transaction strings are then processed using an algorithm to extract the name of the employer. Using this information, I can identify coworkers at the same firm. In most cases, only the name of the firm can be extracted, not the establishment.

Figure B.7: Identification of Primary Employer

Date	Current Employer	Payroll Employer #1	Payroll Employer #2	Payroll Employer #3
...				
5-Jan-17				$500
12-Jan-17				$500
19-Jan-17				$500
26-Jan-17	1	$2,000		$500
2-Feb-17	1			$500
9-Feb-17	1	$2,000		
16-Feb-17	1			
23-Feb-17	1	$2,000		
2-Mar-17	1			
9-Mar-17	1	$2,000		
16-Mar-17				
23-Mar-17				
30-Mar-17	2		$2,000	
6-Apr-17	2			
13-Apr-17	2		$2,000	
20-Apr-17	2			
27-Apr-17	2		$2,000	
...				

The figure provides a hypothetical example of an individual with multiple observed income streams. In this example, my algorithm treats the series highlighted in green as the primary employer from January 26, 2017 through March 9, 2017. There is a break between the start of a new job on March 30, 2017, and this new job then becomes the primary employer. The period from March 16, 2017 through March 23, 2017 will be dropped. In addition, this hypothetical household has weekly income from another source, highlighted in orange. "Total" payroll income over the period January 26, 2017 through February 2, 2017 will be $3,000 ($2,000+$500+$500).

Figure B.8: Biweekly Payroll Timeline

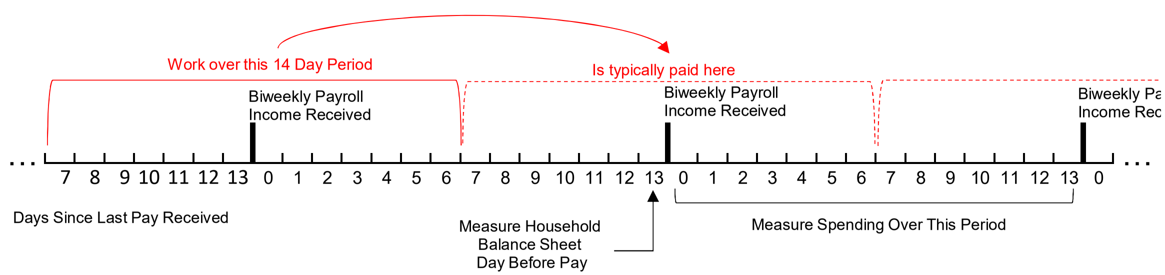


Figure shows the timing of consumption and assets for the sample of biweekly earners. I measure consumption spending over the 14 days following paycheck receipt (black brackets). Measures of the household balance sheet are calculated on the day before payroll income is received (thick black line). In the typical U.S. paycycle income received at date t reflects hours worked in the previous two weeks, usually with a lag of one week in between (red brackets).

Estimation Sample

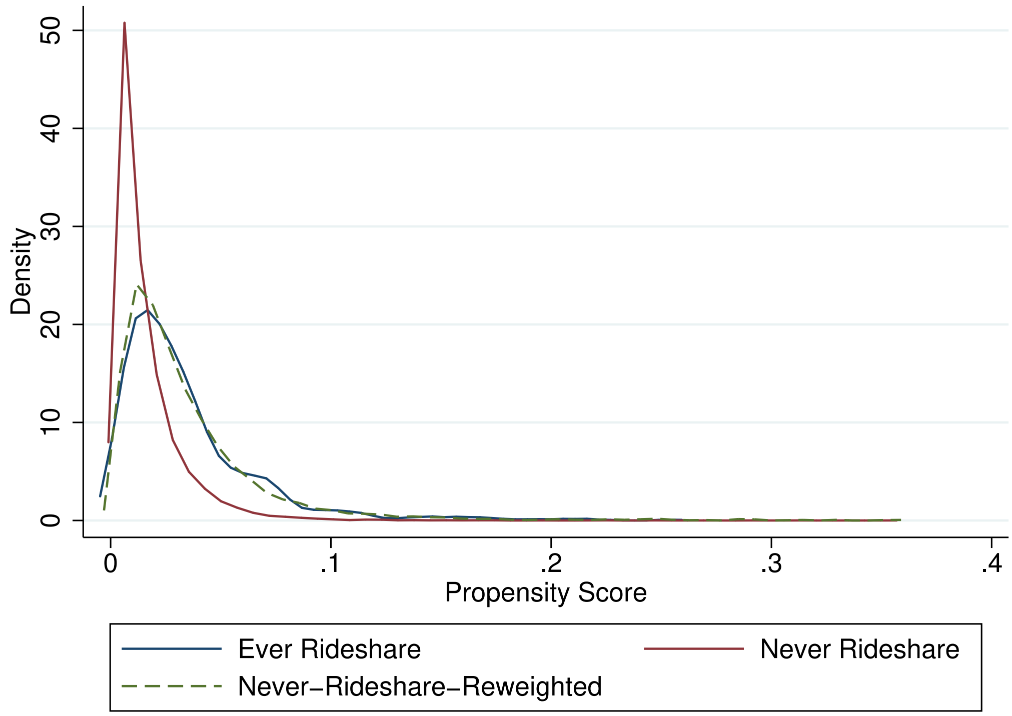
To construct the final sample used in estimation, I restrict to ever-rideshare drivers in the sample at least 6 weeks before I observe the first rideshare income and staying in the sample for at least 4 weeks afterwards. This assures that the pre- and post-period are being estimated off of the same households. It is also necessary to have a longer lead time due to the nature of gig-economy employment: this work tends to be highly variable from week to week. Even if no rideshare economy is immediately observed once a household enters the sample, it's possible the household is just not working that week. Restricting the lead time to at least 6 weeks in the sample before rideshare income is first observed deals with this issue.

B.4 Weighting Control Group to Balance Covariates

The descriptive statistics in Section 2.3 showed that rideshare drivers and their matched coworkers differ in levels of income and assets. While time-invariant level differences can be handled econometrically by fixed effects, the treatment and control groups could differ in important ways, such as their consumption response to income shocks and consumption trends. For instance, the model predicts that consumption growth will be very different for constrained versus unconstrained households. Intuitively, we would not want to compare a CEO with a cashier.

To test whether this is an issue, I reweigh the biweekly payroll sample using inverse-propensity-scores to match covariates for ever-rideshare drivers in 2013. I run a logit regression regressing an ever-rideshare indicator on the following covariates: credit utilization, indicators for city, the time-series standard deviations of log spending, log total income and log payroll income, and quartile indicators of spending, total income, payroll income, bank balances, available credit card balances, and net balances. Quartile indicators are used to be non-parametric and to deal with skewness. I choose 2013 because only few drivers have begun driving at this point. Moreover, because the values are all contained in the same calendar year, it is not necessary to control for calendar time, a problem that would arise if I focused on the actual pre-period before Uber entry into each city. Table B.1 shows the reweighted descriptive statistics, and Figure B.9 shows the propensity scores before and after reweighting. After reweighting, the descriptive statistics, particularly debit variables, move much closer in line.

Figure B.9: Propensity Score: Before and After Reweighting Control Group



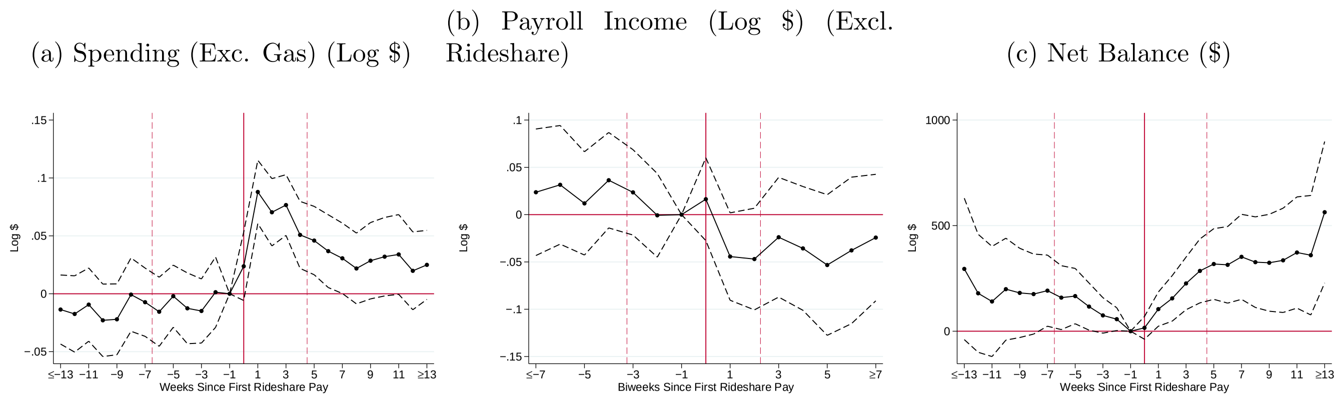
B.5 Additional Results

Continuously Employed

The reduced-form findings for the full sample make it clear that many rideshare drivers face a severe reduction in main job earnings before beginning rideshare. This sample differs from my biweekly sample because I do not restrict to households with strict biweekly income. Results for this group for log weekly spending (excluding gasoline), payroll income over a biweekly period, and weekly net balances, are shown in Figure B.10.

Unlike for the full sample, Figure B.10 Panel A, indicates that this group sees small, long-run gains in consumption spending. Panel B shows no statistically significant short or long-run losses in payroll income. In Panel C, net balances decline in the weeks leading up to starting rideshare, although in the long-run, net balances recover.

Figure B.10: Event Study: Employed Subsample



In these figures, the sample is restricted to households receiving payroll income at least monthly. Panel A plots the event-study coefficients from estimating equation 2.12 for log total spending, excluding gasoline spending. In panel B, the dependent variable is payroll income, which is aggregated over a biweekly period. In panel C, the dependent variable is net balances (bank balance - credit card balance). The area between the dashed vertical lines indicates the coefficients are estimated on a balanced sample. 95% confidence intervals are shown in dashed lines around the main estimates. Dependent variables are winsorized at the 1% level.

Table B.1: Descriptive Statistics (2013)- Reweighted

	(1)		(2)		(3)	
	Rideshare Drivers		Control, Unweighted		Control, Reweighted	
	Mean	Median	Mean	Median	Mean	Median
Spending	2401.28	1902.70	3047.17	2389.36	2611.01	2049.21
SD_t	1299.47	904.20	1722.76	1250.49	1347.94	967.39
Log Spending	7.44	7.44	7.61	7.61	7.49	7.49
SD_t	0.51	0.48	0.58	0.53	0.52	0.49
Income	2515.94	2088.27	3172.48	2545.51	2727.91	2157.23
Log Income	7.52	7.52	7.70	7.72	7.57	7.56
SD_t	0.47	0.44	0.47	0.44	0.46	0.43
Payroll Income	1302.35	1163.27	1787.13	1567.85	1355.88	1211.62
Log Payroll Income	6.98	7.03	7.26	7.33	7.01	7.07
SD_t	0.21	0.16	0.20	0.15	0.21	0.16
Bank Balance	3692.81	803.02	9275.27	2626.49	4299.31	1098.93
CC Balance	4821.91	2464.64	3404.96	1614.84	4781.34	2002.00
CC Utilization Rate	0.45	0.38	0.38	0.29	0.48	0.44
Net Balance	-736.73	-889.24	5852.29	632.57	-543.14	-578.15
Observations	1220		58634		58634	

Long-run Rideshare Drivers

About 25 percent of rideshare drivers appear to permanently cease rideshare within the first quarter after starting. Perhaps they've learned something about their type– that they find driving to have more disutility than they originally thought– or maybe they are kicked out by one of the rideshare companies, such as for having a low rating. The overall results are mixing together a group of rideshare stayers and leavers. I also run the specification focusing on households that maintain at least some payroll income, and whose last rideshare observation is outside of the 1 quarter window. Results (not shown) are almost identical.

Figure B.11: Uber's Staggered Geographic Entry

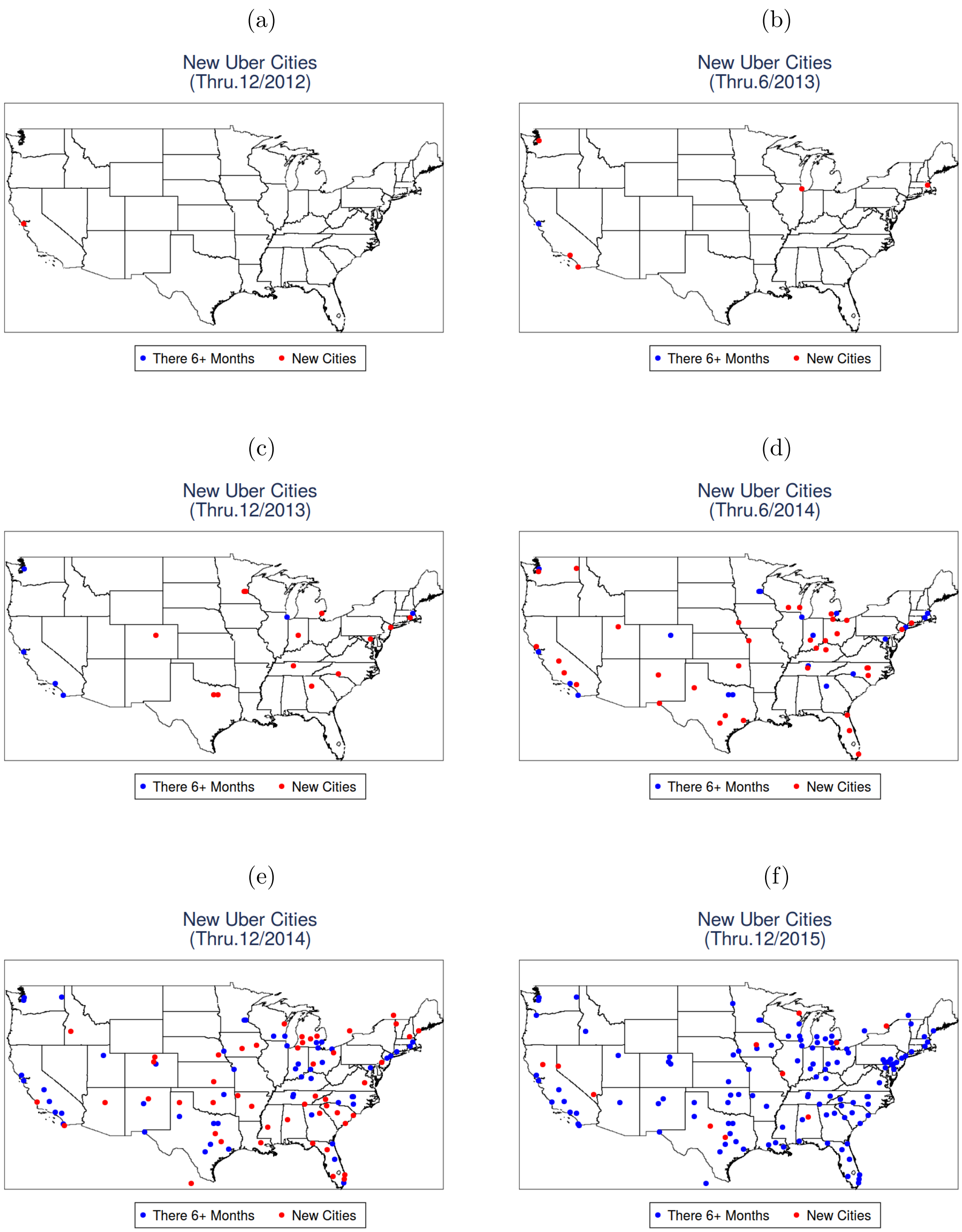


Table B.2: Empirical Markov Transition Matrix

	-1.0	-0.79	-0.59	-0.37	-0.16	0	0.17	0.38	0.59	0.79	1.0
-1.0	0.092	0.018	0.023	0.038	0.101	0.535	0.117	0.04	0.016	0.008	0.012
-0.79	0.02	0.057	0.03	0.045	0.116	0.541	0.119	0.038	0.015	0.009	0.011
-0.59	0.011	0.014	0.069	0.057	0.117	0.542	0.119	0.038	0.015	0.008	0.011
-0.37	0.006	0.006	0.017	0.099	0.142	0.54	0.125	0.038	0.014	0.006	0.007
-0.16	0.004	0.003	0.008	0.031	0.187	0.597	0.121	0.03	0.01	0.005	0.006
0	0.003	0.002	0.005	0.016	0.081	0.754	0.101	0.022	0.008	0.004	0.005
0.17	0.003	0.003	0.006	0.022	0.097	0.607	0.202	0.038	0.011	0.005	0.005
0.38	0.005	0.004	0.008	0.027	0.099	0.552	0.157	0.097	0.028	0.012	0.012
0.59	0.005	0.004	0.009	0.026	0.093	0.524	0.129	0.079	0.076	0.028	0.027
0.79	0.005	0.004	0.01	0.026	0.088	0.52	0.113	0.065	0.056	0.063	0.051
1.0	0.006	0.004	0.009	0.023	0.087	0.515	0.097	0.052	0.042	0.039	0.124