THE UNIVERSITY OF CHICAGO

ESSAYS ON LABOR SUPPLY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY
KUAN-MING CHEN

CHICAGO, ILLINOIS
JUNE 2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am extremely grateful to my advisors, Stéphane Bonhomme, Magne Mogstad, and Alessandra Voena. Chapter 1 is my job market paper. I thank Ming-Jen Lin and Shiau-Fang Chao for helpful comments and valuable support for data. I am also grateful to Yu-Chang Chen, Angela Denis, Ning Ding, Maxwell Kellogg, Jiaming Soh, Kai-Jie Wu, Sung-Ju Wu, Yu-Hsuan Yen, and seminar participants at Labor student group, Econometrics working group, and public-labor seminar at the University of Chicago for detailed comments. I thank Si-Hung Foundation for financial support and the Health and Welfare Data Science Center for data provision.

Chapter 2 is a joint work with Ning Ding, John List, and Magne Mogstad. The opinions expressed in that chapter are those of the authors alone and do not reflect the views of Uber. We appreciate the constructive comments and suggestions from Oystein Daljord, Jonathan Hall, Vishal Kamat, Libby Mishkin, Jesse Shapiro, Winnie van Dijk, and Kerry Smith.

ABSTRACT

This dissertation consists of two parts. The first part concerns adult children's trade-offs between working and providing long-term care (LTC) to parents. How do commonly implemented LTC policies, such as tax deductions, in-kind transfers, and international caregivers' eligibility, affect such trade-offs? What are the welfare effects of these policies? These questions have become increasingly important due to the number of people affected and the costs adult children incur. Using data from Taiwan, I first document that children are 4 percentage point less likely to participate in the labor market when parents' LTC needs arise, with daughters, the less educated, and older children having the largest decreases in labor supply. Only a small share of children return to the labor market if their LTC-needing parents pass away. Motivated by the descriptive findings, I then build and estimate a dynamic labor supply model, combining the descriptive evidence with an exogenous variation in caregivers' prices from a policy reform in Taiwan. The model features costs of returning to work, endogenous health processes, and unobserved heterogeneity in care and labor market skills. Model-based results suggest large costs of returning to work, especially for daughters and the less educated. Typical LTC policies, such as LTC tax deductions and relaxing the eligibility criteria for hiring international caregivers, alleviate the effects of costs of returning to work on labor supply under LTC needs, in part because of work incentives these policies provide.

In the second part, we quantify how labor supply elasticities and reservation wages vary between people and over time, and infer workers' valuation of flexibility in their choices of work hours. Economists and policymakers are keenly interested in these quantities, especially lately with the growth in jobs that offer flexible work schedules. Our study takes advantage of a large natural field experiment at Uber, the largest ride-sharing company. Combining this experiment with high frequency panel data on wages and individual work decisions, we estimate a dynamic labor supply model that let us recover reservation wages, labor supply

elasticities, and workers valuation of flexibility.

CHAPTER 1

UNDERSTANDING ADULT CHILDREN'S LABOR SUPPLY RESPONSES TO PARENTS' LONG-TERM CARE NEEDS

1.1 Introduction

How do adult children trade-off working and providing long-term care (LTC) to parents? How do commonly implemented LTC policies, such as tax deductions, in-kind transfers, and international caregivers' eligibility, affect such trade-offs? What are the welfare effects of these policies? These questions have become increasingly important due to the number of people affected and the costs adult children incur. More than 10% of those aged 65 and over have LTC needs worldwide, and with a global trend of aging population, the number of people with LTC needs will grow substantially. Responding to these needs, governments in OECD countries have been spending 0.3% to 3.5% of GDP on LTC policies annually, and that number is expected to grow with the aging population.[1]

This paper addresses the questions above empirically using a dynamic labor supply framework. I study the Taiwanese context, in which children are expected to be responsible for their parents' care arrangements, a characteristic typical in East Asian countries. Many developed countries implement the policies I analyze, but Taiwan offers several advantages to studying these policies. One of these advantages is that I can exploit a major reform in Taiwan of international caregiver hiring eligibility implemented in September 2012. This reform provides exogenous variation which can be used to identify the opportunity cost of hiring a caregiver. Another advantage is that I can combine multiple data sources, including the Taiwan Longitudinal Study in Aging (TLSA) and link them with the National Health Insurance Research Database (NHIRD) during empirical analysis.

I begin with several descriptive analyses. I estimate dynamic labor supply responses of children when parents' LTC needs arise and when LTC-needing parents pass away. Findings from these analyses guide subsequent modeling choices. Effects of the reform and other data moments recover model primitives. I then use the estimated model to calculate labor supply elasticities and reservation wages, and conduct various counterfactual analyses that quantify

1. Source for LTC needs in OECD countries: [20]

disparities in labor supply paths due to parents' LTC needs. Finally, I calculate fiscal costs, compensating variations, and labor supply responses of typical LTC policies to address the research questions.

In Section 1.2, I introduce the background of LTC, including a definition of LTC needs, the scale of LTC needs, and typical LTC arrangements worldwide. In Section 1.3, I present data, summary statistics, and four findings. The first is that adult children's labor supply drops significantly when parents' LTC needs arise. Compared with those having LTC needs later and with those without LTC needs, the labor market participation decreases by 4 percentage points when parents' LTC needs arise. The drops are persistent and increasingly negative over time. Furthermore, the labor supply starts decreasing before the onset of LTC needs, consistent with a smoothly decaying health process of the parents.

The second finding is substantial heterogeneity in children's labor supply responses when LTC needs arise. Daughters are 300% more likely to leave the labor market than sons are, and 30% more likely than children-in-law. Heterogeneity also exists along education and age dimension. Lower-educated children reduce their labor supply more when parents' LTC needs arise, consistent with the lower opportunity cost of providing care themselves. Younger children decrease their labor supply less in response to the parents' LTC needs. The costs of returning to the labor market might explain this age pattern. Since younger children expect a longer period before retirement and after their parents' deaths, the cost of returning to the labor market more significantly deters younger children from leaving the labor market.

The third finding is that children return to the labor market after their LTC-needing parents pass away, but the probability of returning is 25% smaller in comparison to the drop at the onset of LTC needs. This finding is consistent with costs of returning to work, motivating the choice of a dynamic model that features adjustment costs of entering the labor market. In a static model without such costs, children's labor supply would return to the same level after their LTC-needing parents pass away, as if their parents have never experienced LTC needs.

The fourth finding is that being eligible to hire an international caregiver increases children's labor supply in comparison to those ineligible. Exploiting the reform regarding such eligibility, I find that hiring eligibility increases the labor supply by 6 percentage points immediately. This estimate serves as a key data moment in the model to identify opportunity costs of hiring a caregiver.

The estimates of labor supply effects discussed above are informative of how individuals respond to parents' LTC needs, but they are insufficient for understanding the welfare effects of LTC policies. To understand the policy effects, I build a dynamic labor supply model in Section 1.4 that is informed by the empirical evidence on labor supply effects. In the model, an adult child chooses whether to work and hire a caregiver or not work and provide care herself. The parent's health evolves endogenously according to the care arrangements that the child makes. The key trade-offs that the child faces include the cost of hiring a caregiver, the payoff from the labor market, the parent's health evolution, and the potential costs of returning to the labor market. The model features both observed and unobserved heterogeneity. Sons, daughters, and children-in-law behave differently when dealing with LTC needs. Conditional on the relationship with the care-receiver, individuals vary regarding their abilities in the labor market and providing care to their parents.

I adapt [5] to estimate the model. Beginning with an initial guess of unobserved type distribution, I estimate the selection corrected health and wage processes. Next, I estimate the full model by simulated method of moments. Targeted moments include the share of working individuals conditional on education, parental health, lagged work status for each unobserved type, as well as effects of the eligibility reform. Section 1.5 discusses the estimation procedure and shows that the model replicates critical patterns in the data well. Besides the in-sample model fits, I study an eligibility reform in 2015 and show that the model replicates the out-of-sample reform effects closely.

The model delivers two key insights through counterfactual analyses, discussed in Section 1.6 and 1.7. The first is that LTC needs drive a large share of children out of the labor market, and only some children return after their parents' deaths. Furthermore, these effects from LTC needs show considerable heterogeneity. I begin by comparing two counterfactual scenarios—(i) healthy parents dying immediately without experiencing LTC needs and (ii) parents having LTC needs before their deaths. This comparison shows how much parents' LTC needs change children's career paths. I find that sons and daughters are 5% and 19%, respectively, less likely to participate in the labor market in scenario (ii) than scenario (i). Moving beyond scenarios (i) and (ii), I aggregate the parental health sequences in the data and find a similar pattern, with a 9% decrease for daughters. This magnitude is comparable to fertility effects on female labor supply in the Taiwanese literature. Typical LTC policies, including tax deductions and relaxing the eligibility for hiring international caregivers, reduce permanent leaves from the labor market by providing work incentives. In particular, allowing all children whose parents have moderate ADL to hire an international caregiver cuts permanent leaves from the labor market due to LTC needs by more than half.

The second insight from the model is the vastly different effects from common LTC policies, such as in-kind transfers and tax deductions. The different effects appear in (i) whether children stay in the labor market when parents experience LTC needs and (ii) the set of children benefited from the policies. I analyze in-kind transfers and tax deductions implemented in Taiwan starting from 2017 and 2020, respectively. In-kind transfers provide some hours of care service for those who do not hire caregivers, and tax deductions reduce taxable income for those whose parents have LTC needs. When LTC needs arise, a tax deduction program equivalent to subsidizing 5% of mean annual earning drives 3% fewer people out of the labor market. However, even with means tests, it benefits sons and higher educated individuals more than twice compared to daughters and lower educated individuals. On the other hand, the in-kind transfer program encourages 20% more permanent leaves from the labor market due to LTC needs. Nevertheless, it disproportionately benefits daughters and those with lower education. The stark contrast largely results from work incentives of these policies—the tax deduction only benefits those with income, while the government requires a child to provide care herself to be eligible for the in-kind transfer program.

This paper relates to a growing literature that addresses the economics of LTC. The key questions and findings have been summarized in [63] and [62]. Three strands of the literature are most relevant to the current study. The first strand studies the treatment effects of LTC needs on caregivers, which corresponds to the descriptive analyses in the current paper. [8] survey papers regarding how LTC provision affects informal caregivers' employment and health. Consistent with my findings, most papers find negative labor supply effects of such provision. [34] is the closest to my descriptive analysis. They employ an event study approach similar to my paper. Using Austrian data, they find large negative labor supply responses to unexpected parental health shocks, such as stroke and heart attack. My analyses further complement these results by examining children's labor supply patterns after parents' deaths.

The second strand of the literature evaluates LTC policies using a treatment effect framework. For example, [53] assess the expansion of formal LTC in Norway in 1998. They find that government-provided LTC services substituted for informal care provision. Another example is [34], they find that a reform legalizing migrant LTC workers in Austria in 2007 generated positive labor supply responses from informal caregivers. The reform I study changes rules with explicit health and age criteria, and thus offers suitable control groups to the treated individuals.

The last strand of the LTC literature uses model-based approaches to evaluate LTC policies, and such studies have diverse foci. [6] build an equilibrium model with intra-family bargaining to study LTC subsidies. Consistent with the current paper, they find that demographics of those affected by the policy are essential to determining the welfare effects of LTC policies, such as informal care subsidies. [61] assesses substitution between informal care and LTC insurance. She also finds that families place a large value on cash benefits over in-kind transfers. More closely linked to my setup, [68] builds a dynamic discrete choice model to investigate long-term career costs for daughter caregivers. The author focuses on job search dynamics and directly models the persistence in care provision as a part of the preferences. Similar to current findings, she finds a considerable value in staying in the labor market, in comparison to leaving and returning.

I contribute to the LTC literature in several ways. First, the East Asian context I study is important and mostly unexplored. Besides the large population, traditional norms on care arrangement make children's responses to parents' LTC needs much more salient than other contexts. In addition to the more profound effects, the context I study offers advantages for model identification from clear and strict caregiver hiring regulations. On top of the different context I examine, I contribute to the LTC literature by bring the three strands of the literature together and bridging descriptive analyses, a reform, and the dynamic model for policy analyses. The descriptive analyses connect tightly to the model I construct. The eligibility reform I exploit is directly informative for policy evaluations and useful for recovering structural model parameters. Guided by the descriptive and reform evidence, the model addresses key policy issues that widely apply to many contexts.

This paper also contributes to an extensive literature on immigrant workers. Debates over the costs and benefits of the immigrant workers attract a wide attention in the literature. (For example, [11] and [12].) Although the cost of increasing foreign workers to a destination country has been studied extensively, the benefits of doing so are difficult to measure. [22] and [23] provide examples in which foreign workers provide childcare and induce young women to participate in the labor market. The current paper similarly shows that foreign workers allow domestic workers to substitute labor market participation for time-consuming LTC provision, especially among female workers.

This paper also contributes to the literature on female labor supply and traditional norms, with recent studies investigating how policies affect cultural practices. [7] assesses matrilocality and patrilocality, finding that pension policies reduce the practice of these traditions. The current finding that daughters have the largest labor supply responses to parents' LTC needs reflects traditional social norms in East Asia (see [17] for a discussion). One prominent topic in such literature is whether policies narrow gender gaps in labor force participation. This paper contributes to the literature by suggesting that LTC policies, such as tax deductions and relaxing caregiver hiring criteria, increase female labor market participation.

1.2 Background

In this section, I first describe the definition of LTC needs. I then argue that LTC is an important issue by presenting the scale of the LTC needs. Finally, I discuss common care arrangements and LTC policies.

Definition of LTC Needs. I follow the definition of LTC needs in the literature, which defines it as the assistance necessary to perform at least one Activity of Daily Living (ADL). ADLs refer to the most basic functions of living, including grooming, toilet use, walking, etc.[2] ADL difficulty is commonly used as a major eligibility criterion for LTC insurance and government LTC programs.[3]

Scales and Costs of LTC Needs. A significant share of elderly people have LTC needs, and this share is increasing with age. Approximately 10% of the population aged 65 to 74

2. Standard ADL items include fecal incontinence, urinary incontinence, grooming, toilet use, feeding, transfers, walking, dressing, climbing stairs, and bathing. Difficulties with these activities are highly correlated. See the Appendix for more details on ADL measures.

3. In the United States, many Medicaid programs link eligibility to the number of ADLs. Most LTC-related policies in Taiwan, including those analyzed in this paper, also use ADL difficulties as part of eligibility criteria.

have at least one ADL difficulty, and about one-quarter of those over 75 worldwide have such difficulties. As the global population ages, the share of individuals with ADL difficulties will likely increase. In 2050, more than 30% of the population are expected to be over 60 in developed countries.

Addressing LTC needs is costly from a public policy perspective. Governments' LTC expenditures vary 0.3 to 3.5% of GDP worldwide, and such spending is typically in the form of in-kind transfers, such as residential care services. In comparison, average health spending in OCED countries is 8.8% of GDP. LTC expenditures account for nearly one-fifth of total health spending.

Comparable to other countries, the Taiwanese government estimates that 12.7% of those over 65 have LTC needs. The Taiwanese government spends about 0.3% of GDP on LTC policies, which is lower than that in many other countries, but it is expected to grow rapidly.

Who Provides Care? I divide care provision into hired care service and non-hired care service. The most common hired care service in Taiwan is live-in caregivers. One-third of the LTC-needing families hire live-in caregivers to provide 24/7 LTC service. Nearly all of these live-in caregivers are international caregivers, an arrangement common in East Asia. Another common form of hired caregiver is LTC institutions, such as nursing homes. Approximately one-quarter of LTC is provided by institutions.

Among non-hire, or informal, caregivers, the majority are spouses, sons, daughters, and children-in-law. The distribution of caregivers' relationships with care-receivers is similar across countries. Spouses and children account for similar shares of informal caregivers. Since care-receivers' spouses are usually older and retired, children are the main focuses of LTC needs' labor supply effects.

LTC Policies. I focus on three common LTC policies—(i) expanding or limiting the eligibility of hiring international caregivers, (ii) in-kind transfers, and (iii) tax deductions for

LTC. Among hired caregivers, foreign-born caregivers constitute an essential part of LTC workers, especially in East Asia and Southern Europe. Concerns over the stability of the foreign caregiver supply lie at the core of policy debates in many countries. How much they substitute for informal care provision is essential to evaluating costs and benefits of international caregivers, but the topic remains largely unanswered in the literature.

In-kind transfers and tax deductions for LTC have also been implemented broadly in developed countries. Governments generally provide in-kind transfers by hiring caregivers and assigning them to those who do not hire live-in caregivers. On the other hand, tax deductions usually benefit those who have wage income and do not provide care themselves. Salient policy questions include who benefits from such policies, how the policies (dis)incentivize the labor supply, and whether targeting a specific population increases welfare gains. Taiwan launched an in-kind transfer and a tax deduction program during 2017 and 2020, respectively, with their policy details still being debated. I base counterfactual analyses on these policies and address current discussions.

1.3 Descriptive and Reform Evidence

To analyze the effects of the LTC policies, it is essential to understand how children tradeoff work and care provision decisions in response to parental health statuses. I now present empirical findings on (i) dynamics of children's labor supply around the onset of their parents' LTC needs, (ii) dynamics of children's labor supply after the death of LTC-needing parents, and (iii) effects of the eligibility reform regarding hiring international caregivers.

1.3.1 Data

Main Dataset. This paper's primary dataset is the Taiwan Longitudinal Study in Aging (TLSA), a nationally representative sample of adult residents aged 60 and over from 1989 to 2011. The TLSA is a longitudinal dataset, surveying respondents approximately every three years and representing the Taiwanese counterpart to the Health and Retirement Study (HRS) in the United States.

The TLSA offers detailed information on health, ADL status, and household structure. Notably, it includes the respondents' family members' ages, marital statuses, education, and employment statuses, and such information is repeatedly collected during each wave. Importantly, this information allows me to investigate the effects of LTC needs on family members.

Auxiliary Dataset. I link the TLSA with the National Health Insurance Research Database (NHIRD) from 2007 to 2014. The NHIRD is the administrative record of the universal health insurance system, providing information on basic demographics, death records, and the employment statuses of the population. Importantly, the database can be linked with the TLSA using unique national identification numbers.[4]

The link is useful to this paper in two ways. First, an important reform to eligibility for hiring international caregivers occurred during 2012. Since the TLSA ended in 2011, the link with NHIRD extends data available to 2014 to cover the reform. Second, the TLSA stops collecting information from a respondent after her death. However, the NHIRD allows me to continue tracking her family members' information, an advantage of this linked dataset over the HRS. In HRS, it is difficult to track a family member's information after the respondent's death.

Unit of Observation. I construct my sample so that child-year is the unit of analysis. For example, if a respondent to TLSA has two children, they enter the sample as separate observations while sharing the same parental information. Since the goal is to assess children's labor supply decisions, I restrict the sample to those aged 25 to 65.

4. I describe this link in more detail in the Appendix.

1.3.2 Summary Statistics

	Relationship	(1) All Children Mean	SD	(2) LTC Children Mean	SD
Education	Daughter	9.68	4.20	8.54	4.22
	Son	11.01	3.82	10.03	3.78
	Children-In-Law	10.06	3.80	9.26	3.77
Work	Daughter	0.51	0.50	0.44	0.50
	Son	0.90	0.30	0.84	0.37
	Children-In-Law	0.60	0.49	0.51	0.50
Age	Daughter	44.26	8.04	48.03	7.91
	Son	44.26	8.06	47.87	8.07
	Children-In-Law	40.90	8.40	44.33	9.01
N	Daughter	7,085		2,344	
	Son	7,209		2,340	
	Children-In-Law	2,128		546	

Table 1.1: Summary Statistics

Notes: "LTC Children" includes those whose parents have LTC needs.

Table 1.1 reports summary statistics for the sample. Column (1) shows descriptive statistics for the full sample, and Column (2) restricts the sample to those whose parents have at least one ADL difficulty.

On average, sons have the highest education and work the most. The share of those who work varies considerably between sons and daughters. Approximately half of daughters and 90% of sons are working. The average age of the children is 43 years, and sons received 1.3 years more education than daughters did. Children-in-law are generally younger but otherwise similar to daughters.

In the sample of parents with LTC needs, both the education and the share of working individuals are lower. The difference in education is about a year, and the share of working individuals is about 4 to 7 percentage points lower. The ages are higher for these people, likely because parents with ADL difficulties are older than those without them.

1.3.3 Research Design for Descriptive Labor Supply Dynamics

Overview of Design

In the empirical analysis, I follow [31]'s design to analyze the dynamic labor supply pattern around the onset of parents' LTC needs and deaths. The goal is to compare patterns among adult children who experienced parental health status changes to comparable adult children who did not.

In the discussion that follows, I use labor supply dynamics when parents' LTC needs arise as the example of an outcome to help explain the research design. I refer to those whose parents experienced LTC needs as the "affected group." The effects of such health events on labor supply cannot be read off directly from the affected group because many observed and unobserved variables, such as aggregate time trends and children's age profiles of labor supply, might confound parents' health processes. I therefore construct two baseline groups that have not experienced these health events but are otherwise similar to those in the affected group.

The first baseline group comprises those whose parents have never experienced LTC needs. Guided by the summary statistics, I reweigh the children's age distribution such that the affected and baseline groups share the same children's age distribution. Therefore, the age profile of the labor supply is no longer a concern.

The second baseline group comprises those who would also experience parental LTC needs, but later in the sample. Those who belong to this group stay in the group until their parents' LTC needs arise, so that I avoid comparing two individuals who have both been affected by parents' LTC needs. By comparing the affected group with the second baseline group, I alleviate the concern that unobserved factors of parental health and children's labor supply correlate with and confound the effects of LTC needs. Since parents eventually also have LTC needs in the second baseline group, unobserved factors of parental health are thus similar to those in the affected group.

Formal Description

Formally, the comparison I discuss in this section is a set of difference-in-differences (DiD) estimates, consisting of two steps. The first is to construct the proper affected and baseline groups, and the second is to conduct the standard difference-in-differences procedure period by period. When reporting labor supply dynamics, including Figures 1.1, 1.2, 1.3, 1.4, and 1.6, each point in the figures represents a θ_t in estimation equation:

$$\theta_t = (y_t^T - y_t^C) - (y_b^T - y_b^C), \tag{1.1}$$

where y_t^T is the mean labor supply of the LTC-needing group, or the affected group, at time t, and y_t^C is the mean labor supply of the baseline groups at time t, and b the baseline period for comparison. t is the relative time period, where $t = 0$ denotes the period when LTC needs arise. I compare labor supply responses with the period just before LTC needs arose, setting $b = -1$. I also reweigh the children's age distribution such that the affected and baseline groups share the same children's age distribution. The composition of the baseline group changes across t since once a child's parent's LTC needs arise, the child is removed from the baseline group. By dropping such children from the baseline group, I avoid comparing two individuals whose parents have both had LTC needs already.

After reporting the dynamic patterns, I also show how labor supply responses vary depending on education and age. To focus on heterogeneity and report results compactly, I report, in Table 1.2, estimates from equation:

$$\begin{aligned} y_{it} = \alpha + \beta \text{LTC Need}_i + \gamma \text{Post}_{it} + \delta \text{LTC Need}_i \times \text{Post}_{it} + \\ \eta \mathbf{1}\{X_i = x\} + \theta \mathbf{1}\{X_i = x\} \times \text{LTC Need}_i \times \text{Post}_{it} + \epsilon_{it}, \end{aligned} \tag{1.2}$$

where LTC Need$_i$ equals 1 if in the affected group and 0 if in the baseline group, Post$_{it}$ equals 1 if it is after the period that LTC needs arose, and X_i denotes an individual characteristic, such as education. In this specification, I group periods into those before and after the LTC needs arose. Coefficient θ captures the heterogeneous response.

1.3.4 Labor Supply Dynamics When LTC Needs Arise

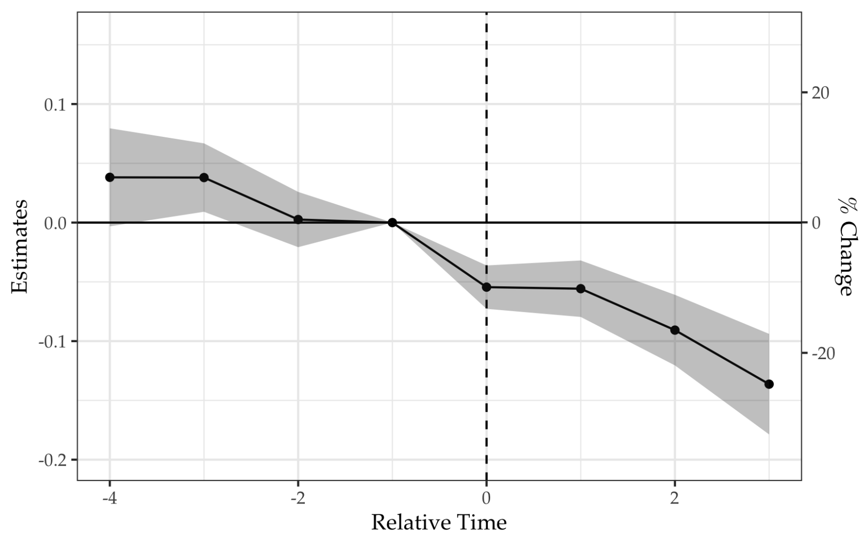


Figure 1.1: Labor Supply Responses for Daughters When LTC Needs Arise

Notes: The event is when parents first report any ADL. The outcome variable is the binary variable of whether a child works. The right y-axis represents the percent change relative to the baseline group mean of the baseline period. The baseline period is -1. Each event time corresponds to a wave of the TLSA. The shaded area represents the 90% confidence interval. Standard errors are clustered at the individual level. The sample includes daughters aged 25 to 65. The baseline group includes those who never have LTC needs and those who had LTC needs later. The samples are reweighed by the propensity score estimated by their age in the estimation.

I first investigate children's dynamic labor supply when parents' LTC needs arise. The magnitude of the responses, whether the responses persist, and whether there are anticipatory effects are important to understanding policy effects. As described previously, I compare the affected group, whose parents experienced LTC needs, with the two baseline groups. I start directly with reporting results by relationship with care-receivers and report the average effects in the Appendix.

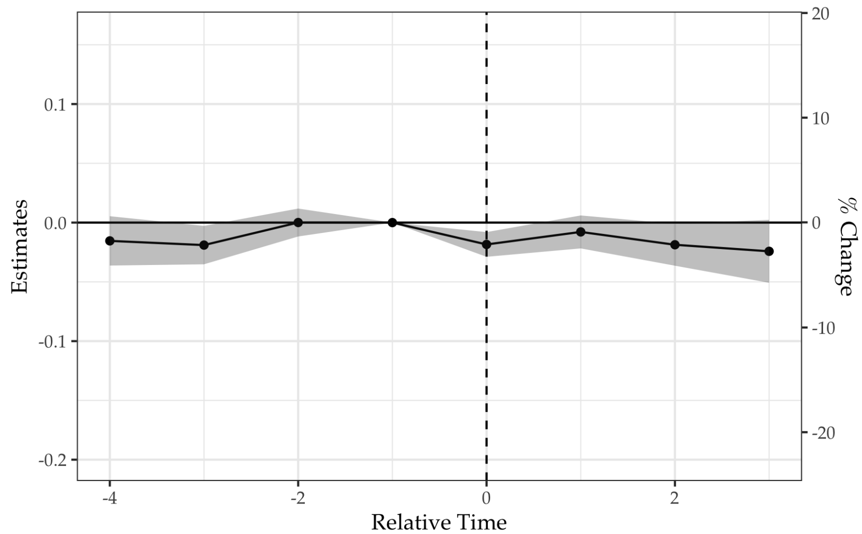


Figure 1.2: Labor Supply Responses for Sons When LTC Needs Arise

Notes: The event is when parents first report any ADL. The outcome variable is the binary variable of whether a child works. The right y-axis represents the percent change relative to the baseline group mean of the baseline period. The baseline period is -1. Each event time corresponds to a wave of the TLSA. The shaded area represents the 90% confidence interval. Standard errors are clustered at the individual level. The sample includes daughters aged 25 to 65. The baseline group includes those who never have LTC needs and those who had LTC needs later. The samples are reweighed by the propensity score estimated by their age in the estimation.

Figure 1.1 suggests a significant drop in ADL onset, followed by further decreases. The decrease is about 5 percentage points at ADL onset, or 10% in daughters' labor supply. The decrease in the labor supply is more than 20% in the long run.

There is also a modest decrease in the labor supply before LTC needs arose. A smoothly decaying health process might generate this pattern. Even before a parent's health status being categorized as LTC-needing, some children have started to respond to this health decay by adjusting labor market participation. This finding guides a modeling of health process that replicates the early adjustments.

Heterogeneity in Labor Supply Responses. In Figures 1.2 and 1.3, I present patterns for sons and children-in-law, respectively, in addition to labor supply responses among daughters. In comparison to daughters, nearly no response from sons is evident, and children-in-law show a large decrease in the labor supply, although the estimates are less precise. This

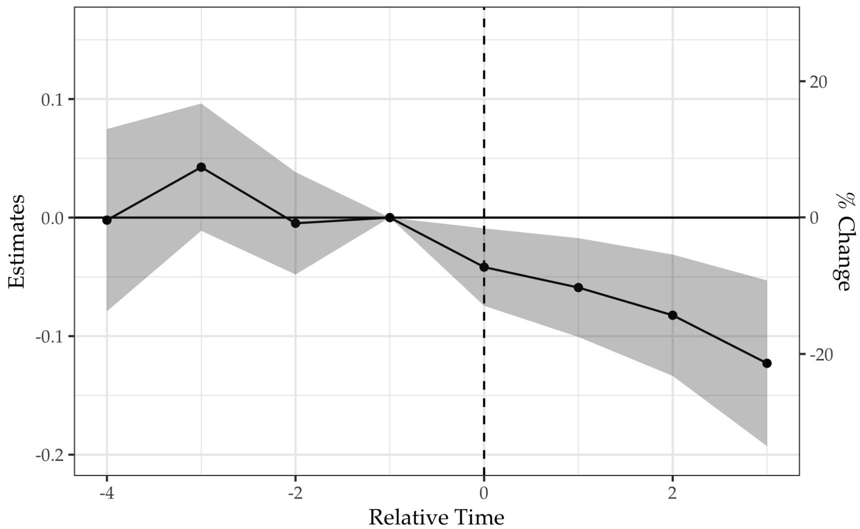


Figure 1.3: Labor Supply Responses for Children-In-Law When LTC Needs Arise

Notes: The event is when parents-in-law first report any ADL. The outcome variable is the binary variable of whether a child works. The right y-axis represents the percent change relative to the baseline group mean of the baseline period. The baseline period is -1. Each event time corresponds to a wave of the TLSA. The shaded area represents the 90% confidence interval. Standard errors are clustered at the individual level. The sample includes daughters aged 25 to 65. The baseline group includes those who never have LTC needs and those who had LTC needs later. The samples are reweighed by the propensity score estimated by their age in the estimation.

heterogeneity in responses suggests the importance of analyzing LTC-related behavior separately for each relationship with care-receivers.

I also present heterogeneous effects by other characteristics. As specified in Equation 1.2, Table 1.2 shows these effects interacting with various children's characteristics. Column (1) indicates that those with greater education are 3 percentage points less likely to drop out of the labor market. Column (2) shows that younger children are less likely to drop out of the labor market. These results suggest that those who are older and less educated are more likely to leave the labor market in response to the parent's LTC needs.

Heterogeneous responses suggest important features that a structural model should capture. Heterogeneity in education is consistent with individuals trading off labor market payoffs for provision of care. Those with less education would have earned less in the labor market and hence have greater chances of providing care to parents when they have LTC needs.

Individuals tending to drop out of the labor market later in their careers more than those in their early careers suggests a model with costs of returning to work. For those in their early careers, it is likely that their parents with LTC needs will not survive until they reach retirement age. If they need to pay a cost to return to the labor market, they are less likely to leave the labor market in the first place. Nevertheless, for those late in their career, they simply retire early to provide care and do not expect to return to the labor market, and hence no returning cost is incurred.

In summary, children reduce their labor market participation when parents' LTC needs arise. The average response is 4 percentage point, but the average masks large heterogeneity. Children who are daughters, less educated, and older are more likely to reduce labor supply. Children who obtained more education or are in their early careers still decrease their labor supply in response to their parents' LTC needs, but on a smaller scale in comparison to groups with opposite characteristics.

	$X_i =$	
	(1)	(2)
	High School	Young
LTC Need × Post × $\mathbf{1}\{X_i = x\}$	0.03	0.03
	(0.01)	(0.01)
LTC Need × Post	-0.05	-0.05
	(0.01)	(0.01)
Post	0.00	0.01
	(0.00)	(0.00)
LTC Need	0.00	-0.02
	(0.01)	(0.01)
$\mathbf{1}\{X_i = x\}$	0.17	0.03
	(0.01)	(0.01)
Intercept	0.61	0.69
	(0.00)	(0.01)
N	928,044	928,044
R^2	0.04	0.00

Table 1.2: Labor Supply Responses When LTC Needs Arise

Notes: The outcome variable is the binary variable of whether a child works. Standard errors are reported in parentheses and are clustered at the individual level. The sample includes sons, daughters, and children-in-law aged from 25 to 65. The baseline group includes those who never have LTC needs and those who had LTC needs later. The samples are reweighed by the propensity score estimated by their age in the estimation. "Young" represents children aged 25 to 40.

1.3.5 Death of the LTC-Needing Parents

In the previous section, I report decreases to children's labor supply when parents have LTC needs. The next question is whether a child returns to the labor market after LTC parents' deaths. This exercise is important to recover the true costs of LTC provision if individuals are prevented from returning to the labor market due to costs of returning to work. I analyze the effect of LTC-needing parents' deaths on children's labor supply to examine whether they return to the labor market after the care provision responsibility ends.

Results. Figure 1.4 shows the labor supply effect after a parent's death. Estimates in a table format appear in the Appendix. I restrict the sample to those whose parents have LTC needs. Similar to analyses on the onset of LTC needs, I construct a baseline group to include those whose LTC-needing parents survive throughout the sample, or those who died later.

In the short run, an increase to the labor supply following LTC-needing parents' deaths is evident, but the increase is not persistent. One explanation is the difficulty of finding permanent employment after leaving for LTC responsibilities. The increase in labor supply is also much smaller than labor supply responses when LTC needs arise.

This pattern is again consistent with a cost of returning to the labor market. Without the cost, the increase should be comparable to the decrease when LTC needs arise. However, the much smaller increase in labor supply response suggests the opposite scenario, in which a high cost of returning to work exists.

Alternative Explanation: Bequest. In addition to the costs of returning to work, wealth effects from bequest should accord with this labor supply pattern. If children inherit a large amount of money from LTC-needing parents, even without the costs of returning to work, they would choose not to participate in the labor market.

However, the parents have few assets. 17.92% of parents reported that they own the houses in which they currently live. Other than the house, only 5.58% of parents with LTC needs reported having total assets of more than 500,000 NTD (or 17,500 USD), approximately the same amount of median annual earning in Taiwan. Bequests thus cannot explain the pattern after parents' deaths.

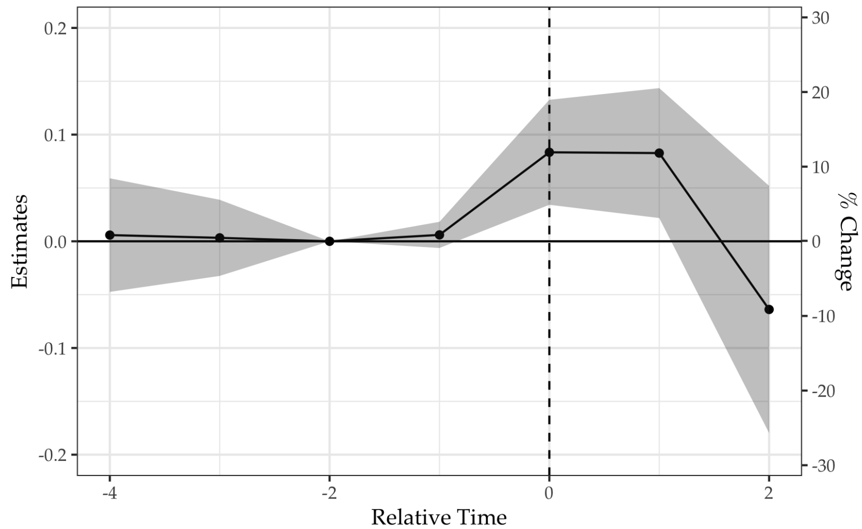


Figure 1.4: Parent Death for Daughter and Work

Notes: The event is when parents with LTC needs pass away. The outcome variable is the binary variable of whether a child works. The right y-axis represents the percent change relative to the control group mean of the baseline period. The baseline period is -2. Each event time corresponds to six months. The shaded area represents the 90% confidence interval. Standard errors are clustered at the individual level. The sample includes daughters aged 25 to 65 whose parents have LTC needs. The baseline group includes those whose parents pass away later. The samples are reweighed by the propensity score estimated by their age in the estimation.

Alternative Explanation: Grandchild. Another explanation for not returning to work is childcare. If the elderly are taking care of their grandchildren, their deaths might simply mean losing a free nanny, but this is not supported by the data. Only 14.23% of the elderly population report that they help take care of their grandchildren. The number is even smaller for LTC-needing elderly people. Therefore, the childcare cannot explain the labor supply patterns either.

1.3.6 Reform in Eligibility for Hiring International Caregivers

Background. The Taiwanese context provides an opportunity to examine how an exogenous change to caregiver hiring prices affects the children's labor supply. I study the effects of a reform to the eligibility for hiring international caregivers on children's labor supply.

In Taiwan, international caregivers play crucial roles in the LTC system. The number of international caregivers grew from nearly none to 259,660, or more than 1% of Taiwan's population, in 2020. Figure 1.5 shows this trend.[5] The vast difference between the number of the international versus domestic caregivers results from their prices. International caregivers are not subject to the minimum wage law or the Labor Standards Act. Most are 24/7, live-in caregivers, and they are approximately 3 times cheaper than domestic caregivers.

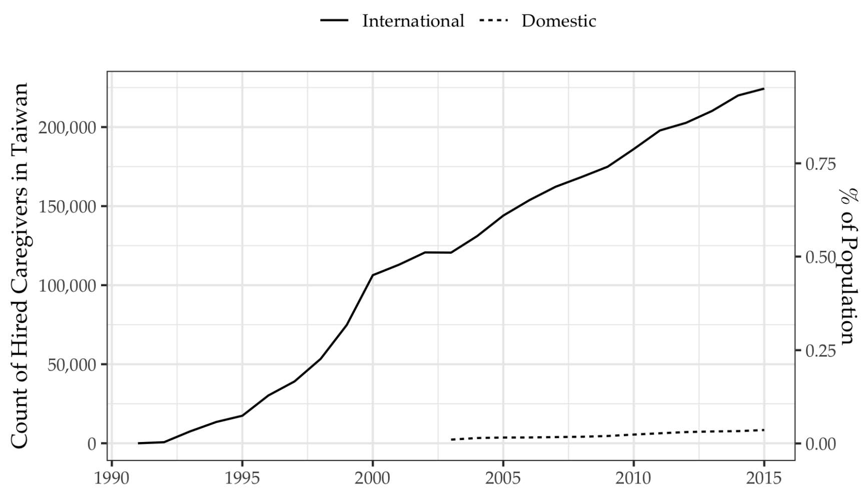


Figure 1.5: Number of International Caregivers in Taiwan

Notes: The left axis represents the number of hired caregivers. The right y-axis represents the number of hired caregivers divided by 2015 total population. Source of data on domestic hired caregivers: [13]. Source of data on international caregivers: [60]

Reform Details. The Taiwanese government heavily regulates hiring and international caregiver. Unlike immigrant workers in the United States, nearly all international caregivers in Taiwan enter the country on a short-term visa and return to their home countries after the end of the contracts. Strict border control also limits the scope of undocumented international caregivers.

5. Taiwan's international caregivers are different from those in the United States, where such caregivers are immigrants who already resided in the country before they were hired. International caregivers in Taiwan mostly work on short-term visas and return to their home countries after the contract ends. The black market is also less of a concern in Taiwan. According to records of the National Immigration Agency, illegal international caregivers who stay in Taiwan comprise approximately 10% of the total stock, much less than the number of undocumented immigrants in the United States.

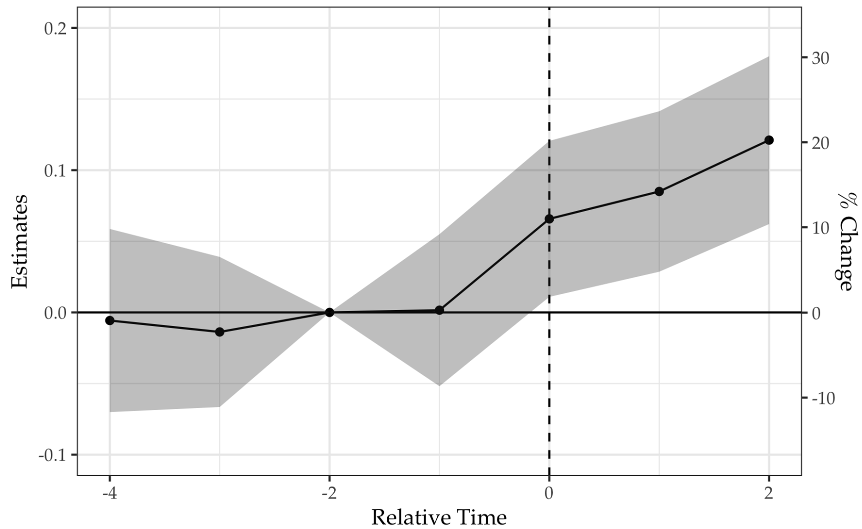


Figure 1.6: Effect of the Reform in Eligibility

Notes: The event is the 2012 reform in the eligibility of hiring. The outcome variable is the binary variable of whether one works. The right y-axis presents the percent change relative to the control group mean of the baseline period. The baseline period is -2. Each event time corresponds to six months. The shaded area represents the 90% confidence interval. The standard errors are clustered at the individual level. The sample includes children aged 25 to 65. The control group consists of those over age 80 and who were already eligible to hire an international caregiver before the reform. The treatment group consists of those over age 80 and who are only eligible to hire an international caregiver after the reform. The samples are re-weighted by the propensity score estimated by their age in the estimation.

Eligibility for hiring a caregiver is a function of the care-receiver's age and health status, but criteria have relaxed over time. The reform I study occurred in September 2012. Before the reform, a care-receiver needed to have severe ADL needs to be eligible to hire an international caregiver. After the reform, the criteria relaxed for the older population. For those over age 80, care-receivers with moderate ADL needs are now eligible to hire international caregivers. This new criterion is much more lenient than the previous.[6]

I restrict my samples to children whose parents are over age 80. The research design remains the DiD design in Equation 1.1. The treatment group is those who became eligible to hire only after the reform, and the control group is those who were already eligible.

6. The official measure of ADL needs is the Barthel index. Severe ADL corresponds to an index lower than 35, and moderate ADL corresponds to an index of lower than 60 but higher than 35. Details appear in the Appendix.

Results. Figure 1.6 shows results. Being eligible to hire international caregivers has a large, positive effect on children's labor supply. In comparison to those who were already eligible, children whose parents are newly eligible have a share of working that is 12 percentage points higher due to the reform. The large effect of the reform again suggests the key trade-off that individuals make between labor market participation and care provision. When hiring a caregiver becomes cheaper, children are less likely to provide care themselves, instead participating in the labor market. No evidence suggests anticipation of the reform, and this unexpected nature is incorporated in the model.

1.3.7 How Descriptive and Reform Evidence Informs Modeling Choices

In the above analysis, I present children's labor supply dynamics in responses to (i) LTC needs of their parents, (ii) the death of LTC-needing parents, and (iii) the reform to eligibility of hiring an international caregiver. There are four main findings. The first is that children's labor supply drops significantly when parents' LTC needs arise. Compared with those having LTC needs later and without LTC needs, the labor supply decreased by 4 percentage points when their parents' LTC needs arose. The effects are persistent and increasingly negative over time. The labor supply started decreasing even before the onset of LTC needs, consistent with a smooth decaying health process.

The second finding is substantial heterogeneity in labor supply responses when LTC needs arose. Daughters are 300% more likely to drop out of the labor market than sons are, and 30% more likely than the children-in-law. Lower-educated children reduce their labor supply more when parents' LTC needs arise, consistent with lower opportunity costs of providing care themselves. Older children also decrease their labor supply more in response to parents' LTC needs. The costs of returning to the labor market might explain this age pattern since younger children expect a longer period before retirement after parents' deaths.

The third finding is that children return to the labor market after their LTC-needing parents pass away, but the probability of returning is 25% smaller in comparison to the drop at the onset of LTC needs. This finding is consistent with costs of returning to work, motivating the choice of a dynamic model that features adjustment costs of entering the labor market. In a static model without such costs, children's labor supply would return to the same level after their LTC-needing parents pass away, as if their parents have never experienced LTC needs.

The fourth finding is that being eligible to hire an international caregiver increases children's labor supply in comparison to those ineligible. Exploiting the reform regarding such eligibility, I find that hiring eligibility increases the labor supply by 6 percentage points immediately. This estimate serves as a key data moment in the model to identify opportunity costs of hiring a caregiver.

1.4 Model and Identification

Motivated by the previous findings in Section 1.3, I build a dynamic labor supply model to understand the policy effects and to conduct counterfactual analyses. Although the estimates above of the labor supply effects are informative for how individuals respond to parents' LTC needs, they do not apply directly to quantifying the effects of typical LTC policies. To assess policy effects, I model key trade-offs individuals face, including consumption, leisure, and parent's health. The model disciplines how individuals value these components under resource constraints. Based on this framework, we can infer behavioral responses and welfare implications of policy experiments by shifting these resource allocations. I present the model and discuss identification in this section.

1.4.1 Individual Problem

I consider an adult child i who maximizes the sum of expected utility in any period t:

$$\max_{D_{it}} V_{it} = \sum_{s=t}^{T} \beta^{s-t} E[u_{is}(C_{is}, L_{is}, H_{is}, D_{is}, D_{is-1})|D_{it}],$$

where u_{it} is the flow utility during period t, β the discount rate, C_{it} consumption, L_{it} leisure, D_{it} choice, and D_{it-1} lagged choice. Individual i has a parent whose health at time t is H_{it}.[7]

During each period, individual i chooses whether to work and hire a caregiver ($D_{it} = 1$), or not work and provide LTC by herself ($D_{it} = 0$).[8] When individual i chooses, she considers both the current period payoff u_{it} and how her choice will affect future payoffs. There is no savings or borrowing in the model. Each period is a year. The model ends at $T = 65$ when individual i retires and the working decision is no longer relevant.

The individual faces following constraints:

$$\begin{aligned}
C_{it} &= D_{it}(W_{it} - P^*_{it}\mathbf{1}\{H_{it} \in \{\text{Any ADL}\}\}), \\
L_{it} &= 1 - aD_{it} - b(H_{it})(1 - D_{it}), \\
P^*_{it} &= \theta_P - \theta_{PE} E_{it}, \\
E_{it} &= E_{it}(H_{it}, X_{H,it}, \text{Reform}_t).
\end{aligned}$$

Consumption C_{it} is earnings minus the expenditure of hiring a caregiver. If a child with an LTC-needing parent chooses to work and hire, her consumption is $W_{it} - P^*_{it}$. If she decides to provide care by herself, then $C_{it} = 0$. Leisure, L_{it}, is the time endowment minus the time needed to spend at work or providing care. a and $b(H_{it})$ are time spent on work and

7. Individual's problem can alternatively be written as

$$\max_{C_{it}, L_{it}, D_{it}} V_{it} = \sum_{s=t}^{T} \beta^{s-t} E[u_{is}(C_{is}, L_{is}, H_{is}, D_{is}, D_{is-1})|C_{it}, L_{it}, D_{it}],$$

and specify individuals' choices as choosing C_{it}, L_{it}, and D_{it}. However, the model structure implies that C_{it} and L_{it} are determined when D_{it} is decided, and thus I write the problem the way above.

8. I discuss alternative specification of individuals' choices in the Appendix.

providing care, respectively, and both are calibrated to data.[9] P^*_{it} denotes the shadow price of hiring a caregiver. Price is a function of eligibility, E_{it}. E_{it} is a function of parent's health H_{it}, age $X_{H,it}$, and the reform in hiring eligibility described in Section 1.3.6.

1.4.2 Preference Specification

An individual cares about her consumption, leisure, and parent's health status. For each individual, i, I specify her flow utility as:

$$u_{it} = \underbrace{\theta_C C_{it}}_{\text{consumption}} + \underbrace{\theta_L L_{it}}_{\text{leisure}} + \underbrace{\sum\nolimits_h \theta_h \mathbf{1}\{H_{it} = h\}}_{\text{parent's health}} - \underbrace{\theta_F D_{it} \mathbf{1}\{D_{it-1} = 0\}}_{\text{cost of returning to work}} + \epsilon_{u,it}(D_{it}).$$

The flow utility is assumed linear. Since savings and borrowing are not part of the model, individuals do not smooth their consumption across time. This is consistent with the linear assumption. I discuss how savings might affect results in the Appendix.

The model corresponds to a unitary household. Parents do not make decisions in the model. There are two reasons for this modeling choice. First, LTC is nearly always expected to be children's responsibility in this context.[10] Second, 37.2% of elderly people with LTC needs have ADL difficulties that resulted from dementia, and thus they are less likely to make economic decisions. Discussions regarding other family members are included in the Appendix.

There is a cost of returning to work in the model, motivated by the previous descriptive results. An individual incurs an adjustment cost, θ_F losses in utility, if she does not work during the previous period and begins working this period. $\epsilon_{u,it}(D_{it})$ denotes the idiosyncratic preference shocks. For example, if an adult child gets sick herself and working

9. a is calibrated to 45 hours per week, the mean hours of a full-time job. b(mild ADL), b(moderate ADL), and b(severe ADL) are calibrated to 100, 87, and 60 hours per week. These numbers are based on data from [28] and [27], respectively.

10. According to [35], more than 80% of people indicated that children are "responsible for taking care of the elderly."

becomes undesirable, she has a small $\epsilon_{u,it}(D_{it} = 1)$ in comparison to $\epsilon_{u,it}(D_{it} = 0)$. Potential experience effects and how they affect results are discussed in the Appendix.

1.4.3 Health Process

A latent parental health index, H^*_{it}, evolves, and health during the next period depends on the choice of care provision, current health, and demographics $X_{H,it}$. Formally, the health process is:

$$H^*_{it+1}(D_{it}) = \begin{cases} \sum_h \gamma_{L,h}(D_{it})\mathbf{1}\{H_{it} = h\} + \gamma_X(D_{it})X_{H,it} + \xi_{H,j(i)}(D_{it}) + \epsilon_{H,it+1}(D_{it}) & \text{if } H_{it} \neq \text{Dead} \\ -\infty & \text{if } H_{it} = \text{Dead}. \end{cases}$$

$X_{H,it}$ includes a parent's gender and age. $\xi_{H,j(i)}(D_{it})$ captures permanent unobserved heterogeneity. The permanent unobserved heterogeneity is type specific, and $j(i)$ denotes individual i's unobserved type. For example, a child of a high ability type will have high $\xi_{H,j(i)}(D_{it})$ in her parent's health process. All parameters are choice specific. $\epsilon_{H,it}(D_{it})$ denotes idiosyncratic health shocks. For example, a serious fall injury is represented by a small $\epsilon_{H,it}(D_{it})$.

To bring the model to the data, I further specify parental health using an ordered dependent variable structure. Observed parental health status H_{it} takes one of five possible values. The best to the worst health conditions are (i) healthy, (ii) mild ADL, (iii) moderate ADL, (iv) severe ADL, and (v) dead. Death is an absorbing state. The three levels of ADL correspond to the cutoff of eligibility for hiring an international caregiver.

$$H_{it} = \begin{cases} \text{Healthy,} & \text{for } m_4 < H_{it}^* \\ \text{Mild ADL,} & \text{for } m_3 < H_{it}^* \le m_4 \\ \text{Moderate ADL,} & \text{for } m_2 < H_{it}^* \le m_3 \\ \text{Severe ADL,} & \text{for } m_1 < H_{it}^* \le m_2 \\ \text{Dead,} & \text{for } H_{it}^* \le m_1. \end{cases}$$

1.4.4 Wage Process

The wage process is a standard AR(1) process with covariates:

$$\log W_{it+1} = \omega_L \log W_{it} + \omega_X X_{W,it} + \xi_{W,j(i)} + \epsilon_{W,it+1}.$$

The next period's wage depends on the current period wage W_{it}, individual demographics $X_{W,it}$, unobserved type $\xi_{W,j(i)}$, and idiosyncratic wage shocks $\epsilon_{W,it}$. $X_{W,it}$ includes age, gender, and education. $\xi_{W,j(i)}$ can be interpreted as labor market skill for type $j(i)$. $\epsilon_{W,it}$ denotes idiosyncratic wage shocks, such as an unexpected promotion.

1.4.5 Timeline and Information Set

Timeline. At the beginning of period t, idiosyncratic preference, health, and wage shocks—$\epsilon_t = (\epsilon_{u,it}(D_{it}), \epsilon_{H,it}(D_{it}), \epsilon_{W,it})$—are realized. Agent i learns her current state variables, $(W_{it}, E_{it}, H_{it}, X_{H,it}, X_{W,it})$. Importantly, she learns the realized wages, eligibility, and parental health for this period. She then forms expectations of future values, $E_t[V_{it+1}]$, where expectation is taken over the distribution of idiosyncratic shocks. She then chooses whether $D_{it} = 1$ or $D_{it} = 0$. The current period ends, and the individual enters the next period.

Information Set. An individual knows her current observable state variables, such as age, gender, parental health status, and wage. She also knows her unobserved type and the value of unobserved permanent heterogeneity, $\xi = (\xi_{H,j(i)}(D), \xi_{W,j(i)})$.

At time t, an individual does not know the exact values of future idiosyncratic shocks, and neither does she foresee any upcoming reforms. However, she knows the health and wage processes, and she also knows the distribution of idiosyncratic shocks, F_ϵ. Therefore, when she makes decisions, she forms correct expectations of future parental health statuses and wages.

1.4.6 Identification of Health and Wage Processes

Identification Challenges. I experience the common identification challenges for the wage process that the literature commonly discusses; I observe wages only when a person works. Furthermore, I allow the health process to evolve differently according to whether a child provides care by herself. Thus, merely regressing observed wages or health statuses on covariates does not recover the underlying processes.

Roy Model Illustration. I use a two-sector Roy model framework and a simplified static model to illustrate the identification challenges and solutions. In this simplified model, there exists labor market ($D_i = 1$) and care provision ($D_i = 0$) sectors. Individuals sort into these sectors according to both observable and unobservable characteristics. Observable characteristics include gender, age, and education, and unobservable characteristics include skills in the labor market, skills with care provision, and access to other care provision support.

To illustrate, I write a static version of the model as:

$$
\begin{aligned}
u_i(D_i = 1) &= W_i + H_i(D_i = 1), \\
u_i(D_i = 0) &= H_i(D_i = 0), \\
W_i &= \omega_X \mathbf{X}_{W,i} + \epsilon_{W,i}, \\
H_i(D_i = 1) &= \gamma_X(D_i = 1)\mathbf{X}_{H,i} + \epsilon_{H,i}(D_i = 1), \\
H_i(D_i = 0) &= \gamma_X(D_i = 0)\mathbf{X}_{H,i} + \epsilon_{H,i}(D_i = 0),
\end{aligned}
$$

The moment conditions that can be derived from the model are:

$$E[W_i|D_i = 1, \mathbf{X}_{W,i}, \mathbf{X}_{H,i}] = \omega_X \mathbf{X}_{W,i} + E[\epsilon_{W,i}|D_i = 1, \mathbf{X}_{W,i}, \mathbf{X}_{H,i}], \tag{1.3}$$

$$E[H_i(D_i = 1)|D_i = 1, \mathbf{X}_{W,i}, \mathbf{X}_{H,i}] = \gamma_X(D_i = 1)\mathbf{X}_{H,i} + E[\epsilon_{H,i}(D_i = 1)|D_i = 1, \mathbf{X}_{W,i}, \mathbf{X}_{H,i}], \tag{1.4}$$

$$E[H_i(D_i = 0)|D_i = 0, \mathbf{X}_{W,i}, \mathbf{X}_{H,i}] = \gamma_X(D_i = 0)\mathbf{X}_{H,i} + E[\epsilon_{H,i}(D_i = 0)|D_i = 0, \mathbf{X}_{W,i}, \mathbf{X}_{H,i}]. \tag{1.5}$$

The choice equation is:

$$D_i = \mathbf{1}\{\omega_X \mathbf{X}_{W,i} + \gamma_X(D_i = 1)\mathbf{X}_{H,i} + \epsilon_{W,i} + \epsilon_{H,i}(D_i = 1) \geq \gamma_X(D_i = 0)\mathbf{X}_{H,i} + \epsilon_{H,i}(D_i = 0)\}.$$

To identify the parameters, the standard Heckman selection procedure applies to this context. Using the choice equation expression, $E[\epsilon_{W,i}|D_i = 1, \mathbf{X}_{W,i}, \mathbf{X}_{H,i}]$, $E[\epsilon_{H,i}(D_i = 1)|D_i = 1, \mathbf{X}_{W,i}, \mathbf{X}_{H,i}]$, and $E[\epsilon_{H,i}(D_i = 0)|D_i = 0, \mathbf{X}_{W,i}, \mathbf{X}_{H,i}]$ can be re-written as inverse Mill ratios. These ratios are functions of $\mathbf{X}_{W,i}$ and $\mathbf{X}_{H,i}$. Controlling for these, the parameters ω_X, $\gamma_X(D_i = 1)$, and $\gamma_X(D_i = 0)$ can be identified.

Excluded Shifters. To avoid relying on identification from parametric assumptions on the ϵ terms, I need shifters of decisions that are excluded from the wage and health equations. For example, in Equation 1.3, variables that shift $E[\epsilon_{W,i}|D_i = 1, \mathbf{X}_{W,i}, \mathbf{X}_{H,i}]$ but do not enter $\mathbf{X}_{W,i}$ are needed. Similarly, in Equation 1.4 and 1.5, shifters that affect decisions but does not enter $\mathbf{X}_{H,i}$ are needed.

I use the parents' health as the shifter for the wage equation, and I use the lagged wage as the shifter for the health equation. The assumption is that lagged parent health does not affect wages directly, and the lagged wage does not affect current parental health directly.

Full Model Implementation. Most of the arguments above go through in the full version of my model, with only two exceptions. First, health status has an ordinal dependent variable structure. However, it is straightforward to accommodate selection correction in this case. Second, the choice equation in the dynamic model has no closed-form solution, but there are semi-parametric approaches that can be used in this context [3]. Estimation details appear in the Appendix.

1.4.7 Identification of the Preference Parameters

I follow the dynamic discrete choice literature for identification of preference parameters in the model. In particular, [56] and [43] discuss general identification results for dynamic discrete choice models with unobserved heterogeneity under assumptions of idiosyncratic preference shocks and finite mixtures. [56] explain why non-parametric identification is impossible in dynamic discrete choices models and that parametric assumptions regarding idiosyncratic preference shocks identify model parameters. [43] show that repeated observations from panel data and information from covariates provide identifying restrictions in finite mixture models.

Linking to results in the literature, I follow the discussion in [2] and list additional formal assumptions for model identification.

Formal Assumptions.

Assumption 1 (IID). *Idiosyncratic preference, health, and wage shocks,* $(\epsilon_{u,it}, \epsilon_{H,it}, \epsilon_{W,it})$, *are independent across individuals, over time, and across one another.*

Assumption 2 (DISTR). *Idiosyncratic preference, health, and wage shocks,* $(\epsilon_{u,it}, \epsilon_{H,it}, \epsilon_{W,it})$, *follow a known distribution.*

Assumption 3 (DISCOUNT). *The discount rate,* β, *is known.*

Assumption 4 (NTYPE). *The number of unobserved types is known and small.*

Assumption (IID) rules out time-varying unobserved types in the model. For example, suppose children have heterogeneous rates when accumulating care provision experience. This generates varying $\epsilon_{H,it}$, correlates over time and violates the assumption. However, the assumption does not rule out persistent shocks in the health or wage process, since there are lagged values included in both processes. Note that the model allows for permanent unobserved types, and Assumption (IID) does not rule out the possibility of constant unobserved labor market and care provision skills.

Assumption (DISCOUNT) and Assumption (DISTR) are common in the literature. I assume the discount factor to be 0.95 per year, as commonly assumed.[11] For computational simplicity, I assume $\epsilon_{u,it}$ follows the type one extreme value distribution, and that $\epsilon_{H,it}$ and $\epsilon_{W,it}$ follow the normal distribution. Assumption (NTYPE) requires the number of unobserved types to be known and small. I assume two unobserved types.

Other standard assumptions that are discussed in the literature directly follow from the model's structure. For example, the model satisfies the additive separability assumption, since the idiosyncratic preference shock, $\epsilon_{u,it}$, is additively separable from the observable

11. [1] discuss recent progress on discount factors and dynamic discrete choice models.

components in the flow utility. Conditional independence is also satisfied, given the specification of the wage and the health processes. Discussed in [43], the number of absorbing versus non-absorbing state variables limits the number of unobserved types in identification. In the current model, only death is an absorbing state. Because all other state variables are non-absorbing, identification conditions are satisfied.

1.5 Estimation

1.5.1 Estimation Procedure

Overview

I adapt [5] to estimate the model. I first predict the type for each individual according to the prior distribution.[12] Given the type, I estimate the selection corrected health and wage processes. I then estimate the full model by simulated method of moments, where the moments are conditional on types. I describe the moments targeted in Section 1.5.2. With these estimated parameters, I update the posterior probability of belonging to a specific type. With the posterior distribution of types, I predict the type for each individual according to the posterior distribution. I then iterate the procedure until the parameters estimated converge.

I assume two unobserved types of children. As summarized in [2], permanent unobserved heterogeneity poses an issue for the initial value. I take the standard solution by allowing the probability of being a type to correlate with the initial distribution of the state variables in the model.

12. For the initial guess of the type distribution, I use K-means clustering on individual mean labor supply and the mean wage.

Type Updates

I describe the procedure of unobserved type estimation in depth. Recall that in the model, $j(i)$ denotes individual i's unobserved type. Let $\pi^{(m)}(j|X_{i1})$ be the probability of being type j conditional on the initial state variable vector X_{i1} at the m-th iteration of the estimation procedure. I predict each individual's type using $\pi^{(m)}(j|X_{i1})$. Conditional on the predicted types, I estimate the health and wage processes as if types were observed. With the processes estimated, I estimate the model parameters with simulated method of moments, where moments are conditional on type j.

Let $\theta^{(m)}$ denote the obtained estimates in the m-th iteration. After obtaining $\theta^{(m)}$ I update the type distribution according to:

$$q^{(m+1)}(j|D_i, X_i) = \frac{\pi^{(m)}(j|X_{i1}) \prod_{t=1}^{T} \mathcal{L}_t[D_{it}, X_{it+1}|X_{it}, j; \theta^{(m)}]}{\sum_{j'} \pi^{(m)}(j'|X_{i1}) \prod_{t=1}^{T} \mathcal{L}_t[D_{it}, X_{it+1}|X_{it}, j'; \theta^{(m)}]},$$

where $\mathcal{L}$ denotes the likelihood function. The updated probability, or the posterior probability, given values of initial state variables is then:

$$\pi^{(m+1)}(j|X_1) = \frac{\sum_i q^{(m+1)}(j|D_i, X_i)\mathbf{1}\{X_{i1} = X_1\}}{\mathbf{1}\{\sum_i X_{i1} = X_1\}}.$$

1.5.2 Results

I begin with the health process results and then results for preference estimates. I report the preference estimates by first discussing target moments and model fit. Since the parameters estimated are themselves difficult to interpret, I report model fit and key economic quantities implied by the model, such as labor supply elasticities and the reservation wages. A table of estimated parameters appears in the Appendix.

Estimates of Health Process

I report the estimated health process by plotting the probability of a health decay or death in Figure 1.7.

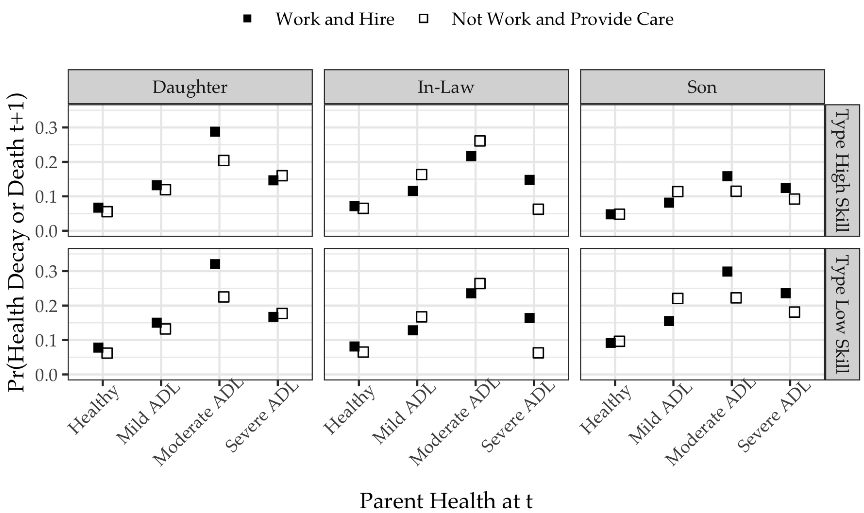


Figure 1.7: Estimates of the Health Process

Notes: The figure plots the probability of parents' health decaying or death, conditional on current health, choice, and demographic groups. Types correspond to various $j(i)$ in the model. Health processes are estimated through selection correction described in Section 1.4.6. The probability of health decay or death is then estimated by simulating data from the estimated health process and calculating the empirical probability.

Figure Setup Explanation. The x-axis represents current parental health status, and the y-axis plots the probability of health decay. The black points show the patterns when a child works and hire a caregiver, and the white points show the patterns when a child does not work and provides care by herself. I plot estimates for various relationships and unobserved types separately.

For example, the first black point in the upper left panel shows that conditional on the parent being healthy and the high type daughter working this period, the probability that the parent has ADL needs or dies during the next period is approximately 0.07. The third white point in the upper left shows that for a high type daughter who does not work and provides care herself for her moderate ADL parent during this period, the probability of her parent having severe ADL or dying during the next period is approximately 0.2.

Probability of Health Decay. I now compare across current health status. For healthy parents without ADL needs, the probability of health decay is consistently 0.07 across all demographic groups. Once a parent has a mild ADL need, the probability of health decay doubles. The probabilities peak at moderate ADL and then drop at severe ADL, since for that condition, the only worse case is death.

The black versus white points denote the patterns of working and hiring versus not working and providing care, respectively. In most cases with moderate and severe ADL, working and hiring leads to worse parental health than not working and providing care. It is reassuring that little difference exists in the estimated probability of health decaying when a parent is healthy. Since no care is needed when a parent is healthy, the probability of a health decay should be similar across the child's choice.

I now focus on demographic patterns. Children-in-law have a different pattern than sons and daughters have. The care provision method shows little difference for mild and moderate ADL, and care provision leads to a very small probability of health decay. The unobserved type also shows a different pattern. In particular, for low-skilled type people, the difference is larger between working and hiring versus not working and providing care.

Model Fit

I now present model fit of targeted moments. When estimating preference parameters, I choose three sets of target moments. These target moments include (i) the share of individuals working conditional on education, (ii) the share of individuals working conditional on parental health status, and (iii) the share of individuals working conditional on lagged work choices and effects of hiring eligibility. I discuss the choice of these target moments and their fits.

Fit of Education. The first set of target moments is the share of children working conditional on education. This set of moments is important since education is a key determinant of a child's wage. A higher-educated individual has greater potential wages, and thus, she is more likely to participate in the labor market and hire a caregiver. Intuitively, this variation provides information on the trade-offs between consumption and care provision.

The model closely replicates data in the share of children working conditional on education. This is shown in Figure 1.8, where white points represent model simulations and black points represent data. In both the model and the data, the share of working children increases as educational attainment increases.

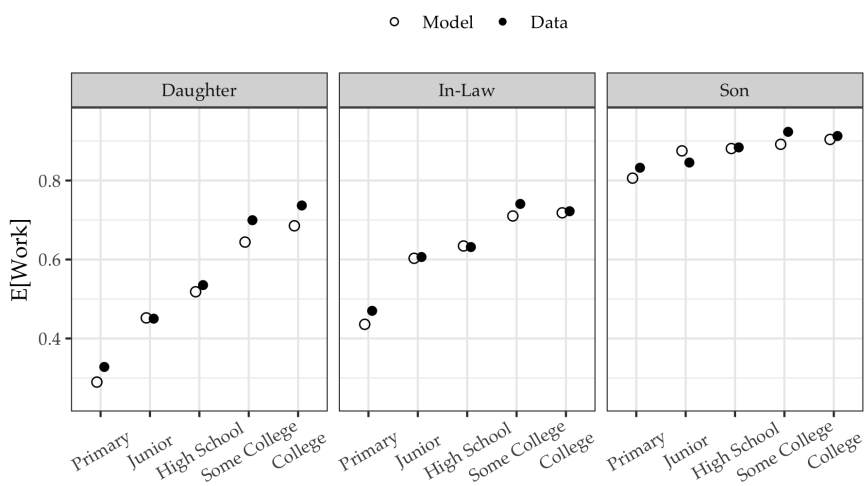


Figure 1.8: Fit of Moment: $E[\text{Work}|\text{Education}]$

Notes: White points represent the data; black points represent the model simulation.

Fit of Health. The second set of target moments is the share of children working conditional on parental health status. Since the model assesses various policy counterfactuals regarding LTC needs, it is essential for it to replicate work decisions conditional on various parental health statuses.

The model captures the relative share of working children conditional on parental health in the data, as shown in Figure 1.9. Consistent with the data, the model predicts that when parents have LTC needs, the share of children who are working is smaller.

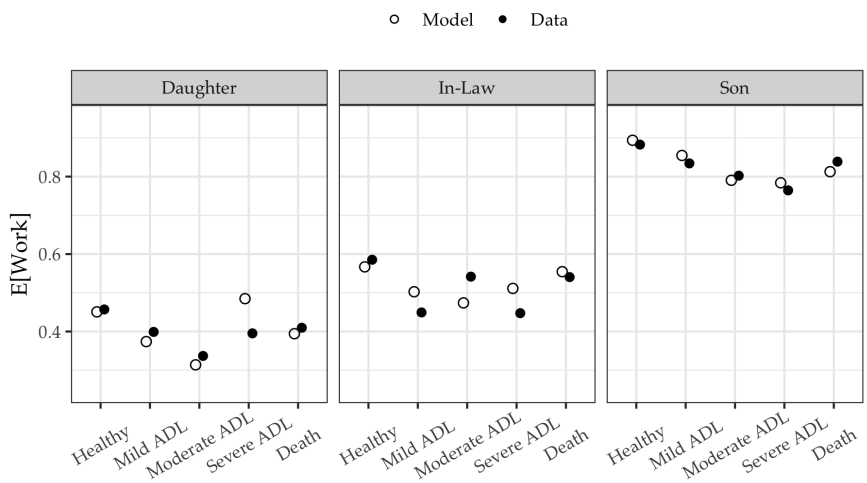


Figure 1.9: Fit of Moment: $E[\text{Work}|\text{Health}]$

Notes: White points represent the data; black points represent the model simulation.

Fit of Persistence and Eligibility. The third set includes moments that capture persistence in the model and effects of eligibility of hiring international caregivers. The share of people working conditional on lagged choices is important to fitting the dynamics of working, and informative for adjustment of the cost parameter, θ_F, in the model. I also target the DiD estimate of the eligibility effect. This moment is important to estimating shadow prices of hiring caregivers and the counterfactual effects of other eligibility criteria.

These moments fit the data reasonably well, as shown in Table 1.3. The model replicates closely the probability of working conditional on working during the last period. The model under-predicts the probability of working conditional on not working last period. However, the predicted probabilities conditional on not working are still much smaller than the predicted probabilities conditional on working. The model also captures effects of being eligible for hiring international caregivers.

Moment	Daughter Data	Daughter Model	Children-In-Law Data	Children-In-Law Model	Son Data	Son Model
$E[\text{Work}_{it}\|\text{Work}_{it-1}]$	0.855	0.918	0.885	0.925	0.954	0.949
$E[\text{Work}_{it}\|\text{Not Work}_{it-1}]$	0.084	0.058	0.091	0.018	0.146	0.123
DiD Eligibility Effect	0.119	0.056	0.119	0.127	0.119	0.080

Table 1.3: Fit of Other Targeted Moments

Notes: $E[\text{Work}_{it}|\text{Work}_{it-1}]$ and $E[\text{Work}_{it}|\text{Not Work}_{it-1}]$ are estimated by empirical probabilities. "DID Eligibility Effect" in the data corresponds to the estimates in Section 1.3.6. The corresponding moment in the model is calculated using the same criteria of eligibility reforms.

1.5.3 Model Validation

Untargeted Moments: 2015 Reform in Eligibility

In addition to the reform in 2012 that I use to estimate the model, there is another reform to eligibility of hiring international caregivers implemented in August 2015. The 2012 reform granted those over age 80 with moderate ADL the permission to hire international caregivers. After the 2015 reform, those over age 85 with mild ADL are also eligible.

This reform provides an opportunity to test the model's performance in predicting policy effects. I assemble a new and independent sample, linking the new 2015 wave of TLSA with NHIRD from 2015 to 2018.[13] This sample is not used elsewhere in the current study.

I use the same DiD design as in Section 1.3.6. For this reform, treatment group is those over age 85 and are eligible after the reform, and control group is those over age 85 and are eligible even before the reform. Results are shown in Table 1.4, with a graphical illustration in the Appendix. The difference in difference estimates suggest a 0.019 increase in children's labor market participation.

The estimated effects are considerably smaller than the 2012 reform. The first reason is that health criteria are different. In the 2012 reform, those whose parents have moderate ADL are benefited, while in the 2015 reform, parents with mild ADL are benefited. For the less

13. The sample is constructed in the same way as the main sample, and the detail appears in the Appendix.

	Reform 2015
Treatment × Post	0.019
	(0.043)
Treatment	0.003
	(0.090)
Post	−0.059
	(0.028)
Intercept	0.528
	(0.059)

Table 1.4: Effects of Reform in 2015

Notes: The outcome variable is the binary variable of whether a child works. Standard errors are in the parentheses and are clustered at the individual level. The sample includes children aged 25 to 65.

severe ADL condition, the substitution between working care providing is smaller. Another reason is the age effect. Children whose parents are over age 85 are older than those whose parents' ages over 80. Baseline labor market participation is smaller for children affected by the 2015 reform.

I simulate the model for the same reform, and compare labor market participation under two different policy. The effect I estimated from my model suggest an average effect of 0.025, and this coincides the estimates from the 2015 reform in the data. Although the estimates from 2015 is less precise due to a smaller number of parents over age 85, the results suggest that the model replicate the out-of-sample effects of reform closely.

Untargeted Moments: Age Profile of Labor Supply

In addition to the reform estimates, I also assess the age profiles of labor supply, which I do not target explicitly. By comparing the data moment with the model prediction, I provide an additional validity check of the estimated model.

Figure 1.10 shows the comparison between the data and model prediction of the life-cycle profile. The predictions fit well, especially for the earlier pattern. For sons and children-in-law, the model over-predicts the share of working individuals near retirement. One explanation for over-predictions is that I do not model savings, pensions, and retirement benefits. When interpreting results from this model, caution is warranted regarding behaviors near retirement. The overall pattern is, nonetheless, close as a set of untargeted moments.

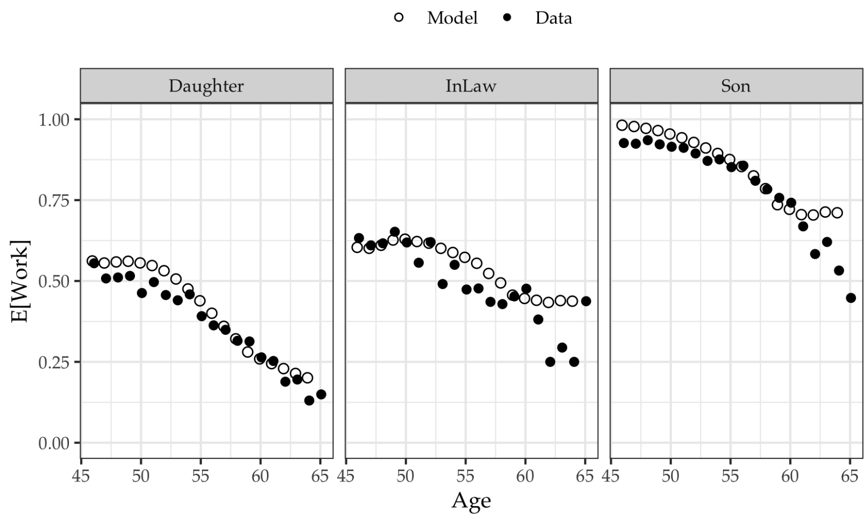


Figure 1.10: Un-targeted Moment: $E[\text{Work}|\text{Age}]$

Notes: The x-axis represents the age of the children. White points represent the data; black points represent the model simulation.

1.6 Economic Mechanism

I describe three sets of results from the model—(i) labor supply elasticities, (ii) reservation wages, and (iii) LTC responsibility and returning to work. These results are useful for understanding the mechanism of the model. They are also building blocks for policy counterfactual analyses in Section 1.7.

1.6.1 Labor Supply Elasticities

Labor supply elasticities from the model are useful in two ways. First, since an extensive literature studies the wage elasticity of labor supply, the elasticity allows us to compare current estimates with those in the literature. Second, many LTC policies are tax reductions or cash subsidies, and thus, labor supply elasticity informs labor supply responses when given these subsidies. For example, if the labor supply elasticity is high when parents have LTC needs, a small wage increase induces individuals to switch from care provision to labor market participation. However, if the labor supply elasticity is low when parents have LTC needs, labor supply responses to wages are small. In this case, policymakers might be less concerned about LTC policies' distortion effects in the labor market.

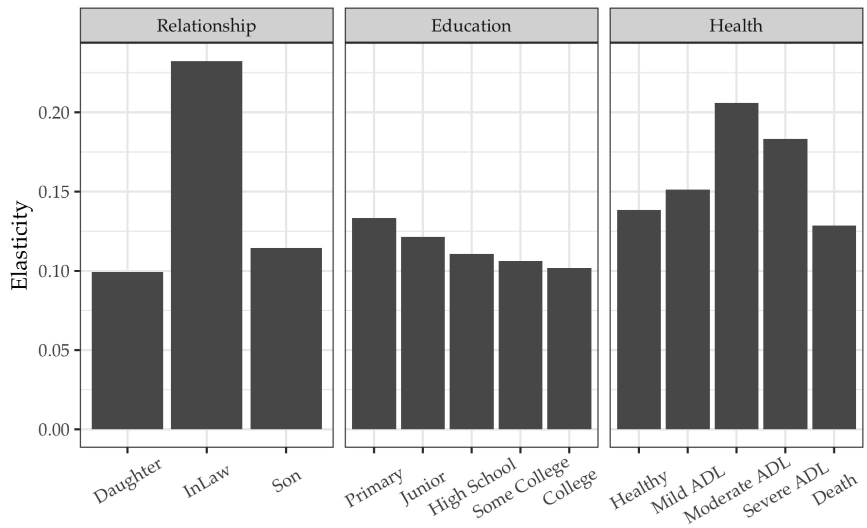


Figure 1.11: Labor Supply Elasticities

Notes: Elasticities are calculated using simulated data.

Results. I follow [24] and calculate extensive margin wage elasticities of labor supply. Results are reported in Figure 1.11. Both daughters and sons have a labor supply elasticity of approximately 0.1, but children-in-law are twice as elastic, likely because they are secondary earners in families and are thus more sensitive to wage changes. A slight downward slope in education exists. The higher educated people have low elasticities. As for heterogeneity in parental health status, an inverted V shape is found, consistent with the level effect—fewer people work when parents have moderate ADL.

These elasticities are comparable to the findings in the Taiwanese literature. [18] find that female labor supply elasticities lie between 0.026 and 0.158. The labor supply elasticity for males is similar to that for females.[14]

1.6.2 Reservation Wage

Reservation wages inform of the wages needed to participate in the labor market. In the model, reservation wage is calculated as the wage needed such that working and hiring a caregiver is indifferent from not working and providing care. The detailed definition is in the Appendix.

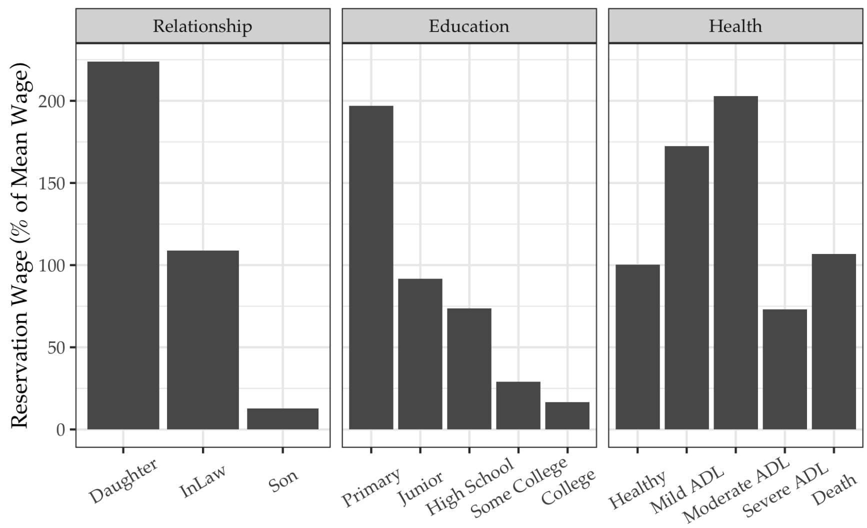


Figure 1.12: Mean Reservation Wage

Notes: The reservation wage is the minimum wage a person requires to make work $D_{it} = 1$ and not work $D_{it} = 0$ indifferent in the model. I normalize the reservation wages reported by the mean wage.

Results. I report mean reservation wages in Figure 1.12. Reservation wages track closely the share of individuals who work. The higher the reservation wages, the smaller the share of individuals who work. Highest to lowest are daughters, children-in-law, and sons. The mean reservation wage is monotonically decreasing in education. For parents with severe ADL, reservation wages are low. This is consistent with the fact that all parents with severe

14. In the U.S. literature, male labor supply elasticities at the participation margin are approximately 0.2. Less consensus has been achieved regarding female labor supply elasticity, but it is generally estimated to be larger than that for males.

ADL are eligible to hire an international caregiver, and hence many children choose to hire one and do not provide care themselves.

Reservation Wage and Policy Effects Illustration. The distribution of reservation wages is also useful for understanding the model. Figure 1.13 and 1.14 show daughters' reservation wage distribution conditional on lagged working statuses, normalized by the mean annual wage. If each person gets the mean wage when participating in the labor market, the area below the curve and left of the vertical line will represent the share of people who are working. The state variables in the model determine where a person is in the distribution. For example, a child worked last period is much more likely to work this period than who did not, and thus the distribution in Figure 1.13 has much smaller reservation wages than the distribution in Figure 1.14

This illustration is also useful for understanding policy effects. In the figures, solid curves represent the distribution of the status quo. In contrast, dashed curves represent the distribution under a tax deduction policy that allows working children with LTC-needing parents to deduct income taxes. The policy shifts the reservation wage distribution to the left, pushing more daughters to participate in the labor market. I provide more details on policy counterfactuals in the next section.

1.6.3 LTC Responsibility and Returning to Work

I assess how many people leave the labor market and do not return due to LTC provision. I consider two scenarios across three periods. In scenario (i), parents are healthy during period one, have moderate ADL needs during period two, and die during period three. In scenario (ii), parents are healthy during period one, die during period two, and die during period three.[15] I then calculate the difference in their labor supply during period three between the

15. Since health processes are endogenous, the counterfactual analysis is conducted through changing potential health outcomes for both choices.

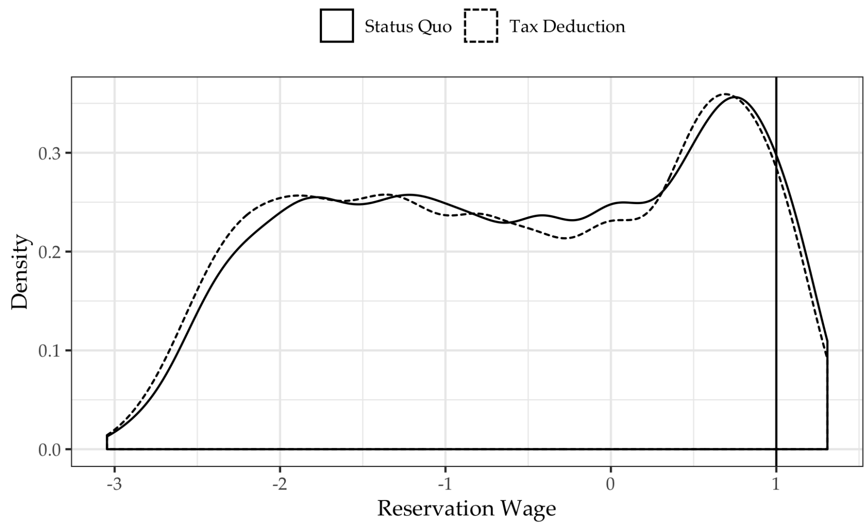


Figure 1.13: Reservation Wage Distribution for Daughters (Worked Last Period)
Notes: The x-axis represents reservation wages, normalized by the mean wage. The black vertical line indicates the mean wage.

scenarios. The difference suggests how much more a person will work, were it not for LTC responsibilities.

Results. I report results in Table 1.5. Column (1) shows the result of counterfactual analysis. The reduction in labor supply is significant for daughters, but also non-trivial for sons and children-in-law. If a daughter experiences parental LTC needs, she participates in the labor market nearly 20% less than in the LTC-free scenario. The difference in labor supply is decreasing in education since the higher educated people do not leave the labor market in the first place. Column (2) shows that these patterns result from the adjustment costs in the model. If I remove adjustment costs θ_F, nearly no difference is evident between the two counterfactual sequences.

Next, I present results using the distribution of parental health sequences in the data. In Column (3) to Column (6), I examine the evolution of parental health. As for the comparison sequence, I again construct counterfactual scenarios in which parents pass away immediately instead of incurring LTC needs. Column (3) shows that during the period in which a parent

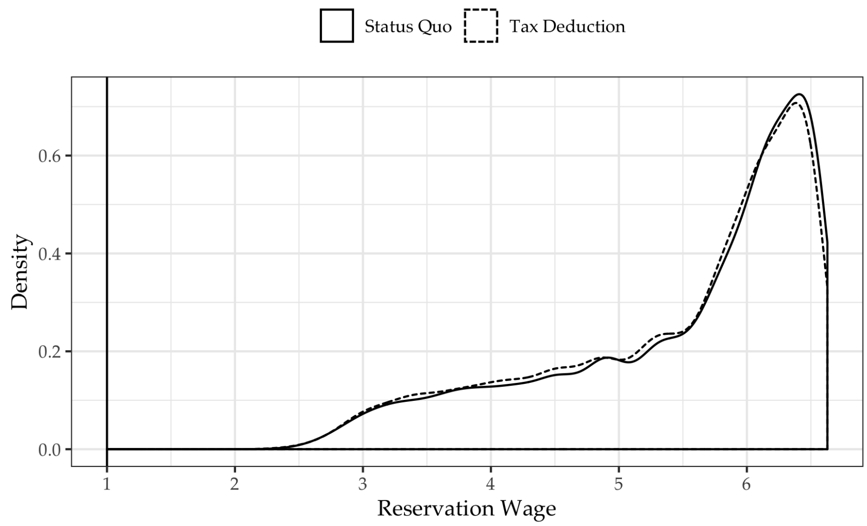


Figure 1.14: Reservation Wage Distribution for Daughters (Not Worked Last Period)
Notes: The x-axis represents reservation wages, normalized by the mean wage. The black vertical line indicates the mean wage.

passes away, daughters whose parents had LTC needs have a 9.3% less probability of working than those whose parents never had LTC needs. Although results are less extreme compared to Column (1), similar heterogeneity is evident for this case. By ways of comparison, this magnitude is similar to fertility effects on female labor supply in the Taiwanese literature. For example, [29] finds a 10% decrease in mother's probability of working when having a third child in Taiwan.

Instead of investigating the period subsequent to parents' deaths, Column (5) shows the long-run effects. Overall, the effects are smaller in the long run, since preference and wage shocks dilute effects from previous LTC needs. However, I still find a 4% smaller labor supply for the daughters in the long run, suggesting how profoundly parents' LTC needs affect careers.

Visualization. Figure 1.15 reports results of the counterfactual analysis. I plot the average decrease found in the short run and in the long run, with the x-axis representing the duration that children experience LTC responsibilities. Consistent with the pattern in Table 1.5, the

Name	Sequence: Healthy, ADL, Dead (1) Baseline	(2) $\theta_F = 0$	Sequence: Aggregate Sequence in Data, Short-Run (3) Baseline	(4) $\theta_F = 0$	Long-Run (5) Baseline	(6) $\theta_F = 0$
Daughter	-19.4	-0.7	-9.3	-0.8	-4.0	-0.1
Children-In-Law	-4.1	-1.1	-1.7	0.3	1.6	-0.6
Son	-5.2	0.1	-6.8	0.6	-2.8	0.6
Primary	-18.1	-1.4	-10.2	-1.3	-5.4	-0.7
Junior	-6.3	-0.3	-7.3	0.5	-3.3	0.5
High School	-6.8	-0.1	-6.8	0.8	-3.1	0.2
Some College	-3.2	-0.7	-6.3	0.8	-4.2	-0.1
College	-3.5	0.4	-5.4	-0.3	-1.9	0.9

Table 1.5: Difference in Labor Supply After Parent's Death

Notes: In "Sequence: Healthy, ADL, Dead," I compare two sequences of parental health outcomes: (i) healthy, moderate ADL, dead, and (ii) healthy, dead, dead. In "Sequence: Aggregate Sequence in Data," I average the differences between pairs of parental health sequences. Each pair of sequence includes (i) healthy at $t = 0$, s periods of ADLs starting from time $t = 1$, and then dead at $t = s+1$, and (ii) healthy at $t = 0$, and then dead at $t > 0$. "Short-Run" reports the comparison of labor supply at time $t = s + 1$, and "Long-Run" reports the average of difference for $t \geq s + 1$. "$\theta_F = 0$" corresponds to results from simulations with $\theta_F = 0$ in the model.

differences are smaller in the long run than in the short run.

The difference is also increasing in the duration of LTC needs. If LTC needs last for only a year, the difference is about 5% in the short run. However, if LTC needs last for 5 years, the difference is approximately 13%. This difference results from expectations the children have regarding their parents' health. Consider a case in which an old parent experiences a severe fall and her health status changes suddenly from healthy to severe ADL. Her son expects that she might pass away in a short time, and thus he likely stays in the labor market and hires a caregiver to avoid the costs of returning to work.

Compensating Variation. I also calculate CV between these two scenarios. The detailed definition is in the Appendix. I find that daughters, sons, and children-in-law demand 11.3%, 3.4%, and 4.1% of mean annual wage to move from the immediate death scenario to the LTC-needing scenario, respectively. Since I compare periods after parents' deaths, the source of

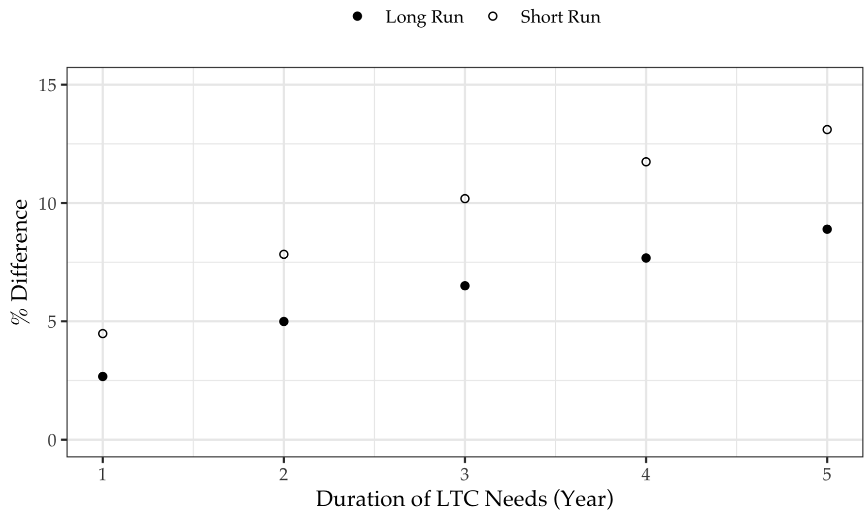


Figure 1.15: Difference in Labor Supply After Parents' Deaths and LTC Duration
Notes: The x-axis represents the duration that parents have LTC needs before death. The y-axis represents how much lower children's labor supply would be after parents' deaths, comparing cases with and without parents' LTC needs. A detailed construction appears in Section 1.6.3.

these CVs is the difference of the cost of returning to work. Many children stopped working to provide care, and they would need to pay a cost to return to the labor market. Therefore, I observe a positive CV to switch to the immediate death scenario.

1.7 Policy Counterfactuals

I now analyze three sets of common LTC policies—(i) reforms to eligibility for hiring international caregivers, (ii) LTC tax deductions, and (iii) in-kind transfers for care-receivers. These policies were all implemented in Taiwan recently, and my goal here is to understand their impacts. I first describe the background, controversy, and debates for each policy. I then revisit how parents' LTC needs affect the long-term labor supply paths of children and assess various policies' implications for this exercise. I also analyze policies' overall labor supply responses under current LTC needs, and I calculate the compensating variation (CV) for each policy and estimate their fiscal costs.

1.7.1 Policy Backgrounds

Reforms on International Caregivers Hiring Eligibility

Foreign-born caregivers constitute an essential part of LTC workers. The core of the policy debate is whether international caregivers serve as a stable source of LTC workers and whether potential competition with domestic professional caregivers hurts local workers.[16] Hiring international caregivers is regulated strictly. Those who want to hire an international caregiver must meet eligibility criteria and apply through the Ministry of Labor.[17] Eligibility criterion is a function of a care-receiver's age, ADL, and disability status. The criterion has relaxed over time, allowing more people to hire international caregivers. Whether the criterion should be further relaxed and who would benefit from such reform remain important topics during policy discussions.

In this section, I simulate two counterfactual reforms. In "Relaxed Eligibility," I analyze a reform that allows everyone with moderate ADL to be eligible, and in "Limited Eligibility," I allow only those with severe ADL to be eligible, even for those over age 80 who were also eligible with only moderate ADL previously. More radical reforms that grant eligibility to everyone or forbid international caregivers are analyzed in the Appendix.

Tax Deductions

Tax deductions and credits for LTC are also common worldwide, including in Belgium, Canada, France, Germany, and Ireland. The Taiwanese government initiated an LTC tax deduction program in 2020, providing a means-tested tax deduction for those with a family

16. These are active debates in Parliament. One parliament member stated, *"Japan is now importing international labor ... Our source countries are similar ... How do we compete with Japan in this market?"* The Director of Workforce Development Agency said, *"We still want them to be complementary...We don't want them to affect domestic labor. Most importantly, we want to build our own LTC system since only that would be a stable source of LTC."* [50]

17. The Ministry of Labor issues visa application permissions to those applying for an international caregiver, and then permission is taken to recruitment agencies to hire an international caregiver.

member who has LTC needs. Each year, depending on the tax bracket, a person can deduct approximately $200 to $500 (or 25% to 61% of minimum monthly wage). This tax deduction is estimated to benefit 0.3 million people, with a tax revenue loss of 2 billion NT dollars, or 0.1% of total tax revenue [51]. As a tax deduction, this policy benefits only those with a job and income. Work incentives and the distribution of benefits are primary topics during discussions of the policy.[18]

The tax deductions I analyze in this section are the same as what the Taiwanese government implemented in 2020. Children who work and hire caregivers could benefit from tax deductions subject to means tests.

In-Kind Transfers

In-kind transfers are common among developed countries, including Canada, Japan, and Portugal [20]. In Taiwan, the in-kind transfer program is part of the LTC 2.0 program, launched in 2017. The program is a means-tested policy that provides in-kind subsidies with broad availability for those with LTC needs.[19] People with disabilities, over age 50 with dementia, or anyone over age 65 with LTC needs are eligible for this program. Importantly, those in nursing homes or who hire caregivers can claim only a minimal amount of transfer. Whether to provide transfers and how to distribute them are central to policy debates.

The in-kind transfers I analyze are similar to the one implemented in 2017. Children who provide care themselves are eligible for in-kind transfers. The transfers for severe, moderate, and mild ADL are 90, 50, and 25 hours of care per month, respectively. I assume that an hour of care provided by a child is equivalent to an hour of care from in-kind transfer.

18. Family members of people with LTC needs are eligible for the deduction. The debate on this policy lies in its scale and how applicable it is. In the form of tax deduction, "those who stay at home and provide care without income will not benefit." [52]

19. Items subsidized include (i) personal and professional care, (ii) transportation to hospitals, (iii) assisted devices purchases and rentals, and (iv) respite care for family caregivers. See [42] for an introduction to the program.

1.7.2 LTC Responsibility and Returning to Work

Specification	(1) Status Quo	(2) Relaxed Eligibility	(3) Limited Eligibility	(4) Tax Deduction	(5) In-Kind Transfer
Daughter	-9.3	-4.7	-9.2	-6.5	-12.0
Children-In-Law	-1.7	7.7	-2.2	-0.4	-5.7
Son	-6.8	-3.7	-6.8	-3.1	-10.6
Primary	-10.2	-4.6	-9.5	-6.5	-13.9
Junior	-7.3	-2.9	-6.9	-3.4	-10.4
High School	-6.8	-3.1	-7.6	-3.9	-10.5
Some College	-6.3	-3.8	-6.8	-3.7	-9.2
College	-5.4	-3.1	-6.0	-2.6	-8.9

Table 1.6: Difference in Labor Supply After Parents' Deaths Under Various Policies

Notes: This table reports short-run returning to work comparisons using the data health sequence under various policies. The details are the same as in Table 1.5. In particular, Column (1) replicates Column (3) in Table 1.5.

Results Under Different Policies. The comparisons among parental health sequences in Section 1.6.3 have vastly different results if different LTC policies are implemented. Table 1.6 reports the comparison in the short run. Column (2) shows that, when eligibility criterion is relaxed, the differences in labor supply after parents' deaths are smaller, resulting from a cheap source of caregivers. A tax reduction also reduces the tendency in which one leaves and returns to the labor market, as shown in Column (4).

Column (3) and Column (5) show that both limiting the international caregivers hiring eligibility and providing in-kind transfers increase labor-market leaving. An individual must provide care herself to be eligible for in-kind transfers, so the program discourages working and hiring.

1.7.3 Labor Supply Responses

I analyze labor supply responses under these policies in comparison to the status quo for children whose parents have LTC needs. By examining how responses differ as a function of observable characteristics, this analysis also identifies the marginal people affected by the various policies.

Reforms on International Caregivers Hiring Eligibility. Labor supply responses to this policy are shown in Column (1) and (2) in Table 1.7. Labor supply responses to an relaxed eligibility are large. When the eligibility is relaxed, sons' labor supply increases by 3.9%, on average. For children-in-law, the number is even higher. Results also vary vastly by education. Higher educated people are less responsive to these policies because they are likely to participate in the labor market under any parental health condition. In contrast, lower-educated children are at the margin. Opposite and almost equally large effects are found when eligibility is limited. In the Appendix, I show that completely open or closed eligibility leads to massive labor supply responses, suggesting that given the current situation, a reform that completely opens or closes the international caregiver market has enormous influences.

Tax Deductions. I report labor supply responses to tax deductions in Column (3) of Table 1.7. A tax deduction has positive effects on the labor supply. However, responses are much smaller in comparison to eligibility reforms. For those whose parents have LTC needs, the labor supply response to this policy is, on average, less than 5%. Sons have larger responses in comparison to daughters and children-in-law, and no clear pattern is evident for education.

In-Kind Transfer. Labor supply responses to in-kind transfers are shown in Column (4) of Table 1.7. In-kind transfers generate negative labor supply responses since they benefit only children who provide care themselves. Negative responses are again larger for lower-educated people. There is also considerable variation in parental health status. Since the program provides many more hours of care services for parents with more severe ADL, responses are larger.

Summary. In addition to the number of international caregivers, analyses above demonstrate large labor supply responses when eligibility criterion is changed. This suggests that international caregivers have already been an essential part of LTC. A second finding is that lower-educated individuals lie at the margin and are responsive to such policies. Elasticity estimates corroborate this finding, but this analysis suggests that the pattern is prevalent under various policies.

Characteristics	(1) Relaxed Eligibility	(2) Limited Eligibility	(3) Tax Deduction	(4) In-Kind Transfer
Daughter	3.6	-8.5	2.9	-4.0
Children-In-Law	7.7	-18.0	2.5	-4.3
Son	3.9	-6.8	5.2	-3.6
Primary	7.8	-10.6	7.1	-4.2
Junior	1.9	-7.1	2.9	-3.3
High School	6.2	-9.2	2.8	-3.5
Some College	6.1	-7.5	4.8	-4.0
College	-0.4	-8.6	3.4	-2.7
Mild ADL			3.4	-2.7
Moderate ADL	4.7	-9.0	6.4	-4.8
Severe ADL			5.6	-5.3

Table 1.7: Labor Supply Responses

Notes: The unit is percent change to the probability of working in comparison to the status quo. The labor supply responses reported are conditional on parents having LTC needs.

1.7.4 Compensating Variation

Another important aspect is how much people value the policies. I simulate the scenarios that implement the policies above, focusing on those whose parents have LTC needs at the first period of the simulated data. I report the average compensation that individuals require during the first period if I remove the policy. I thus report the CV for each policy. I normalize the CVs to the mean annual earnings to ease comparison.

The CV decomposes into two parts for each subgroup. The first is the share of individuals affected by the policy in the subgroup, and the second is how an affected individual value the policy. Taking the tax deduction as an example, the CV for those who benefit from the policy is similar. However, since more sons are working, they have higher overall CV for the policy. Table 1.8 reports total CV, the share of affected individuals, and the CV for affected individuals. Each policy is shown in its own column.

Reforms on International Caregivers Hiring Eligibility. The table shows that daughters and children-in-law value eligibility for hiring more than sons do. Although more sons are working, overall CVs are higher for daughters and children-in-law. The CV of affected individuals is not monotonic in children's education. Two forces operate in the opposite direction. Eligibility to hire an international caregiver benefits higher-educated people more by preventing them from sacrificing higher wages, but eligibility benefits lower-educated people more since they are likely to leave the labor market and return in the future.

Tax Deductions. The scale of CVs is small for a tax deduction policy. Since tax deductions are monetary transfers, CVs are similar across individuals. Small discrepancies are caused by the means-testing design and the tax bracket to which an individual belongs. Variance in the CVs results almost entirely from the share of people affected. For example, since more sons are working, they benefit most from tax deductions.

In-Kind Transfer. Unlike with tax deductions, individuals value in-kind transfers differently. Daughters and children-in-law value the policy more than sons do, and since few sons are benefiting from in-kind transfers, the contrast is more prevalent.

In addition to the heterogeneity of the relationship with care-receivers, in-kind transfers also benefit lower-educated people and those with severe ADL needs more. Total CV of in-kind transfers is monotonically decreasing in education. Since more severe ADL gets more hours of in-kind transfers, total CV increases with LTC needs. Groups that benefit more from in-kind transfers link with economically disadvantaged groups, and this redistributive property could represent the government's argument for this LTC policy.

Name	Relaxed Eligibility			Limited Eligibility			Tax Deduction			In-Kind Transfer		
	Total CV	Affected Share	Affected CV	Total CV	Affected Share	Affected CV	Total CV	Affected Share	Affected CV	Total CV	Affected Share	Affected CV
Daughter	0.017	0.038	0.457	-0.011	0.024	-0.470	0.038	0.437	0.087	0.051	0.534	0.096
Children-In-Law	0.057	0.045	1.282	-0.095	0.068	-1.399	0.037	0.442	0.084	0.075	0.527	0.143
Son	0.010	0.067	0.145	-0.011	0.080	-0.142	0.095	0.824	0.115	0.002	0.047	0.048
Primary.	0.014	0.044	0.324	-0.017	0.035	-0.471	0.048	0.514	0.094	0.049	0.486	0.101
Junior	0.033	0.059	0.560	-0.034	0.053	-0.640	0.081	0.738	0.109	0.029	0.262	0.109
High School	0.020	0.060	0.336	-0.019	0.065	-0.301	0.081	0.745	0.109	0.021	0.207	0.102
Some College	0.008	0.052	0.147	-0.013	0.066	-0.204	0.070	0.629	0.112	0.009	0.106	0.087
College	0.012	0.047	0.261	-0.030	0.105	-0.286	0.044	0.410	0.107	0.008	0.086	0.093
Mild ADL							0.066	0.617	0.107	0.031	0.332	0.094
Moderate ADL	0.133	0.360	0.368	-0.153	0.377	-0.405	0.059	0.587	0.100	0.034	0.292	0.117
Severe ADL							0.054	0.591	0.092	0.038	0.281	0.135

Table 1.8: Compensating Variation

Notes: "Total CV" and "Affected CV" are normalized by mean annual wage. For example, a daughter's total CV for relaxed eligibility, 0.017, means that she requires 1.7% of the mean annual wage to accept removal of this policy. "Affected Share" represents the share of those affected by the policy among children whose parents have ADL needs.

1.7.5 Fiscal Costs

Description of Comparison Exercise. Although eligibility reforms include no fiscal costs, tax deductions and in-kind transfers are costly for the government to implement. In Table 1.9, I compare both policies' fiscal costs when they are implemented. The calculation does not include administrative costs, and I assume full take-up for both policies.

Policy	Costs			Benefits	
	(1) Per Beneficiary Spending ($USD/Year)	(2) Normalized Total Spending	(3) Tax Revenue from Behavior Changes	(4) Beneficiary Mean CV	(5) % Benefited Among ADL
(a) In-Kind Transfer	149.71	0.240	-0.00067	0.079	0.324
(b) LTC Tax Deduction	330.13	1.000	0.00059	0.099	0.612
(a)/(b)	0.453	0.240	-1.136	0.899	0.529

Table 1.9: Costs and Benefits

Notes: "Normalized Total Spending" sets the total spending on LTC tax deduction to 1.

Results. Column (1) shows average spending on those benefited. Spending on tax deductions is more than twice as large as on in-kind transfers per beneficiary. Since more people are working and hiring, total spending on in-kind transfers is only a quarter of that on tax deductions, as Column (2) shows.

One might argue that tax deduction could incentivize work, and hence the real cost will be smaller due to increased tax revenue. However, additional tax revenues from labor supply responses is only 0.00059 times total spending on the policy, as shown in Column (3). Similarly, additional tax losses from discouraging work by an in-kind transfer policy are negligible. The CV generated by the in-kind transfer policy is about 89.9% of the one generated by tax deductions for an average beneficiary. The cost of the in-kind transfer program per beneficiary is only 45.3% of the tax deduction policy, suggesting that in-kind transfer represent the more cost-effective program.

1.8 Conclusion

I assess the children's labor supply responses to elderly's LTC needs, analyzing the effects of LTC policies on such responses. Using data from Taiwan, I first document that children are 4 percentage point less likely to participate in the labor market when parents' LTC needs arise, with daughters, the less educated, and older children having the largest decreases in labor supply. Only a small share of children return to the labor market if their LTC-needing parents pass away.

Motivated by the descriptive findings, I then build and estimate a dynamic labor supply model, combining the descriptive evidence with an exogenous variation in caregivers' prices from a policy reform in Taiwan. The model features costs of returning to work, endogenous health processes, and unobserved heterogeneity in care and labor market skills. Model-based results suggest large costs of returning to work, especially for daughters and the less educated.

By simulating commonly implemented LTC policies, including changing eligibility criteria for hiring international caregivers, LTC tax deduction, and in-kind transfers, I find vastly different labor supply responses to LTC needs and welfare implications. Relaxing or restricting eligibility of hiring international caregivers will have huge impacts on LTC arrangements and children's welfare. Tax deductions and in-kind transfers have different effects, appearing in whether children stay in the labor market when parents experience LTC needs and the set of children benefited from the policies. In particular, tax deductions keep more children in the labor market and mostly benefit sons, while in-kind transfers drive more children out of the labor market permanently and benefit daughters. The different effects largely result from the work incentives these policies provide.

The Taiwanese government recently began expanding community-based LTC institutions, trying to provide more diverse and flexible LTC services that focus on professional and preventive care. If these services could be accessed easily, labor-intensive and low-skilled focused care could change in the future. The potential effects are beyond the scope of this paper, but they would be interesting and important issues for future studies.

1.A Data Construction Details

1.A.1 ADL Measure Construction

I construct the ADL measure using the TSLA based on the eligibility rule for hiring international caregivers. The eligibility rule uses the Barthel Index as a measure of ADLs. The index maps the performance of ten ADL items to a scale between 0 and 100; the lower the index, the more severe the health condition.

The ten items that the Barthel Index considers include grooming, feeding, transfers, toilet use, walking, dressing, climbing stairs, bathing, urinary incontinence, and fecal incontinence. The index considers each of these ten items separately and sums them up. For example, if a person is capable of climbing stairs by herself, she gets 10 points from that item. If she needs supports from someone to climb stairs, she gets 5 points, and if she cannot climb stairs even with support, she gets 0 points.

Severe ADL corresponds to those with a Barthel Index of 0 to 35. Moderate ADL corresponds to 35 to 60, and mild ADL corresponds to 60 to 95. The TLSA includes questions regarding feeding, transfers, toilet use, walking, dressing, climbing stairs, and bathing. On average, the correlation between ADL items is approximately 0.7. Since all ADL items are highly correlated, I assume that individuals have difficulties with grooming, urinary, and fecal incontinence whenever they report any other ADL difficulties. The assumption does not create an issue for descriptive analysis. One concern is the eligibility criteria, and I tend to overstate the severity of an individual's health if bias exists in the measurement. In that case, estimates of the effects of eligibility represent a lower bound, since some of those labeled as treatment groups are in fact control groups.

1.A.2 NHIRD Construction

The NHIRD provides a link to TLSA data. The link is created using parents' national identification number in TLSA data. Since the national identification number is unique for each individual, parents' information can be linked perfectly.

The NHIRD also provides information on the family structure. Due to the design of the National Health Insurance, a person becomes a dependent of one of her family members if she does not have a job. I can thus infer the family relationship from this dependent structure. When I track children's information after parents pass away, I rely on the dependent structure to infer children's information.

Under this structure, one concern may is that the set of children I track is incomplete, and hence estimation of labor supply effects after parents pass away is biased. However, a child identified through a parent is more likely to bear LTC responsibility, and thus, if this set of children leads to biased estimates, I would overestimate returning to work given that these children are more responsive to LTC-related events. This means that the cost of returning to work would play an even more important role than analyses currently suggest.

1.B Additional Descriptive Results

In Section 1.3, I provide descriptive evidence visually, and I present estimations in Table 1.10. These estimates are equivalent to an average of estimates before and after corresponding events.

	ADL Needs	Death	Reform
Affected Group × Post	−0.04	−0.03	0.12
	(0.01)	(0.04)	(0.05)
Affected Group	−0.02	0.10	−0.08
	(0.01)	(0.05)	(0.09)
Post	0.01	−0.00	−0.07
	(0.00)	(0.00)	(0.04)
Intercept	0.70	0.73	0.66
	(0.00)	(0.02)	(0.07)
N	928,044	566,286	4,835
R^2	0.00	0.00	0.00

Table 1.10: Average Effects

Notes: The outcome variable is the binary variable of whether a child works. Standard errors are in the parentheses and are clustered at the individual level. The sample includes sons, daughters, and children-in-law aged 25 to 65. The samples are reweighed by the propensity score estimated by their age in the estimation.

I report results when LTC needs arise, using both baselines in Section 1.3. I report results for the two baselines separately in Figures 1.16 and 1.17.

1.C Alternative Model Specification

1.C.1 Choice Specification

The choice is assumed to be binary in my model. This vastly simplifies the model identification. This simplification rules out the case in which one hires a caregiver but does not work. In the data, approximately 10.8% of the children report that they hire a caregiver but do not work.

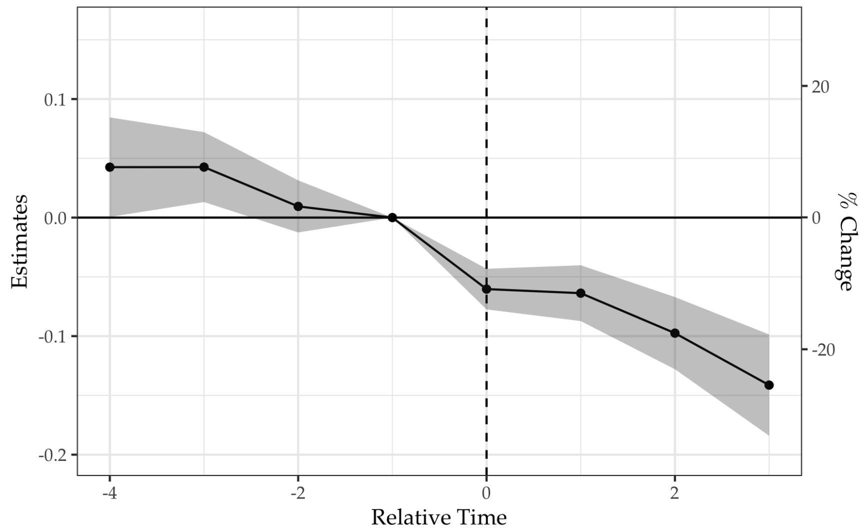


Figure 1.16: Labor Supply Responses for Daughters When LTC Needs Arise (First Baseline)
Notes: The event is when parents first report any ADL in the data. The outcome variable is the binary variable of whether a child works. The right y-axis represents the percent change relative to the baseline group mean of the baseline period. The baseline period is -1. Each event time corresponds to a wave of the TLSA. The shaded area represents the 90% confidence interval. Standard errors are clustered at the individual level. The sample includes daughters aged 25 to 65. The baseline group consists of those whose parents never have LTC needs. The samples are reweighed by the propensity score estimated by their age in the estimation.

There are 32.9% of children who work but do not report hiring a caregiver. However, the average time one needs to take care of their parent for mild ADL is 60 hours per week. It is hard for those people to have a full-time job and take care of their parents simultaneously. Most likely, these children ask relatives or other siblings to provide a certain amount of care. The model is consistent with this possibility. However, the price of hiring a caregiver should be interpreted as a shadow price.

1.C.2 Savings

In the LTC literature, most papers that include savings in their model focus on how parents save to insure against future ADL shocks. (For example, [61] and [6].) In the current study, the focus is on children's decisions. In addition, as discussed in Section 1.3.5, most parents have few assets in the data, and thus parents' savings should be less concerning.

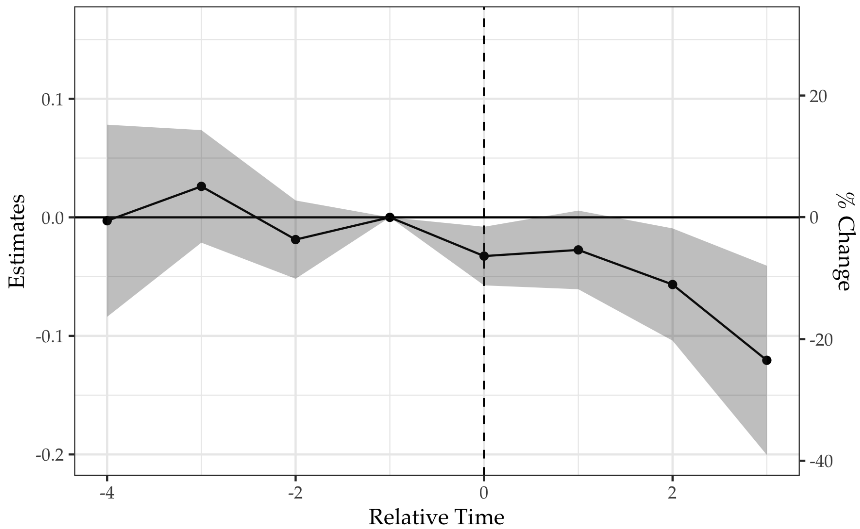


Figure 1.17: Labor Supply Responses for Daughters When LTC Needs Arise (Second Baseline)

Notes: The event is when parents first report any ADL in the data. The outcome variable is the binary variable of whether a child works. The right y-axis represents the percent change relative to the baseline group mean of the baseline period. The baseline period is -1. Each event time corresponds to a wave of the TLSA. The shaded area represents the 90% confidence interval. Standard errors are clustered at the individual level. The sample includes daughters aged 25 to 65. The baseline group consists of those whose parents have LTC needs later. The samples are reweighed by the propensity score estimated by their age in the estimation.

Some papers in the literature that also focus on children from saving decisions as the current study. (For example, [68].) The main reason that I do not include savings in the model is data limitation; there is no good asset information for children in my data. To address this concern, I build a stylized two-period model that includes saving decisions. With this simplified setup, it is possible to infer in which direction would savings shift results.

A Stylized Model with Saving Decision

Consider a two-period for individual i. Individual i's problem is to maximize the lifetime utility:

$$U_i = C_{i0}^{\alpha} L_{i0}^{1-\alpha} + \beta C_{i1}^{\alpha} L_{i1}^{1-\alpha},$$

where C_{i0} denotes i's consumption at period 0, L denotes leisure, and β denotes discount factor. Individual i faces the following constraints:

$$W(D_{i1} + rD_{i0}) = C_{i1} + rC_{i0} + PD_{i1}H_{i1} + rPD_{i0}H_{i0},$$
$$L_{i0} = 1 - aD_{i0} - b(1 - D_{i0})H_{i0},$$
$$L_{i1} = 1 - aD_{i1} - b(1 - D_{i1})H_{i1},$$

where W denotes wage, r denotes interest rate, D denotes individual's work decision, H_{i0} and H_{i1} denote indicator of parents' LTC needs, and P denotes the price of a hiring caregiver. The first constraint links total spending and total earnings in both periods, and the rest constraints specify time usages as in the main model.

I show that when parents have LTC needs, children work less in the world allowing savings than in the world not allowing savings. For simplicity, I assume that individuals know that $H_{i0} = 0$ and $H_{i1} = 1$.

First consider the world with savings. Children must have positive savings, since in the second period their parents need LTC. Due to the curvature in the utility function, children smooth consumption by saving in the first period. Denote the amount an individual saves as S.

Next, consider the world without savings. Since the only decision is whether $D_i = 0$ or $D_i = 1$, I write the consumption under $D_i = 1$ as C_{Work} and the consumption under $D_i = 0$ as $C_{\text{Not Work}}$. The utility comparison an individual makes in the second period is then $C^{\alpha}_{\text{Work}}(1-a)^{1-\alpha}$ versus $C^{\alpha}_{\text{Not Work}}(1-b)^{1-\alpha}$. She will work if and only if $C^{\alpha}_{\text{Work}}(1-a)^{1-\alpha} \geq C^{\alpha}_{\text{Not Work}}(1-b)^{1-\alpha}$. Similarly, for an individual in the world with savings, she will work if and only if $(C_{\text{Work}} + rS)^{\alpha}(1-a)^{1-\alpha} \geq (C_{\text{Not Work}} + rS)^{\alpha}(1-b)^{1-\alpha}$.

With the decision rules above, I discuss the implications of savings in the model. I focus on the decisions in period two. Consider a case where an individual works in a world with savings but does not work in a world without saving. Then the following conditions must be satisfied:

$$(C_{\text{Work}} + rS)^\alpha (1-a)^{1-\alpha} \geq (C_{\text{Not Work}} + rS)^\alpha (1-b)^{1-\alpha},$$
$$(C_{\text{Work}})^\alpha (1-a)^{1-\alpha} \leq (C_{\text{Not Work}})^\alpha (1-b)^{1-\alpha},$$

which implies that

$$\frac{C_{\text{Work}}}{C_{\text{Not Work}}} \leq \frac{C_{\text{Work}} + rS}{C_{\text{Not Work}} + rS},$$

and thus $C_{\text{Work}} \leq C_{\text{Not Work}}$. If a child's wage is higher than the price to hire an caregiver, then this condition is not satisfied. On the other hand, if we consider a case where a child works when not allowed to save but does not work when allowed to, then we have the opposite implication.

In summary, analyses above suggest that by allowing for savings in the model, it is more likely to observe children with low wages to leave the labor market due to parents' LTC needs.

1.C.3 Alternative Household Structure

Motivation

My model has a unitary household. In the main model, there is only a child making all the decisions and a parent whose only role is to be taken care of. In the LTC context, it is reasonable to have the elderly parent not participating in the decision process. However, one may argue that there are potentially other members in the household. The main model is compatible with this setup, since the price of hiring a caregiver is an estimated shadow price, that includes the possibility of hiring another household member.

Nevertheless, it is still interesting to uncover the heterogeneity in household structures and explore its implication for the counterfactual analysis. This might also be useful for interpreting the model. For example, we have seen daughters and children-in-law having the largest responses to parents' LTC needs. An implication is to think of the model as a model for secondary earners.

Setup

The extended model I consider is as follows:

$$u_{it} = \theta_C C_{it} + \theta_L L_{it} + \sum\nolimits_h \theta_h \mathbf{1}\{H_{it} = h\} - \theta_F D_{it}\mathbf{1}\{D_{it-1} = 0\} + \epsilon_{u,it}(D_{it}),$$

subject to the following constraints:

$$\begin{aligned}
C_{it} &= D_{it}(W_{it} - P^*_{it}\mathbf{1}\{H_{it} \in \{\text{Any ADL}\}\}),\\
L_{it} &= 1 - aD_{it} - b(H_{it})(1 - D_{it})(1 - \theta_{LM}M_{it}),\\
P^*_{it} &= \theta_P - \theta_{PE}E_{it} - \theta_{PM}M_{it},\\
E_{it} &= E_{it}(H_{it}, X_{H,it}, \text{Reform}_t).
\end{aligned}$$

The flow utility remains unchanged from the main model in the paper. In the constraints individual faces, M_{it} is an indicator of whether this household has people other than the child and the parent. $M_{it} = 1$ if the child or the parent's spouse is also in the household. $M_{it} = 0$ if neither the child or the parent has a spouse living in the same household.

This additional member enters the model as a potential helping hand. First, suppose one decides to provide LTC by oneself. The amount of time that a child needs to spend on providing LTC depends on whether there are other members. With a helping hand, the amount of time needed to provide care by oneself drops from $b(H_{it})$ to $b(H_{it})(1 - \theta_{LM})$.

Second, the shadow price of hiring a caregiver also changes in this extension. As described in the previous section, the shadow price can also be interpreted as the price of hiring a relative or friend. In this case, the shadow price would be θ_{PM} less since there is an additional household member that one can potentially hire.

Results

I estimate this version of model and highlight the difference between $M_{it} = 1$ versus $M_{it} = 0$. The share of children working conditional on parents' health and whether there is an additional member is shown in Figure 1.18.

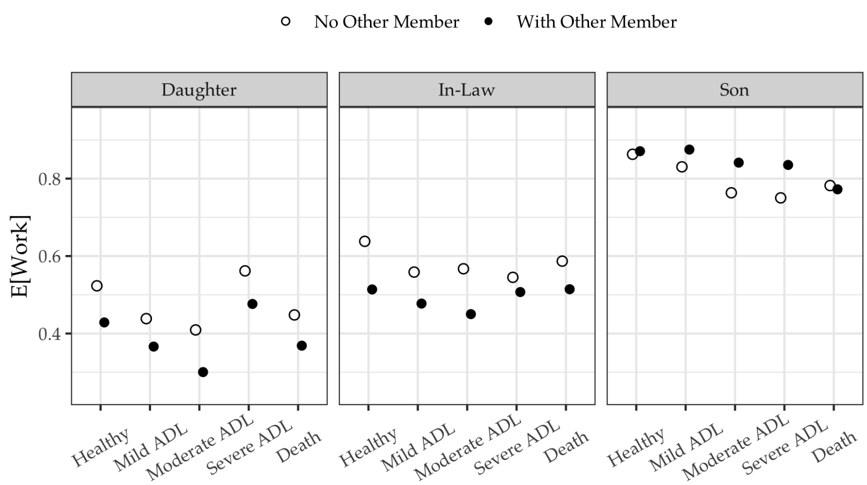


Figure 1.18: Share of Children Working Conditional on Parental Health and Additional Member

Notes: The white points plot the case without other member ($M_{it} = 0$), while the black points plot the case with other member ($M_{it=1}$).

As shown in the figure, the effect of an additional member is mostly a parallel shift in share of individuals working conditional on different parental health statuses. Interesting patterns lie in the different relationships with the care-receivers. For daughters and children-in-law, the presence of an additional member decreases the share of individuals working for any parental health status. In contrast, for sons the effect of additional members goes the opposite direction.

The result is consistent with the interpretation that daughters and the children-in-law are the secondary earners in a household. When there are no other members, they act more alike as primary earners. However, when there are other members, their behavior diverges.

Next, I explore this extension's implication to my counterfactual analysis. Figure 1.19 presents the results. We observe that the most differences are generated from those without other members. In the case with other household members, there are much fewer people who leave the labor market and do not return due to parents' LTC needs. As discussed in the previous section, my main specification presents an average effect. This extension further shows the large burdens and huge effects for those without additional helping hands.

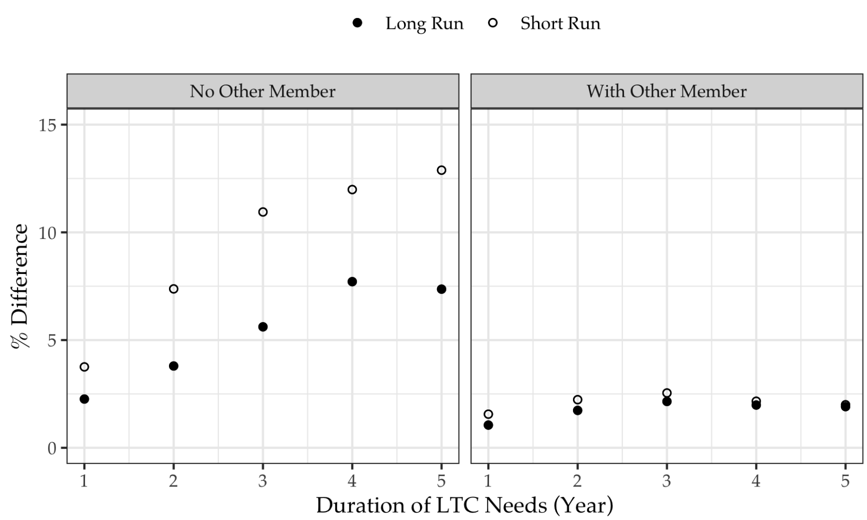


Figure 1.19: Difference in Labor Supply After Parent's Death and LTC Duration

Notes: The x-axis plots the duration that the parents have LTC needs before death. The y-axis plots how much lower the children's labor supply would be after parents' death, comparing the cases with and without parents' LTC needs. "No Other Member" corresponds to those without other member in the household ($M_{it} = 0$), while the "With Other Member" corresponds to those with other member in the household ($M_{it} = 1$). The detailed construction is described in Section 1.6.3.

1.C.4 Experience and Human Capital

Motivation

One possible extension of my model is to add the human capital and experience aspects. The effects of experience in the labor market has been studied in classical papers in labor economics, such as [46].

In the context of my paper, experience in the labor market may play a role in the returning to work decision after experiencing parents' LTC needs. Conditional on the lagged choice, are the experienced more likely to keep participating in the labor market? Or would it be other case that the more experience is less likely to keep supplying labor? One way of the other, the experience effect would affect the pattern of returning to work.

My current main model corresponds to a special case in which experienced or not is binary. The experience is fully depreciated after one stops working for one period. All the possible experience effects and human capital accumulation are loaded into the adjustment cost term in the model. In this extension, I explore how individuals with different experiences may respond to parents' LTC needs.

One major limitation in extending the model to incorporate the experience aspect is data. I do not observe the full work history, nor do the TLSA collect information on children's labor market experiences. The only proxy to the labor market experience is to use the observed work duration in the panel data.

In addition to the data limitation, this extension is going to increase the size of the state space. Currently, the state variable related to experience is whether one worked in the last period. To record the experience in the labor market, the size of the state variable will increase accordingly.

Setup

In response to these limitations, I take a calibration approach and extend my model to explore the possible experience effects. The model I consider is as follows:

$$u_{it} = \theta_C C_{it} + \theta_L L_{it} + \sum\nolimits_h \theta_h \mathbf{1}\{H_{it} = h\} - \theta_F D_{it} \mathbf{1}\{D_{it-1} = 0\} + \underbrace{\theta_{Exp} D_{it} Exp_{it}}_{\text{experience effect}} + \epsilon_{u,it}(D_{it}),$$

where Exp_{it} denotes individual i's labor market experience at time t. This labor market experience follows a deterministic accumulation process:

$$Exp_{it+1} = Exp_{it} + D_{it} - (1 - D_{it}).$$

The flow utility is similar to the main specification. However, there is an additional experience term $\theta_{Exp} D_{it} Exp_{it}$ entering the flow utility. One interpretation of θ_{Exp} is simply the wage return to labor market experience. However, by allowing the experience term to enter directly in the utility function, I allow for a more general return to experience, such as job amenity or flexible work arrangements.

The experience process is simple. By working for an additional period, one's experience increases by one. If one does not work this period, then her experience depreciates by one. A possible further extension is to allow for asymmetry in accumulating and depreciating experience stocks, but I stick to the above specification for simplicity.

Results

I take the parameter estimates from the main model, and then calibrate θ_{Exp} to the expectation of work conditional on the experience constructed from the observed duration of the data, Exp_{it}.

By calibrating the model to daughters' results, I find the Exp_{it} to be 0.140. First, the positive sign suggests that the more experience one has in the labor market, the more like she chooses to work. This is true even conditional on whether she worked in the last period. Second, the magnitude is large. According to the consumption parameter, θ_C, an additional year of experience translates into approximately 10% of wage increase.

To further understand the implication of experience effects to my model, I conduct the main counterfactual analysis in Section 1.6.3. The results are shown in Figure 1.20. For the less experienced, they have much smaller probability of being in the labor market compared with the scenario without parental LTC needs. On the other hand, for the more experienced, the differences are much smaller. The different patterns for these two groups of individuals result from the high return of labor market experiences.

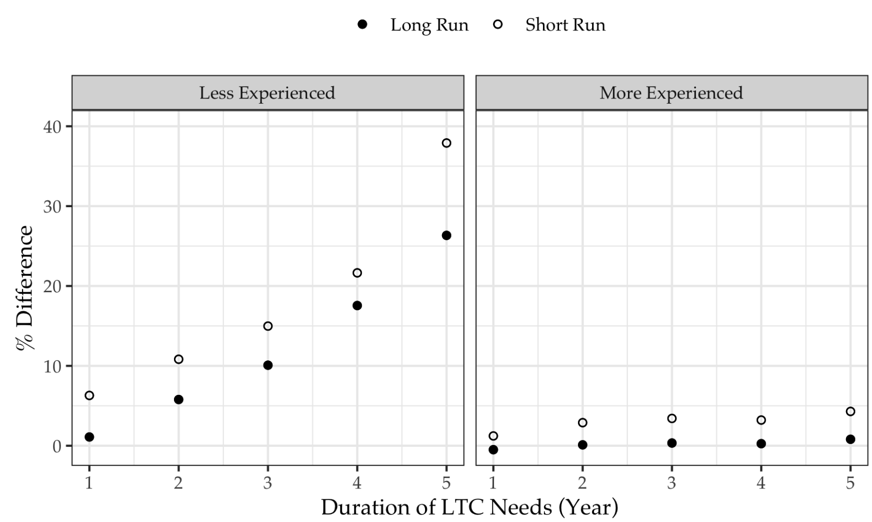


Figure 1.20: Difference in Labor Supply After Parent's Death and LTC Duration

Notes: The x-axis plots the duration that the parents have LTC needs before death. The y-axis plots how much lower the children's labor supply would be after parents' death, comparing the cases with and without parents' LTC needs. "Less Experienced" corresponds to those with experience level smaller than the mean experience level, while the "More Experienced" corresponds to those with experience level larger than the mean experience level. The detailed construction is described in Section 1.6.3.

While the main results in my paper present the average pattern for all experience levels, this exercise informs us of more potential heterogeneity. If the policymakers aim at preventing the permanent leave of labor market due to LTC needs, one aspect they could consider is to target those with less experience.

1.D Estimation Details

1.D.1 Parameters Estimated

Estimates for preference parameters in the model appear in Table 1.11. Standard errors in parentheses are calculated following [54] and [36].

	Daughter	Son	Children-In-Law
θ_C	1.21	4.18	1.78
	(0.02)	(0.04)	(0.09)
θ_L	2.00	4.54	4.15
	(0.06)	(0.09)	(0.10)
$\theta_{h=\text{Death}}$	-1.27	16.04	4.14
	(0.09)	(0.15)	(0.08)
$\theta_{h=\text{Severe ADL}}$	3.09	12.84	0.71
	(0.04)	(0.06)	(0.09)
$\theta_{h=\text{Moderate ADL}}$	10.71	10.85	5.81
	(0.12)	(0.13)	(0.11)
$\theta_{h=\text{Mild ADL}}$	7.79	14.12	16.52
	(0.08)	(0.19)	(0.10)
θ_P	1.70	2.52	3.56
	(0.11)	(0.12)	(0.05)
θ_{PE}	1.66	1.39	6.01
	(0.18)	(0.23)	(0.11)
Intercept	-3.84	-9.47	-5.55
	(0.09)	(0.04)	(0.13)
θ_F	25.65	18.47	24.57
	(0.04)	(0.31)	(0.11)

Table 1.11: Preference Parameter Estimates

Notes: Standard errors appear in parentheses. The definition and calculation follows [54] and [36].

1.D.2 Graphical Illustration of 2015 Reform

The point estimates and the magnitude from the reform are close to the prediction from model, although the reform effects of 2015 are less precise compared to the 2012 one due to a smaller affected population and the health requirement of the reform. Figure **??** shows the estimates from the data.

1.D.3 Reservation Wages

In the model, reservation wage is calculated as the wage needed such that working and hiring a caregiver is indifferent from not working and providing care. To formally define reservation

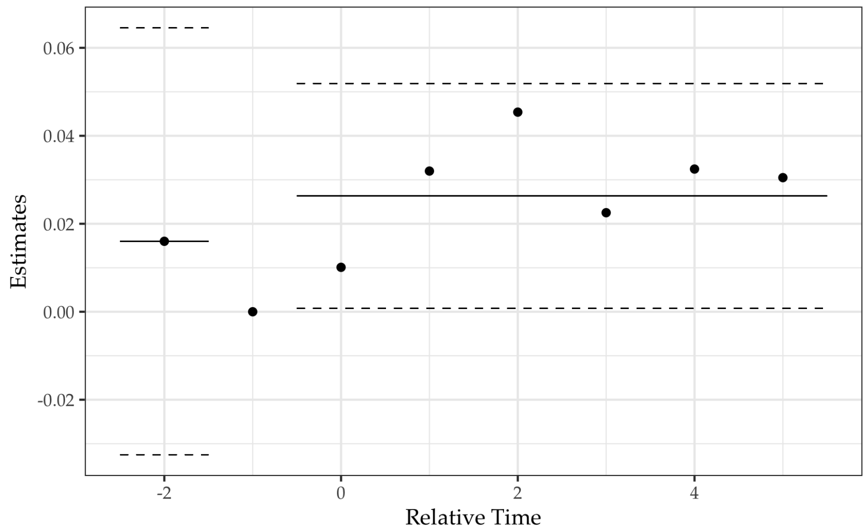


Figure 1.21: Effect of the Reform in Eligibility

Notes: The event is the 2015 reform in the eligibility of hiring. The outcome variable is the binary variable of whether one works. The baseline period is -1. Each event time corresponds to six months. Solid lines represent mean of estimates before and after the reform. Dashed lines represents the 90% confidence interval. The standard errors are clustered at the individual level. The sample includes children aged 25 to 65. The control group consists of those over age 85 and who were already eligible to hire an international caregiver before the reform. The treatment group consists of those over age 85 and who are only eligible to hire an international caregiver after the reform.

wages, recall the individual problem:

$$\max_{D_{it}} V_{it} = \sum_{s=t}^{T} \beta^{s-t} E[u_{is}(C_{is}, L_{is}, H_{is}, D_{is}, D_{is-1})|D_{it}],$$

and the expression can be also written as follows:

$$\max_{D_{it}} u_{it}(D_{it}) + V_{it+1}(D_{it}),$$

and if I expand u_{it} and replace C_{it} with budget constraints, we have:

$$\max_{D_{it}} \theta_C D_{it}(W_{it} - P^*_{it}\mathbf{1}\{H_{it} \in \{\text{Any ADL}\}\}) + \theta_L L_{it} +$$
$$\sum\nolimits_h \theta_h \mathbf{1}\{H_{it} = h\} + \theta_F D_{it}\mathbf{1}\{D_{it-1} = 0\} + \epsilon_{u,it}(D_{it}) + V_{it+1}(D_{it}).$$

Consider the case when $D_{it} = 1$ and $D_{it} = 0$ separately. If $D_{it} = 1$, then individual's value is:

$$\theta_C(W_{it} - P^*_{it}\mathbf{1}\{H_{it} \in \{\text{Any ADL}\}\}) + \theta_L(1-a) + \\ \sum\nolimits_h \theta_h \mathbf{1}\{H_{it} = h\} + \theta_F \mathbf{1}\{D_{it-1} = 0\} + \epsilon_{u,it}(D_{it} = 1) + V_{it+1}(D_{it} = 1).$$

If $D_{it} = 1$, then individual's value is:

$$\theta_L(1 - b(H_{it})) + \sum\nolimits_h \theta_h \mathbf{1}\{H_{it} = h\} + \epsilon_{u,it}(D_{it} = 0) + V_{it+1}(D_{it} = 0).$$

Reservation wage is defined as the wage such that an individual is indifferent between $D_{it} = 1$ and $D_{it} = 0$. That is, reservation RW_{it} is defined as the RW_{it} that satisfies the following:

$$\theta_C(RW_{it} - P^*_{it}\mathbf{1}\{H_{it} \in \{\text{Any ADL}\}\}) + \theta_L(1-a) + \\ \sum\nolimits_h \theta_h \mathbf{1}\{H_{it} = h\} + \theta_F \mathbf{1}\{D_{it-1} = 0\} + \epsilon_{u,it}(D_{it} = 1) + V_{it+1}(D_{it} = 1) \\ = \theta_L(1 - b(H_{it})) + \sum\nolimits_h \theta_h \mathbf{1}\{H_{it} = h\} + \epsilon_{u,it}(D_{it} = 0) + V_{it+1}(D_{it} = 0).$$

The reservation wage defined does not involve future wages, and hence it does not affect values of V_{it+1} in the above equation. The only place RW_{it} term shows up is in the very first part of the equation. As a result, reservation wage RW_{it} is well-defined.

1.D.4 Compensating Variation

CV for a policy is defined as the compensation needed for an individual to reach her initial utility after I remove the policy. Formally, consider the following expression of an individual's problem with certain policy at the first period:

$$\tilde{v}_{i1} = \max_{D_{i1}} \tilde{u}_{i1}(D_{i1}) + \tilde{V}_{i2}(D_{i1}),$$

where $\tilde{v}_{i1}$ is the optimized value, and I use tilde to represent flow utility and values under the policy. The counterpart value where no policy is in effect is:

$$v_{i1} = \max_{D_{i1}} u_{i1}(D_{i1}) + V_{i2}(D_{i1}).$$

Given the linear flow utility specification in the model CV is simply:

$$CV_i = \frac{\tilde{v}_{i1} - v_{i1}}{\theta_C},$$

where θ_C is in the denominator because that translate utility into monetary unit.

1.E Additional Policy Counterfactuals

In this section, I consider the counterfactual analysis which (i) allows everyone with LTC needs to hire an international caregiver (open eligibility), and (ii) forbids anyone to hire an international caregiver (no eligibility). These extreme eligibility rules might induce general equilibrium effects. In the analyses I abstract from the potential general equilibrium effects and show results for differences in labor supply after parents' deaths, labor supply responses, and compensating variation.

As shown in tables below, completely open or closed eligibility leads to massive labor supply responses, suggesting that given the current situation, a reform that completely opens or closes the international caregiver market has enormous influences.

Specification	(1) Status Quo	(2) Open Eligibility	(3) No Eligibility	(4) Relaxed Eligibility	(5) Limited Eligibility
Daughter	-9.3	10.2	-20.1	-4.7	-9.2
Children-In-Law	-1.7	24.6	-16.5	7.7	-2.2
Son	-6.8	-2.0	-12.0	-3.7	-6.8
Primary	-10.2	6.5	-19.3	-4.6	-9.5
Junior	-7.3	3.5	-13.6	-2.9	-6.9
High School	-6.8	3.0	-14.7	-3.1	-7.6
Some College	-6.3	0.1	-12.1	-3.8	-6.8
College	-5.4	1.8	-12.2	-3.1	-6.0

Table 1.12: Difference in Labor Supply After Parents' Deaths Under Various Policies

Notes: This table reports short-run returning to work comparisons using the data health sequence under various policies. The details are the same as in Table 1.5. In particular, Column (1) replicates Column (3) in Table 1.5.

Characteristics	(1) Open Eligibility	(2) No Eligibility	(3) Relaxed Eligibility	(4) Limited Eligibility
Daughter	44.1	-20.4	3.6	-8.5
Children-In-Law	50.3	-44.9	7.7	-18.0
Son	7.7	-12.3	3.9	-6.8
Primary	31.8	-24.9	7.8	-10.6
Junior	18.1	-17.1	1.9	-7.1
High School	17.1	-15.6	6.2	-9.2
Some College	13.4	-15.2	6.1	-7.5
College	11.5	-13.4	-0.4	-8.6
Mild ADL	21.1			
Moderate ADL	18.0	-15.9	4.7	-9.0
Severe ADL		-18.3		

Table 1.13: Labor Supply Responses

Notes: The unit is percent change to the probability of working in comparison to the status quo. The labor supply responses reported are conditional on parents having LTC needs.

Name	Open Eligibility			No Eligibility			Relaxed Eligibility			Limited Eligibility		
	Total CV	Affected Share	Affected CV	Total CV	Affected Share	Affected CV	Total CV	Affected Share	Affected CV	Total CV	Affected Share	Affected CV
Daughter	0.358	0.373	0.959	-0.082	0.111	-0.742	0.017	0.038	0.457	-0.011	0.024	-0.470
Children-In-Law	1.103	0.457	2.412	-0.254	0.138	-1.837	0.057	0.045	1.282	-0.095	0.068	-1.399
Son.	0.212	0.737	0.287	-0.043	0.223	-0.194	0.010	0.067	0.145	-0.011	0.080	-0.142
Primary.	0.326	0.471	0.691	-0.049	0.081	-0.604	0.014	0.044	0.324	-0.017	0.035	-0.471
Junior	0.428	0.556	0.769	-0.140	0.207	-0.673	0.033	0.059	0.560	-0.034	0.053	-0.640
High School	0.432	0.597	0.723	-0.084	0.207	-0.405	0.020	0.060	0.336	-0.019	0.065	-0.301
Some College	0.466	0.617	0.756	-0.083	0.281	-0.294	0.008	0.052	0.147	-0.013	0.066	-0.204
College	0.428	0.669	0.639	-0.150	0.270	-0.555	0.012	0.047	0.261	-0.030	0.105	-0.286
Mild ADL	0.504	0.703	0.717									
Moderate ADL	0.244	0.369	0.661	-0.218	0.374	-0.581	0.133	0.360	0.368	-0.153	0.377	-0.405
Severe ADL				-0.355	0.701	-0.506						

Table 1.14: Compensating Variation

Notes: "Total CV" and "Affected CV" are normalized by mean annual wage. For example, a daughter's total CV for open eligibility, 0.358, means that she requires 35.8% of the mean annual wage to accept removal of this policy. "Affected Share" represents the share of those affected by the policy among children whose parents have ADL needs.

CHAPTER 2

RESERVATION WAGES AND WORKERS' VALUATION OF JOB FLEXIBILITY: EVIDENCE FROM A NATURAL FIELD EXPERIMENT

2.1 Introduction

The last decade has witnessed a dramatic increase in the prevalence of flexible working, either via workers entering the Gig Economy or historically traditional jobs becoming more flexible, allowing the worker to choose specific hours or where to work. These changes raise several questions of both policy and practical importance. How do labor supply elasticities and reservation wages vary across days of the week and hours of the day? To what extent do labor supply elasticities and reservation wages differ between people such as men and women or old and young? How do different workers value the ability to customize work schedules? While both economists and policymakers are keenly interested in these questions, credible answers have been hindered by a lack of high frequency panel data on wages and work decisions as well as by the difficulty of identifying how labor supply elasticities, reservation wages, and the value of flexibility vary between people and over time.

The goal of our paper is to answer the above questions while addressing both the measurement and the identification challenges. The context of our study is the largest ride-sharing company, Uber. Our work draws on three strengths of this environment.[1] First, Uber allows a driver to work anytime she is willing to accept the wage she would be paid in the market. Second, we have access to high frequency panel data on the wage an individual is paid and her decision to work.[2] Third, via a large natural field experiment we observe reactions to exogenous variation in expected market wages across individuals and over time.

Combining the panel data with the experiment, we first estimate individuals' labor supply responses to exogenous changes in expected market wages. These experimental findings motivate and guide our modeling of the labor supply of the drivers. The primitives of the

1. [37] describe the labor market for Uber's drivers. They find that drivers cite flexibility as a reason for working for Uber and that many drivers report that Uber is a part-time activity secondary to more traditional employment.

2. As in [21] and [14], we calculate the "wage" in an hour as a driver's total earnings in that hour divided by minutes worked (i.e. the number of minutes for which a driver has the app on and is available for accepting requests).

model are recovered from a combination of the experimental estimates and other data moments. We use the estimated model to compute how labor supply elasticities and reservation wages vary between people and over time and to perform counterfactual analyses. These analyses allow us to infer the drivers' willingness to pay for flexible work arrangements.[3]

In Section 1.2, we describe the labor market for Uber's drivers and the natural field experiment. The analyses of the experiment yield three main findings, which we present in Section 1.3. The first main finding is that the labor supply responses vary systematically both across people and over time. In order to discover these heterogeneous effects, we apply the method of [15] to estimate an instrumental variables model with a full set of interactions between the endogenous regressor of interest, wages, and the pre-determined covariates, gender, hours of the day, and days of the week. A clear pattern of heterogeneity emerges: Labor supply is most responsive in the evenings; men and older drivers have, on average, larger responses than other drivers. Taken together, these results suggest it is key to allow preferences of the drivers to vary by gender and age for each demographic group across hours of the day and days of the week.

The second main finding is that drivers do not only increase their labor supply during the periods with exogenously higher wages but also in the hours preceding and following these periods. This finding of anticipatory and persistent responses to increases in expected wages is consistent with forward looking drivers with fixed costs of starting to drive. In the presence of such adjustment costs, a static labor supply model is insufficient to analyze the behavior of the drivers. A dynamic model is needed to capture the connection between the decision to drive in the current period and future utility.

The third main finding is that unobserved determinants of wages, if ignored, lead to a significant downward bias in the estimated labor supply responses. In particular, OLS esti-

3. In the working paper version of this paper (https://www.nber.org/papers/w27807), we also perform another counterfactual experiment which allows us to examine how preference heterogeneity and adjustment costs influence the effectiveness of driver incentives that Uber may offer.

mates show much weaker associations between labor supply and wages than the experimental estimates. This downward bias is consistent with demand being high when it is costly for the drivers to work. Including fixed effects for workers, days of the week, and hours of the day reduces the bias, but the labor supply elasticities remain too small. This finding suggests that idiosyncratic factors, such as weather conditions and entertainment events, may create high demand while, at the same time, make driving more costly or difficult.

The experimental estimates provide key data points for recovering reservation wages, labor supply elasticities and the value of job flexibility, but do not by themselves tell us these quantities. To do so, we develop, in Section 4, a dynamic model of labor supply. This model builds on the experimental findings and accommodates important features of the market, including uncertainty about wages and costs of driving in the future, the possibility of a job other than driving for Uber, and fixed costs of starting to drive. When taking the model to the data, we allow for both observed and unobserved heterogeneity across drivers, and we allow market wages to be correlated with the cost of driving in a given period. Even with these considerations, it is possible to prove identification of the primitives of the model given the panel data and the experiment that creates exogenous variation in expected market wages.

We use the EM algorithm to find the maximum likelihood estimates of the model parameters. The parameter estimates suggest significant costs of starting to drive and considerable observed and unobserved heterogeneity in preferences. Conditional on age and gender, there appears to be three types of drivers: The 'infrequent driver' who only drives occasionally; the 'full-time driver' who drives regularly both in the evening and during the day; and 'the evening driver' who rarely drives during the day, possibly because she has a daytime job. To assess the importance of substitution between Lyft and Uber, we compare the results for all drivers to those we obtain from a subsample of drivers who are ineligible to drive for Lyft. It is reassuring to find that both the experimental estimates and the estimated model parameters do not materially change when we restrict attention to this subsample.

The model delivers two key insights, presented in Section 5. The first insight from the model is that reservation wages, and thus the shadow prices of time, vary a lot both over time and across people.[4] For an average driver at a typical day, the reservation wage is relatively low during the day, starts increasing in the late evening, peaks at around 4 a.m., and then declines gradually until 9 a.m. By way of comparison, there is little variation in reservation wages across days of the week: On average, the reservation wage is only a few percent lower on weekdays than during the weekends. Holding day of the week and hour of the day fixed, there is also a great deal of variation in reservation wages across people. On average, reservation wages of women are 106 percent higher than male reservation wages. There is also a great deal of heterogeneity conditional on observables: The infrequent drivers have much higher reservation wages than the full-time drivers, while the evening drivers demand relatively high wages to drive during the day.

The second insight from the model is that drivers would demand much higher wages if they had to commit to pre-set work schedules. We quantify the importance of two distinct types of job flexibility. One is the ability to set a customized work schedule, so that each driver may plan to work only when her expected reservation wage is lower than the expected wages. We quantify the value of this type of flexibility by removing certain hours of the day or days of the week from the choice set of the workers. Our findings suggest that drivers are particularly averse to restrictions on what hours of the day to work. By way of comparison, constraining drivers to work only on the weekends or only on weekdays would require a modest increase in wages. The other type of flexibility we consider is the ability to adjust the schedule from day to day or even hour to hour in response to unexpected changes to offered wages or costs of driving. We measure the value of this flexibility by

4. It is important to recognize that a driver's reservation wage should be interpreted as a shadow price of time that reflects not only leisure possibilities but also alternative economic activities such as home production or other jobs.

restricting drivers to stick to the work schedule they prefer before observing any shocks to wages and preferences. Our findings suggest that Uber drivers, especially those who are female, benefit significantly from the possibility to adapt work schedules to unexpected events. Taken together, these results suggest that job flexibility is a central component of the total compensation of ride-sharing companies like Uber.

Our paper is primarily related to a large literature on labor supply. The models, data, and findings have been summarized and critiqued in multiple review articles including [64], [47], [10], [45] and [16]. Most models of labor supply are concerned with the problem of choosing how much to work, not when to work. In many of these models, there are no hours restrictions, and each worker supplies labor until the wage she would face in the market equals the value she places on her time, the reservation wage. When taking such models to the data, labor supply elasticities and reservation wages are typically inferred from differences in work hours across people given their observed wages. There are, however, several concerns with this revealed preference argument. One of these concerns is that both theory and evidence suggest restrictions on hours choices stemming from the demand side of the market. This concern motivates a large body of work that incorporates hours restrictions in models of labor supply under the assumption that the analyst has full or partial knowledge about the probability distributions of either offered or desired hours of work.[5]

To avoid making questionable assumptions about hours restrictions, we take advantage of the fact that Uber is a platform on which drivers, once approved, are free to choose their work hours. There are no minimum-hours requirements and only modest constraints on maximum hours. As a result, our estimates of extensive margin labor supply elasticities and reservation wages are not confounded by hours restrictions from the demand side of the market. Instead, the estimated elasticities capture the sensitivity of the decision to supply labor in a given hour to anticipated and exogenous changes in hourly market wages. The wage changes we

5. See [9] and the references therein for details.

consider are modest and temporary, so that lifetime wealth is approximately unchanged. Thus, our setting allows us to recover estimates of extensive margin Frisch elasticities per hour and elasticities of intertemporal substitution (IES) between hours, which in our model differ due to adjustment costs.

Averaging over time and across drivers, we find an extensive margin Frisch elasticity of 0.65, and an IES of 0.45. The estimated Frisch elasticity is significantly larger than what is typically reported in micro studies that ignore or make assumptions about hours restrictions from the demand side of the market [16]. Our IES falls in the range of 0.22 and 0.60, comparable to those by [32] and [33]. By contrast, [4] report estimates of IES close to one. Their estimates are based on a comparison of the commission-based compensation model of Uber and the conventional taxi contract. However, as emphasized by [58] in their review of the literature, it is difficult to compare the estimates of IES across studies, in part because the restriction on hours may vary but also because the accounting period differs (e.g. days, weeks, or years).

The closest study to ours is arguably the work of [14]. Like us, they take advantage of the fact that Uber has virtually no hours restrictions.[6] Thus, [14] argue, one can recover how reservation wages vary across people and time by relating the probability an individual drives in a given time period to the mean prevailing market wage for that period. Using a multivariate probit model with time-varying thresholds for work decisions, they estimate driver-specific reservation wages, and then decompose these reservation wages into predictable and unpredictable components. Armed with the estimates from this static labor supply model, they calculate the surplus from driving for Uber and the surplus changes that would result from requiring the driver to instead work specific patterns of hours.

Our paper complements and extends the model and analyses of [14] in several important

6. There are also other papers using data from Uber. [4] study how workers' view the commission-based compensation model of Uber as compared to traditional taxi compensation contract. [19] estimate consumers' demand and surplus from Uber rides. [21] study the determinants of the gender earnings gap amongst Uber drivers.

ways. First, motivated by our natural field experimental results, we develop, identify, and estimate a dynamic model of labor supply with fixed costs of starting a shift. Second, we allow market wages to be correlated with the unobserved cost of driving in a given period. Third, we use a natural field experiment to identify the primitives of the model. Fourth, we allow permanent heterogeneity both by the drivers' observable characteristics and according to their unobserved latent types. Empirically, we find that these modeling choices are important to match the data as well as for the estimated reservation wages and the counterfactual analyses. Our paper also offers a complementary perspective on the heterogeneity in reservation wages. [14] model and estimate the heterogeneity in reservation wages as arising from idiosyncratic preferences. We show there is a systematic and predictable pattern in the reservation wages by not only the day of week or hour of the day but also according to the gender, age, and type of driver. This pattern is useful to better understand who benefits from flexible work arrangements, and, as a result, it may also help improve the design of driver incentives and inform discussions over recent policy proposals about regulation and pay rules for ride-sharing companies.

Our paper also relates to a body of work on the labor supply of taxi drivers. The primary goal of this work is to estimate the wage elasticity of daily hours of work to test if labor supply behavior is consistent with reference dependence. The work is summarized and critiqued in [33]. He also replicates and extends existing work. His findings suggest that reference dependence is not an important factor in the daily labor supply decisions of taxi drivers.

Some of our findings are similar to those reported in [33]. For example, much of the variation in hourly wages is predictable based on the day of the week and the hour of the day, and drivers are more likely to work when market wages are high. Other findings differ. For instance, [33] finds that the probability of ending a shift depends strongly on hours worked. We do not find support for such fatigue being empirically important for the behavior of Uber drivers. However, the environment and decision problem of Uber drivers differ in important ways as compared to taxi drivers. In particular, accumulated hours worked in a given day tend to be a lot higher for taxi drivers, and, as a result, fatigue could be more salient for whether they continue driving or end a shift. By comparison, the labor supply of Uber drivers is best described by a combination of adjustment costs in terms of starting to drive and heterogenous reservation wages, especially by hour of the day and type of driver.

Another literature to which we relate is the research on how individuals value workplace amenities such as job flexibility. Survey evidence shows that workers state that they are willing to take lower pay for more flexible jobs (e.g., [38]; [65]; [30]; [55]). However, recovering the workers' actual valuation of job flexibility from naturally occurring data has proven difficult for several reasons. One challenge is that firms may pay differently simply because they employ workers of different quality. A second challenge is that observed wage variation across firms may reflect workplace amenities other than job flexibility. Most research to date tries to address these issues by controlling for worker and firm characteristics, hoping that any remaining wage variation across firms is due to job flexibility.[7]

Even if these controls were sufficient to address concerns about omitted variables bias, it is important to observe that additional assumptions or data are needed to draw inference about workers' valuation of job flexibility. Wage differentials across firms could reflect imperfect competition in the labor market, not workplace amenities. Additionally, in standard models of equalizing differences, such as [66], the observed wage differentials are the market prices of amenities, providing only information on the valuation of marginal workers. [49] develop, identify and estimate an equilibrium model of the U.S. labor market with two-sided heterogeneity where workers view firms as imperfect substitutes because of heterogeneous preferences over workplace amenities. The estimated model makes it possible to distinguish between and draw inference about imperfect competition, compensating differentials, and

7. [59], [14] and [40] review this literature.

the distribution of worker preferences over amenities. The empirical findings suggest one needs to be cautious in extrapolating the valuation of amenities among marginal workers, as measured by the compensating differentials, to the valuation of inframarginal workers, who extract a significant amount of surplus or rents from workplace amenities. The importance of worker heterogeneity in the value of amenities like job flexibility is consistent with both our findings and those in [14].

Our findings on job flexibility complement recent evidence that uses a stated preference approach to infer workers' preferences based on their choices between pairs of exogenously assigned hypothetical jobs with different combinations of amenity levels and pay (e.g., [59]; [69]). For example, [59] use a discrete choice experiment in hiring for a U.S. call center to estimate the willingness to pay for alternative work arrangements relative to traditional office positions. A significant number of workers state that they are willing to give up a substantial share of their wages to avoid a schedule set by an employer on short notice. By comparison, the stated willingness to pay for choosing when to work is relatively low.

[40] provide complementary evidence from a revealed preferences approach to estimating workers' valuation of flexibility. They combine data from a natural field experiment conducted on a Chinese job board with survey and observational data. The experimental job ads differ randomly in offering jobs that are flexible regarding when and where one works. Both the survey evidence and the experimental estimates suggest that workers are willing to take lower pay for more flexible jobs. For instance, application rates are significantly higher for flexible jobs, conditional on the salary offered. [40] argue that a natural field experiment offering real jobs to real job seekers has several advantages over alternative approaches.[8] The participants in the natural field experiment are actually searching for jobs, properly incentivized to respond in ways most likely to get them the jobs they want, and unaware they were under scrutiny in a scientific study. The natural field experiment that we study

8. See [39] for a broader discussion of the advantages of natural field experiments.

have the same advantages. In addition, we can measure the participants' valuation of flexibility in terms of their actual work decisions, unlike the job board experiment that does not capture the final outcomes of the search process (such as callbacks for interviews, job offers, and actual remuneration).

Finally, methodologically, we join a set of recent studies that combines field experiments with structural methods to uncover key counterfactuals (see, e.g., [25]; [26]). In doing so, we highlight how the combination of theory and field experiments can be used to evaluate a wide range of economic issues (see also [48]).

2.2 Background and Experiment

We now review the labor market for Uber's drivers before describing the natural field experiment.

Uber Marketplace

Uber's rideshare platform is the largest service provider in the ride-sharing market in the U.S. In 2016, for example, it had a market share of about 83% of the U.S. consumer ride-sharing market. Uber and Lyft combined owned nearly 97 percent of this market. Uber connects riders and drivers through its app. Once a ride request is made, the app contacts nearby drivers for the ride. Drivers would see the rider's location. While drivers are incentivized to maintain a high acceptance rate, drivers can decide whether to accept this trip.

Drivers are effectively free to choose when and how much to work. There are no minimum-hours requirements and only modest constraints on maximum hours. Drivers are paid according to a fixed, non-negotiated formula. As described in detail later, workers earn a base fare per trip plus amounts for how long and how far they drive. On top of this standard fare, Uber offers fare multipliers when the demand for rides is sufficiently high compared to the supply of drivers (commonly referred to as surge pricing) or if the drivers are participating in the randomized experiment. Both the variation in surge multipliers and the randomness in the arrival of rider requests lead to variability in effective market wages per hour.

In most U.S. cities, there are relatively few barriers to becoming an Uber driver. While the exact requirements vary from city to city, drivers must typically fill out online paperwork, undergo a background check, and meet certain driver and vehicle requirements. In the years (2016-2018) and cities (Boston, Chicago, San Francisco) we consider, one of these requirements is that the vehicle's model year is 2001 or newer. By comparison, Lyft required that the vehicle must have a year model of 2003 or newer. As a robustness check, we take advantage of this difference in eligibility requirements to assess the importance of substitution between Lyft and Uber.

Experiment Setup: Guaranteed Surge Level

Our natural field experiments arises from the so-called Guaranteed Surge Levels (GSL hereafter). The GSL is essentially a fare multiplier that Uber randomly offers to a subset of drivers to increase their expected market wages during certain hours. Drivers who were active in the past 28 days and have completed at least 40 trips are eligible to receive fare multipliers through GSL.

The experimental setup is as follows. Uber divides each week into 2 blocks: Block 1 starts from Monday 4:00 a.m. and ends on Friday 3:59 a.m., while Block 2 goes from Friday 4:00 a.m. to 3:59 a.m. the next Monday. Over the course of a given block, GSL is switched on for a subset of hours. In Figure 2.1, we show an example of Block 1. During the example block, the highlighted hours, such as Monday 5:00–6:00 a.m., are chosen as experiment hours where drivers receive hour-specific fare multipliers. We refer to consecutive experiment hours as an experiment window. We refer to the schedule of GSL experiment hours within a block as a GSL menu. These menus vary across blocks. Drivers learn about the GSL menu, via email and/or the Uber app, the night before a block starts. Figure 2.2 shows an example of a GSL menu with multiple experiment windows, making clear to drivers when and for how long GSL will be switched on in an upcoming block.

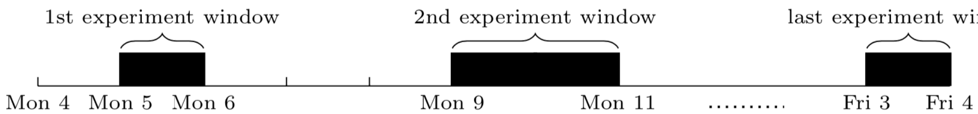


Figure 2.1: Example of a GSL Menu of Block 1 with Multiple Experiment Windows

Notes: Highlighted hours denote the hours when the GSL experiment is switched on.

For each block, eligible drivers are randomly assigned to treatment and control groups where the treatment group receives 0.1 higher GSL fare multipliers than the control group for all the experiment hours within the block. Consider again the example in Figure 2.1. Monday 5:00–6:00 a.m. and Monday 9:00–11:00 a.m. are two experiment windows. Suppose the control group drivers receive 1.1× fare multiplier in the first window and 1.3× fare multiplier in the second window. Since the treatment group always receives 0.1 higher GSL fare multipliers than the control group, the treated drivers would then be receiving 1.2× fare multiplier in the first window and 1.4× fare multiplier in the second window. At the end of each block, drivers are re-randomized into treatment and control groups for the next block.

On average, an experiment window lasts for about 5 hours, and there are about 7 experiment windows per block. Across blocks, there is variation in the days and hours of the experiment windows. In total, around 40 percent of the hours in our sample are subject to the GSL experiment. Thus, the GSLs generate considerable variation in expected wages at different days of the week and at various hours of the day.

Trip Earnings, Wages and Work Decisions

The unit of observation in our analysis is an individual driver at a given hour. Thus, we measure labor supply and wages on an hourly basis. We define total minutes worked per hour as the number of minutes for which a driver has the app on and is available for accepting

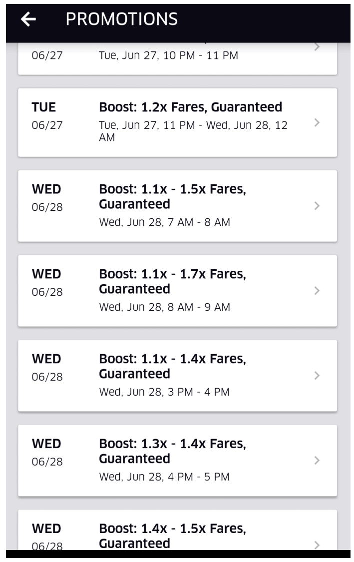


Figure 2.2: Example of a GSL Announcement to Drivers

Notes: The screenshot shows a GSL announcement to drivers on the Uber Driver App. The announcement clearly shows the hours when the GSL will be switched on and the corresponding GSL multiplier.

requests in that hour. In other words, a driver is said to be working if she is actively searching for rider requests.[9]

The observed hourly wage rate is measured as a driver's total earnings divided by the number of minutes worked in an hour, multiplied by 60. The total earnings in an hour are measured as the sum of trip earnings a driver receives, where each trip earnings is determined by the following formula:

$$\text{Trip Earning} = \max(\text{GSL}, \text{Surge}) \times \text{Baseline Fare}.$$

In this formula, GSL is the experimental fare multiplier, Surge is the demand-driven fare multiplier, and BaselineFare is the baseline trip earnings following Uber's fixed compensation

9. This is the same definition of working as in [14]. A driver is active if the driver-side app is turned on and she is available to accept requests for rides. This is to be distinct from a "browsing" mode in which the app is on but the driver has not indicated a willingness to accept rides.

rule:

$$\text{Baseline Fare} = \text{Fixed Price} + (\text{Price per Minute} \times \text{Minutes}) + (\text{Price per Mile} \times \text{Miles}).$$

As in [14], our measure of wages does not net out the variable costs of operating a vehicle. Therefore, our reservation wages should be interpreted as a gross quantity. It is important to observe, however, that labor supply decisions are based on the differences between expected wages and reservation wages, which do not depend on assumptions regarding the incorporation of time-invariant operating costs.

Drivers base labor supply decisions on expected hourly wages rather than the realized wages that we observe. To construct measures of expected wages, we predict the hourly wage a driver is likely to face in each hour. As a first step, we calculate the wage multiplier, defined as $\max(\text{GSL}, \text{Surge})$, from our detailed data on GSL and Surge. Next, we calculate the pre-multiplier wages as the observed hourly wages divided by the calculated wage multipliers. We then fit the following regression model to the panel data on pre-multiplier wage:

$$\tilde{W}_{it} = \alpha_i + \kappa_{h(t)} + \epsilon_{it}$$

where t is an hour, $h(t)$ is the hour of the week at t, i denotes a driver, α_i and $\kappa_{h(t)}$ are driver and hour-of-week fixed effects, and $\tilde{W}_{it}$ is the pre-multiplier hourly wage. For each hour t, we fit the model with the panel data up to $t-1$, and use the estimated $\hat{\alpha}_i$ and $\hat{\kappa}_{h(t)}$ to compute a predicted value for $\tilde{W}_{it}$ for every worker i in each hour t. The predicted hourly wage is then constructed as the product of the predicted pre-multiplier wage and the calculated wage multiplier.

To assess how well our prediction model performs, we compare it to alternative approaches using a cross-validation procedure with details in Appendix 2.A. This procedure repeatedly divides samples into a training sample and a testing sample. For each approach, we use the training samples to estimate a prediction model, and then we use the estimated model to form a prediction and calculate the out-of-sample mean squared errors on the testing samples. In addition to our current prediction model, we consider several alternative approaches, including a matching procedure with a K-means clustering method. Our prediction model performs considerably better than the alternative approaches.

2.3 Data and Experimental Findings

In this section, we describe the data and present the findings from the experiment.

Description of the Data

Our analyses are based on panel data of UberX and Uber Pool drivers who are eligible for the GSL experiment in Boston, Chicago, and San Francisco. In Boston and Chicago, we observe all these drivers. In San Francisco, we have data for a random subsample of 35 percent of the eligible drivers. In Boston and San Francisco, the duration of our data spans from October 2016 to March 2018. In Chicago, we have only one year of data, covering October 2016 to March 2017. For each driver, we observe gender, age, type of vehicle, minutes worked per hour, trip earnings, and fare multipliers.

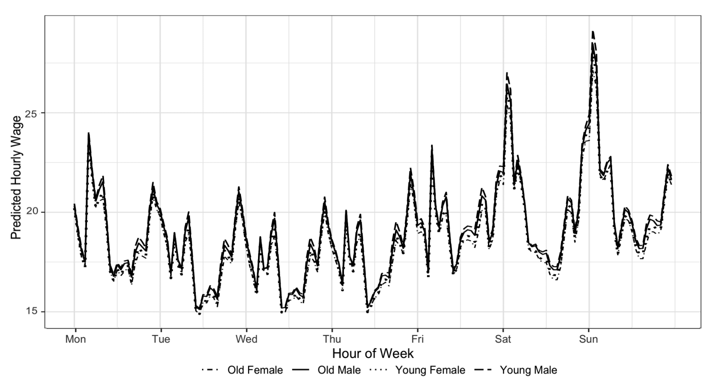


Figure 2.3: Predicted Hourly Wages Across Hours of the Week by Demographic Groups

Notes: We compute the average predicted wage at every hour of the week and for each demographic group.

Our sample covers 333,172 drivers. A quarter of these are female, and the median age is about 38. The average observed wage is \$19.29 per hour. In Figure 2.3, we plot the predicted hourly wage over time according to gender and age. We define drivers as young if they are younger than 38 years old. Consistent with [14], most of the heterogeneity in hourly wages is due to hours of the day. It is also evident that hourly wages tend to be higher in the weekends and that male drivers have only slightly higher hourly wages as compared to female drivers.

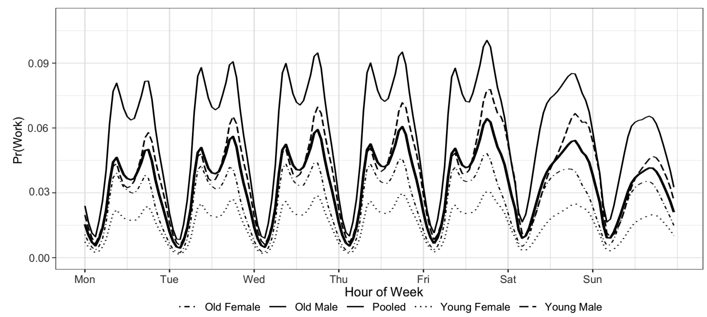


Figure 2.4: Probability of Working Across Hours of the Week by Demographic Groups

Notes: In this figure, we compute the share of active drivers at every hour of the week and for each demographic group. We define active drivers as those who work any positive number of minutes in a given hour. "Pooled" refers to the combined sample across demographic groups.

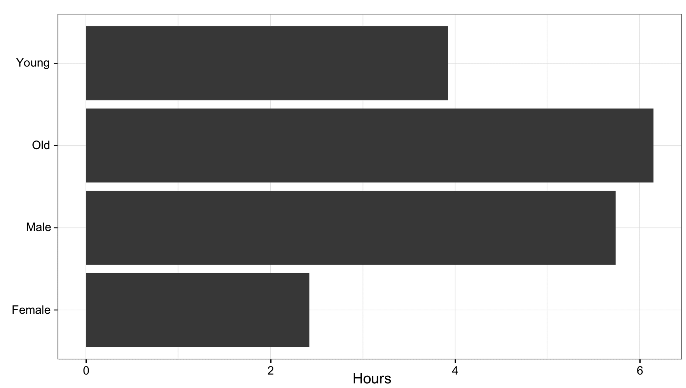


Figure 2.5: Average Number of Hours Worked per Week by Demographic Groups

Notes: In this figure, we compute the average number of hours worked for each demographic group. An hour worked is defined as an hour where a driver works for any positive number of minutes.

Figure 2.4 plots the probability of working over time for all drivers and by subgroup. The work probability varies considerably across hours of the week and days of the week. There are also distinct differences in the probability of working by age and gender. Conditional on age, male drivers are much more likely to work, especially during the daytime. Holding gender fixed, old drivers tend to work more than young drivers. In Figure 2.5, we show the average number of hours worked per week by age and gender. In expectation, drivers work nearly five hours per week. However, males drivers work twice as many hours per week as female drivers, and young drivers work 36 percent less than older drivers.

Checking Covariate Balance

In a properly implemented, randomized experiment with a sufficiently large sample size, we expect the treatment and control groups to be balanced in their distribution of pre-treatment variables. To assess this, we check the covariate balance by regressing the treatment status in the GSL experiment on the pre-treatment characteristics of the drivers:

$$D_{it} = X_{it}'\beta + u_{it}$$

where D_{it} is an indicator variable that is equal to 1 if driver i is assigned to the treatment group in the GSL experiment at time t, and X_{it} is a vector of covariates that include gender, age, number of trips completed, and past wages. All these covariates are measured in the week before the randomization. Table 2.1 reports the estimates. In Column 1, we regress treatment status on each characteristic separately. There is no evidence of systematic differences between drivers in the treatment and control groups in the characteristics considered. In Column 2, we regress treatment status on all the characteristics in a multiple regression. Consistent with the randomization, we cannot reject the null hypothesis that all coefficients are zero.

Pre-determined Var.	(1) **Separate Regression** Var. on Treatment	(2) **Joint Regression** Treatment on Var.
Female ($\times 10$)	-0.0039	0.0000
	(0.0025)	(0.0001)
Age ($\times 100$)	-0.0003	-0.0651
	(0.0009)	(0.0426)
Wage Last Week ($\times 100$)	0.0051	-0.0000
	(0.0025)	(0.0000)
Trips Completed ($\times 1000$)	-0.0000	0.0536
	(0.0000)	(0.0259)
N (Blocks)	22,318,255	22,318,255
R^2		0.0000
F Statistic		1.8450
p value - F		0.1171

Table 2.1: Balance Tests

Notes: In Column 1, we regress each driver characteristic on the treatment status separately, and each row represents a separate regression. In Column 2, we regress the treatment status on all the characteristics in one regression. We use the F-test to examine whether one or more of the coefficients of these four pre-determined variables are significantly different from zero. For interpretability, we scale the regression coefficients and the standard errors of the driver's gender by 10, the driver's age and past wages by 100, and the driver's total number of trips completed by 1000.

First Stage: Effects of GSL on Expected Market Wages

Figure 2.6 presents the treatment effects of the GSL experiment on the expected market wages during the experiment hours. These treatment effects are obtained by OLS estimation of the predicted hourly wage on a dummy variable of being in the treatment group. These estimates form the first stage in the IV estimation of the effects on labor supply of exogenous changes in predicted wages. The shaded area in the figure indicates the hours with the experiment switched on. The figure shows that assignment to the GSL experiment increases the average predicted hourly wage by around forty cents or, equivalently, around 1.8 percent.To examine if the wage effects are persistent across experiment hours, we divide each experiment window in half, and then estimate the treatment effects separably for each half. The estimates suggest little, if any, changes in the wage effects across hours within the experiment window.

Reduced Form: Labor Supply Responses to GSL

We now document that drivers respond to the GSL experiment both during the experiment windows and in the hours preceding and following these periods. This is done by OLS estimation of labor supply on a dummy variable of being in the treatment group. These effects will be the reduced form estimates in the IV estimation of the effects on labor supply of exogenous changes in predicted wages. We consider labor supply responses along two margins: In each hour, we measure if the driver worked at all and the number of minutes she worked.

In Figure 2.7, we plot the effects of the GSL experiment on the labor supply responses of the drivers before, during, and after the experiment window. These effects are represented by the solid lines. The dotted lines represent the changes in wages due to the GSL experiment. We find significant changes in labor supply during the experiment window. During this window, the treated drivers increase the employment rate per hour and the hours of work by about one percent as compared to the control drivers. There is also some suggestive evidence of persistent effects outside the experiment window. For example, in the hour following the experiment window, the labor supply is half a percent higher for the treated group as compared to the control group.

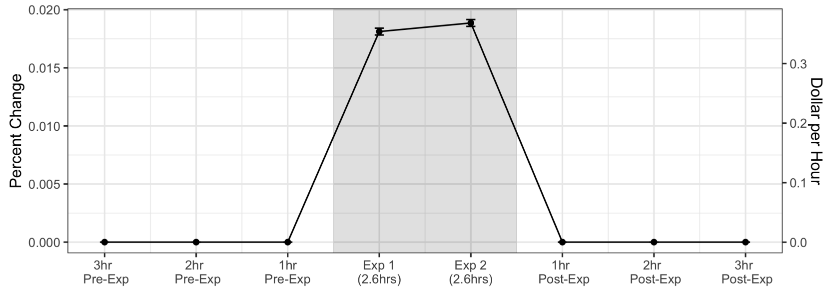
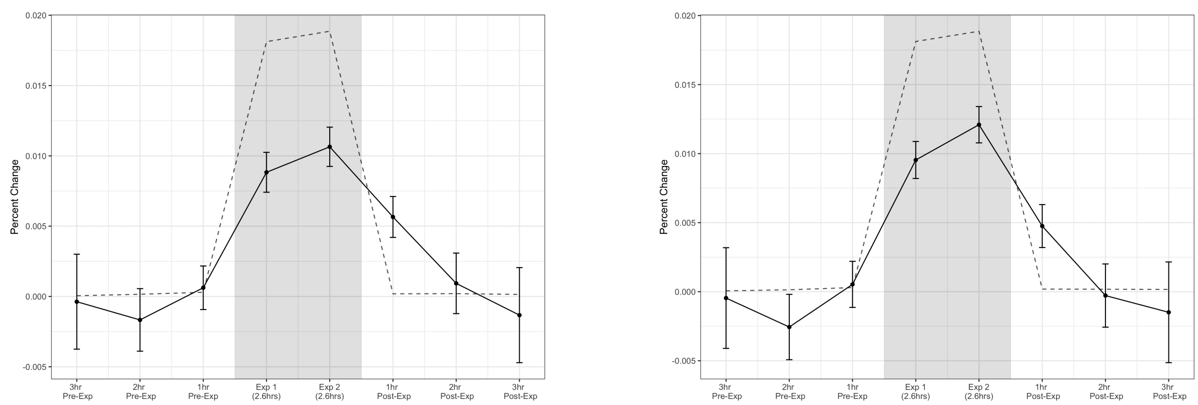


Figure 2.6: Treatment Effects of the GSL Experiment on the Expected Market Wages

Notes: In this figure, we present the estimates of the GSL experiment on the expected market wages during the experiment hours (in the hours before and after the experiment window, the effect is zero). The estimation is performed separately for the first half (Exp 1) and the second half (Exp 2) of the experiment windows. On average, the experiment windows last 5.2 hours. The shaded area in the figure indicates the experiment hours with the GSL switched on. The y-axis on the left shows percent differences between the treatment and the control groups. It is computed as the treatment effect on the predicted hourly wage divided by the predicted hourly wage of the control group. The y-axis on the right reports the estimated effects in dollars per hour. The bars indicate the 90% confidence intervals calculated from subsampling bootstrap.



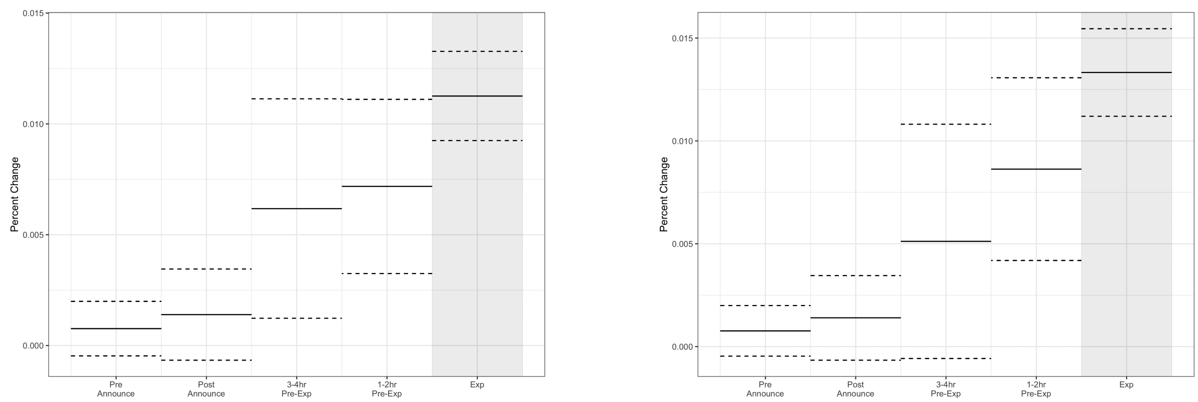
(a) Working

(b) Minutes Worked

Figure 2.7: Treatment Effects on Labor Supply Responses

Notes: In this figure, we illustrate the changes in the probability of working and minutes worked. We recenter all the experiment windows and plot the x-axis the same way as in Figure 2.6. The estimation is performed separately for the first half (Exp 1) and the second half (Exp 2) of the experiment windows. On average, the experiment windows last 5.2 hours. The shaded area in the figure indicates the experiment hours with the GSL switched on. The solid line represents the difference in labor supply responses, and the dashed line represents the difference in predicted hourly wages. The bars indicate the 90% confidence intervals.

When interpreting the estimated effects in the hours preceding and following the experiment window, it is important to recognize that most blocks have several GSL experiment windows. As a result, the hours preceding and following a given experiment window are likely to be confounded by other GSL experiment windows. For example, in 30 percent of the blocks in our sample, there exists at least two experiment windows that are no more than an hour apart. To address this concern, we take advantage of the re-randomization at the start of every block and restrict attention to the experiment windows preceding and following the re-randomization. By estimating the labor supply responses around these experiment windows, we avoid confounding anticipatory and persistent responses with other GSL experiment windows.

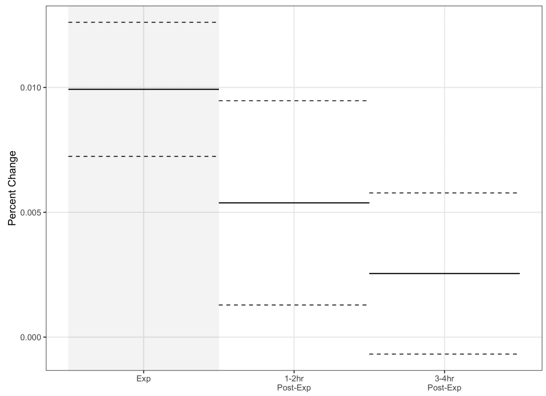


(a) Working

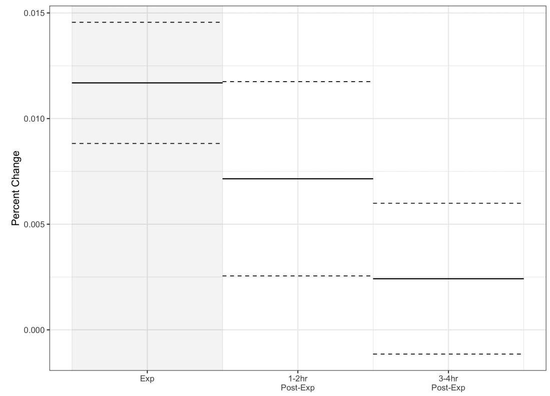
(b) Minutes Worked

Figure 2.8: Treatment Effects on Labor Supply Responses in First Experiment Windows

Notes: We illustrate the labor supply responses immediately after re-randomization. We pool all the first experiment windows for estimation. The point estimates are constructed in the same way as Figure 2.7. The shaded area in the figure indicates the experiment hours in the first experiment windows. The "Pre Announce" period contains all hours before the announcement of the GSLs, while the "Post Announce" period contains the hours after announcement up to 5 hours before the first experiment windows. The dashed bars indicate the 90% confidence intervals.



(a) Working



(b) Minutes Worked

Figure 2.9: Treatment Effects on Labor Supply Responses in Last Experiment Windows

Notes: We illustrate the labor supply responses immediately before re-randomization. We pool all the last experiment windows for estimation. The point estimates are constructed in the same way as Figure 2.7. The shaded area in the figure indicates the experiment hours in the last experiment windows. The dashed bars are the 90% confidence intervals.

In Figures 2.8 and 2.9, we present the estimates from the experiment windows preceding and following the re-randomization. Even four hours prior to the experiment window, we see the treated drivers are more likely to be working as compared to drivers in the control group. The anticipatory response is most pronounced in the hour just before the experiment window. The same holds true for the persistent responses. In the first few hours following the experiment, the labor supply of the treated drivers remains significantly higher than the drivers in the control group. As time passes, these differences decline, and four hours after the experiment window the treated drivers work as much as the control drivers.

IV Results: Labor Supply Responses to Exogenous Changes in Expected Market Wages

We now turn attention to how the labor supply of drivers responds to exogenous changes in expected market wages. This is done by 2SLS regression with labor supply as the dependent variable, predicted wages as the treatment variable, and the experiment as the instrument.

The resulting 2SLS estimates correspond to the ratio of the reduced form and the first stage estimates reported above. Table 2.2 presents the IV estimates. In the first three columns, we measure labor supply as any work in a given hour. The last three columns measure labor supply as minutes worked during an hour. In Columns 1 and 4, we use the data from all the experiment hours. Columns 2 and 5 restrict attention to the hours surrounding the first experiment window in a block, whereas Columns 3 and 6 consider only the hours surrounding the last experiment window in a block.

The first stage estimates in Columns 1 and 4 are very precise and show that the GSL experiment raises the predicted hourly wage by nearly forty cents. More importantly, the IV estimates in these columns imply that a $10 increase in hourly wages would raise the share of workers that drive in a given hour by 1.4 percentage points or, equivalently, by 27 percent. By comparison, this increase in hourly wages would raise the amount of minutes that an average driver works by about 30 percent, from 2.26 to nearly 3 minutes per hour. The columns other than 1 and 4 quantify the labor supply responses of the drivers in the hours preceding and following the exogenous changes in predicted wages. We find that these anticipatory and persistent responses are significant and economically relevant. For both measures of labor supply, the anticipatory and persistent responses are about a third of the size of the responses during the experiment hours.

To better understand what drives the increase in labor supply, it is useful to decompose the IV estimates into responses on the extensive (any work in a given hour) and the intensive margin (minuted worked conditional on working in a given hour). Concretely, we decompose the estimate in Column 4 of Table 2.2 as follows:

$$\underbrace{\frac{\partial}{\partial W_t}E(\text{MinutesWorked}_t)}_{0.680} = \underbrace{\frac{\partial}{\partial W_t}E(\text{MinutesWorked}_t|\text{Work}_t) \times Pr(\text{Work}_t)}_{0.064\ (9.4\%)}$$
$$+ \underbrace{E(\text{MinutesWorked}_t|\text{Work}_t) \times \frac{\partial}{\partial W_t}Pr(\text{Work}_t)}_{0.616\ (90.6\%)}$$

where Work_t is an indicator variable that is equal to one if the driver works in a given hour. Using the drivers in the control group, we calculate $E(\text{MinutesWorked}_t|\text{Work}_t)$ and $Pr(\text{Work}_t)$ while $\frac{\partial}{\partial W_t}E(\text{MinutesWorked}_t)$ and $\frac{\partial}{\partial W_t}Pr(\text{Work}_t)$ are taken from the estimates in Columns 4 and 1, respectively. The results suggest that responses at the extensive margin account for nearly all the increase in the amount of minutes worked during an hour. This finding suggests it is important to model the driver's decision to work or not in a given hour, not the amount of minutes she works within an hour.

Heterogeneity in Labor Supply Responses

So far, we have focused on the average labor supply responses across all drivers. However, these average impacts miss a lot: The labor supply responses vary systematically both across people and over time. In order to discover these heterogeneous effects, we apply the method of [15] to estimate an IV model with a full set of interactions between the endogenous regressor of interest, wages, and the pre-determined covariates, gender, hour of the day, and day of the week. Since there are 168 hours per week, we have 168 bins for the time dimension. For age, we divide the sample into four equally sized groups: Younger than 30 years, 30 to 38 years, 38 to 48 years, and older than 48 years. Since gender is binary, we therefore get 1,344 mutually exclusive and collectively exhaustive groups.

We sort these 1,344 IV estimates in an increasing order. In Figure 2.10 we illustrate these sorted IV estimates. The x-axis represents the percentile rank in the distribution of

Dependent Var.	Working			Minutes Worked		
Panel A: IV						
	All	First	Last	All	First	Last
Experiment Hrs	0.0142	0.0167	0.0113	0.6801	0.8130	0.5558
	(0.0014)	(0.0009)	(0.0009)	(0.0654)	(0.0382)	(0.0394)
Anticipation		0.0053			0.2451	
		(0.0009)			(0.0447)	
Persistence			0.0036			0.1800
			(0.0009)			(0.0403)
Control Mean	0.0525	0.0474	0.0375	2.2600	1.9484	1.5636
of Dep. Var.	(0.0002)	(0.0002)	(0.0002)	(0.0084)	(0.0088)	(0.0075)
Panel B: First Stage						
	All	First	Last	All	First	Last
First Stage	0.3608	0.3193	0.3288	0.3608	0.3193	0.3288
	(0.0037)	(0.0057)	(0.0062)	(0.0035)	(0.0054)	(0.0061)
Control Mean	19.5091	20.3611	19.7459	19.5091	20.3611	19.7459
of First Stage	(0.0138)	(0.0196)	(0.0189)	(0.0133)	(0.0186)	(0.0180)

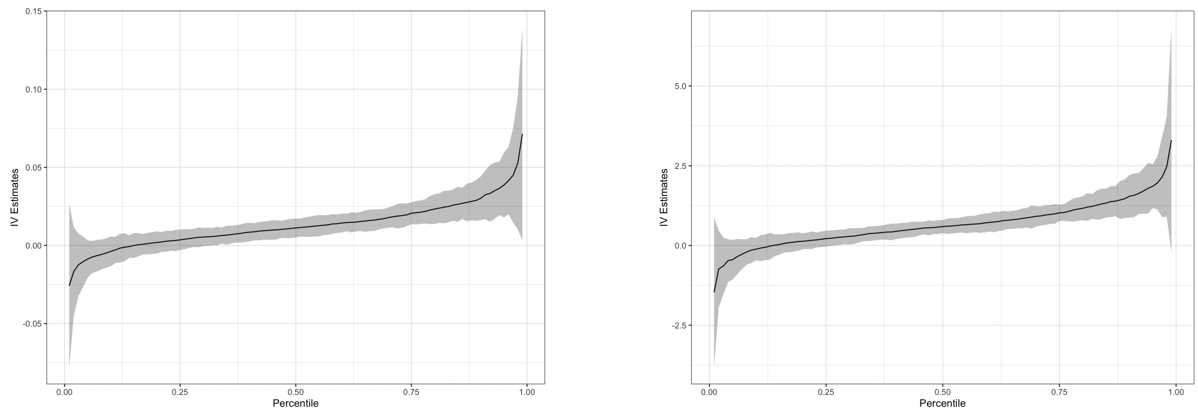
Table 2.2: Labor Supply Responses: IV and First Stage Estimates

Notes: "Control mean" is the expected outcome for the control group. IV is estimated as the increase in the probability of working per hour (or minutes worked per hour) per $10 increase in the predicted hourly wage. The standard errors of the IV estimates and the first stage are estimated by bootstrap. Column 2 and Column 5 show the estimates for the first experiment windows after re-randomization. Column 3 and Column 6 show the estimates for the last experiment windows before re-randomization.

the estimated effects, and the y-axis shows the estimated effect sizes. Panel (a) graphs the estimated effects for working and Panel (b) shows the estimated effects for minutes worked. It is evident that the IV estimates vary widely across groups. At the 25th percentile, the estimated effect for working is 0.0036, while the value at the 75th percentile is 0.021, almost 6 times as large. The estimated effect for minutes worked at the 75th percentile is almost 5 times as large as the value at the 25th percentile.

A natural question is what drives the heterogeneity in labor supply responses documented in Figure 2.10. To answer this question, we begin by examining the time dimension. In Figure 2.11a, we plot the average responses across all drivers for different hours of the day and for weekends versus weekdays. There is substantial heterogeneity along the time dimension.

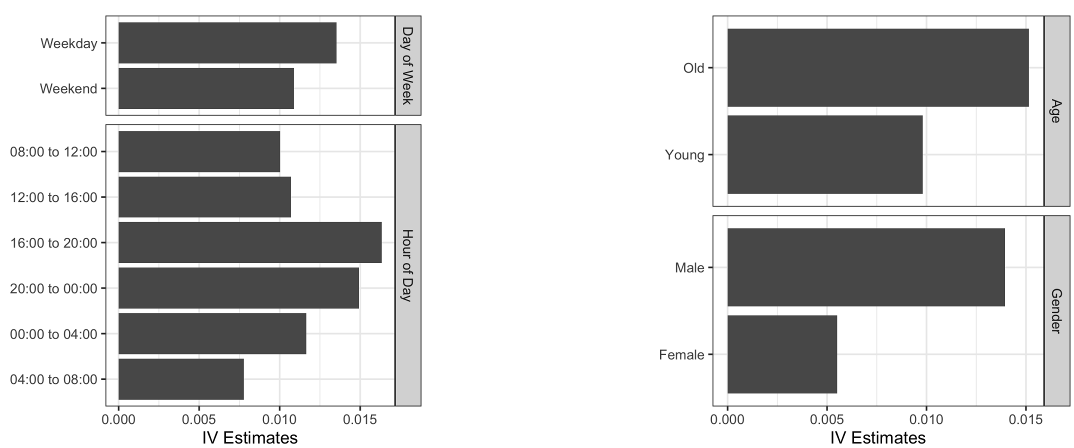
Most of this variation comes from the differences across hours of the day rather than days of the week. When performing formal statistical tests, we can strongly reject the null hypotheses of equal average responses across hours of the day and across weekends versus weekdays.



(a) Working (b) Minutes Worked

Figure 2.10: Heterogeneity in IV Estimates

Notes: In this figure, we plot the distribution of the heterogeneous treatment effects. The solid line indicates the IV estimates measured as the increase in the probability of working (or minutes worked) w.r.t. a \$10 increase in the predicted hourly wage. The estimates and the 90% bootstrap uniform confidence bands are derived following [15] based on the linear model with full saturation of observed heterogeneity in hours of the week, young versus old, and male versus female.



(a) Over Time (b) Across Demographic Groups

Figure 2.11: Heterogeneity in IV Estimates over Time and Across Demographic Groups

Notes: Figure (a) compares the IV estimates across hours of the day, and weekdays versus weekend. Figure (b) compares the IV estimates across demographic groups. The unit of the IV estimates is the increase in the probability of working w.r.t. a \$10 increase in the predicted hourly wage.

As shown in Figure 2.11b, there is also considerable heterogeneity by gender and age. On average, male and old drivers have larger responses than young and female drivers. The differences across gender are larger than those by age. Again, the null hypotheses of equal average responses by gender or age are strongly rejected in the data.

We conclude the analysis of heterogeneity by examining how much of the heterogeneity in the estimated labor supply responses can be explained or accounted for by various covariates. We begin by regressing the 1,344 IV estimates reported in Figure 2.10 on indicator variables for time. In Table 2.3, We find that day of the week explains as little as 1.2 percent of the variation in the labor supply responses. By comparison, measuring time through indicators for hours of the day increases the R-squared to 12.4 percent. There are only small gains in explanatory power from including indicators for both hour of the day and day of the week in a separable fashion. By way of comparison, interactions between hours of the day and days of the week are empirically important to explain the pattern of heterogeneity in labor supply responses, increasing the R-squared from 13.7 percent to 34.9 percent. Also including indicator variables for gender and young (defined as younger than 38) further increases the R-squared by a few percentage points. By comparison, a flexible regression model with interactions between the day of the week, the hour of the day, gender and young explains nearly 70 percent of the variation in labor supply responses. Taken together, these results suggest it is key to let preferences of the drivers vary by gender and age, and for each demographic group, across hours of the day and days of the week.

	Explanatory Power (R^2)	
Independent Variable	Working	Minutes Worked
Day of Week	0.012	0.012
Hour of Day	0.124	0.124
Day of Week + Hour of Week	0.137	0.137
Day of Week × Hour of Day	0.349	0.349
(Hour of Day × Day of Week) + (Young × Gender)	0.384	0.384
(Hour of Day × Day of Week) × (Young × Gender)	0.691	0.718

Table 2.3: Explanatory Power in Regressions of Labor Supply Responses on Covariates for Time and Demographics

Notes: We regress the IV estimates of labor supply responses on the time, age, and gender dummies, and we report the R-squared in this table. Each number in this table corresponds to a separate regression. Each regression is weighted by the inverse of the variance of the IV estimates.

Comparison with OLS estimates

Table 2.4 compares OLS estimates of the labor supply responses to the IV estimates we obtain using the experiment. These results suggest that unobserved determinants of wages, if ignored, lead to a significant downward bias in the estimated labor supply responses. In particular, the OLS estimates in Column 1 show much weaker associations between labor supply and wages than the IV estimates. This downward bias is consistent with demand being high when it is costly for the drivers to work. Including fixed effects for workers, days of the week, and hours of the day reduces the bias, as shown in Column 2. However, the labor supply elasticities remain too small. This finding suggests that idiosyncratic factors, such as weather conditions and entertainment events, may create high demand while, at the same time, make driving more costly or difficult.

Dependent Variable: Working			
	OLS	FE	IV
Estimates	0.0099	0.0125	0.0142
	(0.0003)	(0.0003)	(0.0014)
	OLS = IV	FE = IV	
p-value	0.0001	0.1143	
Dependent Variable: Minutes Worked			
	OLS	FE	IV
Estimates	0.5223	0.6296	0.6801
	(0.0157)	(0.0180)	(0.0654)
	OLS = IV	FE = IV	
p-value	0.0005	0.2066	

Table 2.4: OLS and IV Estimates of Labor Supply Responses

Notes: In this table, we report the OLS, FE, and IV estimates of the labor supply responses. OLS is estimated by regressing the outcome variables on predicted hourly wages. The standard errors for the OLS estimates are clustered at the drivers level, and the standard errors for the IV estimates and the p-values are estimated by bootstrap. The fixed effects include the driver fixed effect and the hour of week fixed effect.

2.4 Dynamic Model of Labor Supply

The experimental estimates provide key data points for learning about labor supply elasticities, reservation wages and the value of job flexibility, but do not by themselves tell us these quantities. In order to recover the labor supply elasticities and reservation wages and to infer the value of job flexibility, we now develop, identify, and estimate a dynamic model of labor supply. In this section, we present this model and discuss the parameter estimates. In Appendix 2.C, we compare these estimates to those produced by more restrictive models, including a static labor supply model. This comparison highlights how several of our modeling choices – including adjustment costs, permanent observed and unobserved heterogeneity, and the field experiment to address wage endogeneity – are key not only to match the data but also for the estimates of the reservation wages and for the results from the counterfactual analyses.

Model Setup

Driver's Problem

We model the driver as living infinitely many periods where each period is an hour. In each period t, the driver decides whether to work $a_{it} = 1$ or rest $a_{it} = 0$, taking into account both the current period payoff $U_{it}(a_{it})$ and how her choice in t will affect the payoffs in the future $\tau > t$. In each period t, a driver chooses a_{it} in order to maximize the expected sum of discounted flow payoffs:

$$\max_{a_{it}} \mathbb{E}\left[\sum_{\tau=t}^{\infty} \rho^{\tau} U_{i\tau}(a_{i\tau})|a_{it}\right]$$

where i indexes a driver, τ indexes hour, $U_{i\tau}(a_{i\tau})$ is the flow payoff associated with choice $a_{i\tau}$, and ρ is the discount rate. The expectation is taken over the future values of $U_{i\tau}(a_{i\tau})$ given the current choice a_{it} for $\tau \geq t+1$.

4.1.2 Preferences

We sort drivers into subgroups based on their age and gender, $X = (\mathbf{1}\{\text{Female}\}, 1\{\text{Young}\})$. For a driver in an observed subgroup $X = x$ who works in a given city, the flow payoff associated with action a_{it} is given by

$$U_{it} = \left\{ \begin{array}{ll} \gamma w_{it} + \beta_{h(t)} + \mu \mathbf{1}\{a_{it-1} = 0\} + \eta_{j(i),h(t)} + \xi_{it} + \epsilon_{1it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{array} \right\}$$

where i is a driver, t is a calendar hour (e.g., 2018/10/10, 9 a.m.) and $h(t)$ is an hour of a week at time t (e.g., Monday 9 a.m.). We also include city fixed effect in the empirical specification of U_{it} to allow for systematic differences in the costs of driving across cities. For notational simplicity, we suppress these fixed effects as well as the conditioning on X.

A driver's flow payoff from work depends on the wage she may earn, w_{it}, and the time-specific shifter of the cost of driving at a given hour of the week, $\beta_{h(t)}$. The empirical counterpart of w_{it} is the predicted hourly wage as described in Section 2.2. If the driver did not work at $t-1$, she needs to pay an adjustment cost to start to work, μ. The parameter $\eta_{j(i),h(t)}$ captures the unobserved type j of driver i at time $h(t)$. For example, full-time drivers who are more likely to drive at all times have higher $\eta_{j(i),h(t)}$ at all $h(t)$ than infrequent drivers, while evening drivers have higher $\eta_{j(i),h(t)}$ only in the evenings. Driver type $\eta_{j(i),h(t)}$ is known to the drivers themselves but unobserved to the analyst.

A driver's flow payoffs from the choices at t are also affected by a set of choice-specific preference shocks, ξ_{it}, ϵ_{1it} and ϵ_{0it}, which are revealed to the driver at the beginning of t. The component ξ_{it} captures the unobserved preference shocks that may correlate with wages. For example, an individual may dislike to drive in periods with heavy traffic or poor weather and these conditions may also covary with the demand for rides and thus offered wages. The components ϵ_{1it} and ϵ_{0it} capture the idiosyncratic preference shocks that are independent of wages.

The equation for the predicted hourly wage of worker i in t is specified as follows:

$$w_{it} = \delta_{h(t)} + \delta_0 m_t + \delta_1 m_t \times z_{it} + u_{it}$$
$$u_{it} \sim G_u(u_{it} | u_{it-1}, a_{it-1}, z_{it-1}, h(t-1), m_{t-1}, c(i))$$

where m_t is an indicator of whether t is an experiment hour and z_{it} is an indicator for being assigned to the treatment group in an experiment hour, $\delta_{h(t)}$ is a fixed effect for hours of week, and u_{it} represents the unobservable determinants of wages. By including m_t in the wage equation, we allow the wage to be different when a GSL experiment is switched on. The parameter δ_1 captures the exogenous change in the wage for the treated drivers during experiment hours.

The wage in our model evolves as a first-order Markov process. We allow persistence in wages by letting u_{it} depend on u_{it-1}, together with lagged choice a_{it-1}, lagged treatment z_{it-1}, lagged experiment hour m_{t-1} and city $c(i)$. We allow a driver's unobserved type $\eta_{j(i),h(t)}$ to depend nonparametrically on her initial wage and work decision. The endogeneity of wages arises if $Cov(\xi_{it}, u_{it}) \neq 0$. Thus, an exogenous wage process is a special case of our model in which the costs of working do not covary with market wages, $\delta_0 = \delta_1 = 0$.

4.1.3 Timeline and Information Set

Recall that at 4 a.m. every Monday and Friday, drivers are randomized into treatment and control groups. Drivers are then informed of when and for how long GSL will be switched on in an upcoming block. We specify the timeline within and between blocks as follows.

Timeline Between Blocks. At the beginning of each block, a driver i learns her treatment status z_{is} and the experiment hours m_s for all hours s in the block. At the same time, the driver forms expectations about her treatment status and the experiment hours in future blocks. The driver then sequentially makes labor supply decisions, taking into account the current flow payoff and the continuation values. The drivers are re-randomized at the start of the next block. Figure 2.12 presents an example of the timeline between blocks.

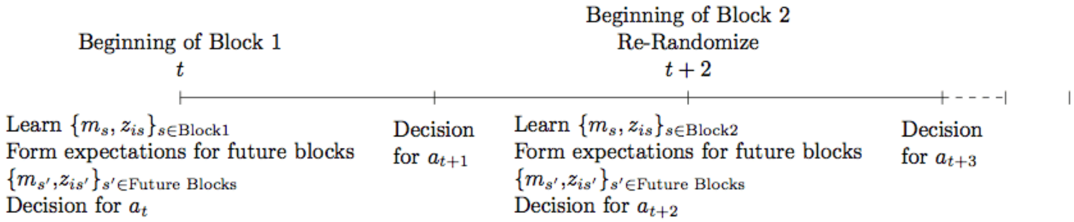


Figure 2.12: Example of Timeline with 2 Blocks and 2 Periods per Block

Timeline Within a Given Block. The decision timeline within a block is given as follows. Let t be an arbitrary period within a given block. At the beginning of t, driver i learns the realization of u_{it} (equivalently, her market wage w_{it}) and the realizations of the preference shocks $(\xi_{it}, \epsilon_{1it}, \epsilon_{0it})$. Based on these realizations, she forms expectations about future values of $u_{it'}, \xi_{it'}, \epsilon_{1it'}, \epsilon_{0it'}$. Next, she makes the work decision for period t, taking into account the current flow payoffs and how her decision at t will affect her future flow payoffs.

Assumptions and Identification

Our identification argument combines a control function based on the experiment with fairly standard assumptions in the dynamic discrete choice literature. In this section, we briefly discuss the key assumptions and the outline of the identification argument. The details are in Appendix 2.B.

The identification argument begins by making the following assumptions:

Assumption 5. *Control Function Assumption*

(Instrument Exogeneity) $z_{it} \perp (\epsilon_{0it}, \epsilon_{1it}, \xi_{it}, u_{it})$

(Joint Normality) $\begin{pmatrix} u_{it} \\ \xi_{it} \end{pmatrix} \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho_{u\xi} \\ \rho_{u\xi} & 1 \end{pmatrix})$

Under joint normality of u_{it} and ξ_{it}, we can rewrite ξ_{it} as

$$\xi_{it} = \frac{\rho_{u\xi}}{\sigma} u_{it} + \psi_{it} = \frac{\rho_{u\xi}}{\sigma}(w_{it} - \delta_{h(t)} - \delta_0 m_t - \delta_1 m_t \times z_{it}) + \psi_{it}$$

where $\psi_{it} \sim \mathcal{N}(0, 1-\rho_{u\xi}^2)$ and $\psi_{it} \perp u_{it}$ by construction. We define a new state variable $\phi_{it} \equiv w_{it} - \delta_{h(t)} - \delta_0 m_t - \delta_1 m_t \times z_{it}$. Thus, the flow payoffs of the problem become

$$U_{it} = \left\{ \begin{array}{ll} \gamma w_{it} + \beta_{h(t)} + \mu \mathbf{1}\{a_{it-1} = 0\} + \eta_{j(i),h(t)} + \frac{\rho_{u\xi}}{\sigma}\phi_{it} + \nu_{it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{array} \right\}$$

$$\begin{aligned} w_{it} &= \delta_{h(t)} + \delta_0 m_t + \delta_1 m_t \times z_{it} + u_{it} \\ u_{it} &\sim G_u(u_{it}|u_{it-1}, a_{it-1}, z_{it-1}, h(t-1), m_{t-1}, c(i)) \end{aligned}$$

where $\nu_{it} = \psi_{it} + \epsilon_{1it}$.

In addition to Assumption 5, we make the following set of assumptions which are often invoked in the literature on dynamic discrete choice (see [67], [57], [44]):

Assumption 6. *Standard dynamic discrete choice assumptions*

(IID) $\epsilon_{0it}, \epsilon_{1it}$ are iid across i, t

(EXOG) $\nu_{it} \perp (w_{it}, \eta_{j(i),h(t)}, \phi_{it})$

(CI-X) State transition probability F satisfies

$$F(w_{it+1}, \phi_{it+1}|a_{it}, a_{it-1}, w_{it}, \phi_{it}, h(t), j, \nu_{it}) = F(w_{it+1}, \phi_{it+1}|a_{it}, a_{it-1}, w_{it}, \phi_{it}, h(t))$$

(DISTR) Distributional assumption on ϵ_{0it} and ν_{it}, and independence: $\epsilon_{0it} \perp \nu_{it}$.

(DISCOUNT) ρ is known

(REACH) All states are reachable at any given time

$$F(w_{t+1}, \phi_{t+1}|a_t, a_{t-1}, w_t, \phi_t, h(t), j) > 0 \ \forall a_t, a_{t-1}, w_t, \phi_t, h(t), j$$

(NTYPE) The number of types is small.

Assumption 6 follows the standard assumptions in the dynamic discrete choice literature. The only difference is that, in our setup, the state-dependent unobserved preference shock, ξ_{it}, can be re-written as a combination of an observed state, ϕ_{it}, and an idiosyncratic component ψ_{it}, as a result of Assumption 5. Under the restrictions (IID), (EXOG), (CI-X), (DISTR), (DISCOUNT), we can identify the parameters in the flow payoff in the absence of unobserved heterogeneity (see [67], [57]). The restrictions (REACH) and (NTYPE) are made to incorporate unobserved heterogeneity in the model. Under these restrictions, the structural parameters in the flow payoffs are identified (see [41], [44]). The restriction (REACH) imposes that the entire support of wages at $t+1$ has a positive probability conditional on the state at t. In our problem, the restriction requires that at each hour of the week and conditional on the lagged choice and the wage in the previous period, a driver may get any wage in the support with a strictly positive probability. The restriction (NTYPE) limits the number of unobserved types among drivers. We allow for three unobserved types of drivers. For the restriction (DISTR), we assume ϵ_{0it} and ν_{it} are both distributed as T1EV.

It is useful to observe that some of the restrictions in Assumption 6 are not that strong in our setting. For example, the restriction (CI-X) implies that, conditional on choices, the transition probability of the state variables is the same across unobserved types and independent of transitory shocks. To understand this restriction, consider a driver who gets tired when working in period t. As a result, she may drive slower or take fewer trips and thus earn less if she chooses to continue driving in t+1. Restriction (CI-X) permits such a scenario. The restriction (DISCOUNT) requires the discount rate to be known to the analyst. In our setting, the data suggests that temporal decisions are primarily driven by heterogeneity in preferences over when to work $\beta_{h(t)}$, not discounting of future payoffs over a relatively short period of time. Thus, we think the discount rate plays a minor role for the behavior we observe. Assuming an annual interest rate as 5 percent, we set the hourly discount rate $\rho = 1/(1 + \frac{0.05}{365\times 24})$.

4.2.1 Value Function and Structural Equation

We now describe the solution to the model and the structural equation. Under the standard conditions, the driver's problem can be characterized by the Bellman equation:

$$V(s_{it}) = \max\left\{V(a_{it} = 1, s_{it}), V(a_{it} = 0, s_{it})\right\} \tag{2.1}$$

where s_{it} is a vector of observed state variables and the unobserved type, $V(s_{it})$ is the (ex-ante) value function for driver i who is in state s_{it}, and the choice-specific conditional value functions $V(a_{it} = 1, s_{it})$ and $V(a_{it} = 0, s_{it})$ are defined as follows:

$$V(a_{it} = 1, s_{it}) = s_{it}'\theta + \xi_{it} + \epsilon_{1it} + \rho\mathbb{E}V(s_{it+1})$$

$$V(a_{it} = 0, s_{it}) = \epsilon_{0it} + \rho\mathbb{E}V(s_{it+1})$$

where $\theta = (\gamma, \beta_{h(t)}, \mu, \eta_{j(i),h(t)})$ is a vector of structural parameters in the flow payoff of work. The expectation is taken over future states and actions, $s_{i\tau}$ and $a_{i\tau}$ $\forall\tau \geq t+1$, as well as future preference shocks $(\epsilon_{0i\tau}, \epsilon_{1i\tau}, \xi_{i\tau})$ $\forall\tau \geq t+1$, conditional on s_{it} and a_{it}, according to the evolution of states $F(s_{it+1}|s_{it}, a_{it})$ and the distribution of shocks. Thus, with equation (2.1), we can describe the driver's decision rule as follows: At the beginning of each period t, driver i learns state s_{it}, and chooses to work if and only if $V(a_{it} = 1, s_{it}) \geq V(a_{it} = 0, s_{it})$.

4.2.2 Estimation

The first step in our estimation procedure is to use OLS to estimate the parameters of the wage equation, $\delta_{h(t)}$, δ_0, δ_1, and then to obtain the empirical counterpart of u_{it} as measured by the residuals. We then non-parametrically estimate the transition probability of u_{it} given lagged state variables $(u_{it-1}, a_{it-1}, z_{it-1}, h(t-1), m_{t-1})$.

We follow the two-stage estimator of Arcidiacono and Miller (2011) to estimate the model with unobserved types by maximizing the log likelihood of the finite mixture model:

$$\{\hat{\theta}, \hat{\pi}\} = arg\,max_{\theta,\pi} \sum_{i=1}^{N} \ln[\sum_{j=1}^{J} \pi_j(s_{i1}) \prod_{t=1}^{T_i} l(a_{it}|s_{it}, j, \hat{p}, \theta)] \tag{2.2}$$

where $s_{it} = (w_{it}, \phi_{it}, a_{it-1}, h(t))$ is the vector of observed states, $\hat{p}$ is a vector of empirical conditional choice probabilities, $\pi(s_{i1})$ is the population probability of type j conditional on initial state s_{i1}, and θ is a vector of the model parameters. The number of unobserved types, J, is assumed to be known and we set $J = 3$. Let $l(a_{it}|s_{it}, j, \hat{p}, \theta)$ denote the likelihood contribution of driver i at time t. We can express the likelihood as follows:

$$l(a_{it}|s_{it}, j, \hat{p}, \theta) = \frac{a_{it}e^{[s_{it}+\rho(\tilde{s}(a=1,s_{it})-\tilde{s}(a=0,s_{it}))]'\theta+\rho(\tilde{e}(a=1,s_{it})-\tilde{e}(a=0,s_{it}))} + (1 - a_{it})}{1 + e^{[s_{it}+\rho(\tilde{s}(a=1,s_{it})-\tilde{s}(a=0,s_{it}))]'\theta+\rho(\tilde{e}(a=1,s_{it})-\tilde{e}(a=0,s_{it}))}}$$

where $\tilde{s}(a, s)$ and $\tilde{e}(a, s)$ are known functions of the state s, the conditional choice probabilities $P(a|s)$, and the state transition probabilities $F_s(s|a, s)$. We first estimate the empirical counterparts of the conditional choice probabilities and the transition probabilities of the state variables. Next, we initialize $\hat{\theta}, \hat{\pi}_j(s), \hat{q}_{ij}$ for all $\forall i, j$ by estimating the dynamic model without unobserved types, and set $\hat{\pi}_j(s_1)$ equal to $\hat{\pi}_j(w_1, a_0)$. We then update $\hat{\pi}(s), \hat{q}_{ij}$, $\hat{p}(s, j)$ and $\hat{\theta}$ for all i, j based on the EM algorithm. We refer to Appendix 2.B for further details about the estimation procedure.

Model Fit and Estimation Results

Before we present the estimation results, we examine how well our estimated model fits the data. To examine the fit of the model, we focus on the probability of working conditional on the treatment status and other state variables.

Figure 2.13 plots the probability of working by treatment status across hours of the week, days of the week, and hours of the day. We integrate the model predicted probabilities across all states, except for treatment status and time. The model predicted probabilities fit the data counterparts very well.

In addition to the working pattern by treatment status across time, we also examine the fit of the model by treatment status across demographic groups and unobserved types of drivers. As shown in Appendix 2.E.1, our model predicts well the working patterns across these dimensions.

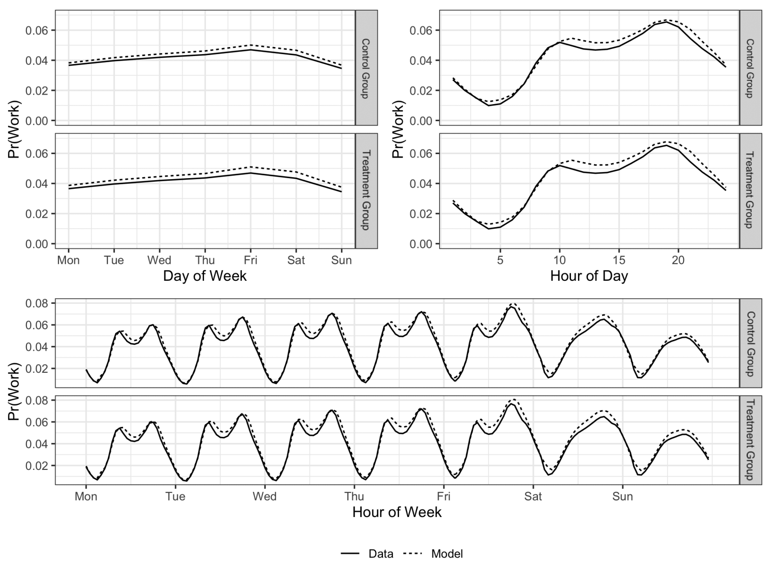


Figure 2.13: Model Fit of Probability of Working by the Treatment Status over Time
Notes: The solid line plots the data, and the dashed line plots the prediction from the model.

Our model also captures the important dynamic component, lagged choices, in a driver's labor supply decision. Figure 2.14 shows the probability of working conditional on the lagged work decision, integrated over the observed heterogeneity and all other states. The probability of working differs significantly depending on whether a driver worked or not in the previous period. Through the lens of the model, this difference produces sizable costs of starting to drive.

The dynamic component is captured by fixed costs of starting to drive in our model.

However, one might worry that there are other important sources of adjustment costs. For example, [32] and [33] argue that fatigue is important to understand the behavior of taxi drivers. The reason is that the probability that a taxi driver ends a shift depends strongly on hours worked. In the Uber setting, there is little, if any, evidence of such dependence, as evident from Figure 2.26. Given the weak relationship between the probability of stopping to drive and cumulated hours worked, we decided against including fatigue in the model. Arguably, a small improvement in model fit does not justify to further enlarge the state space, which would significantly increase the computational costs of solving the model.

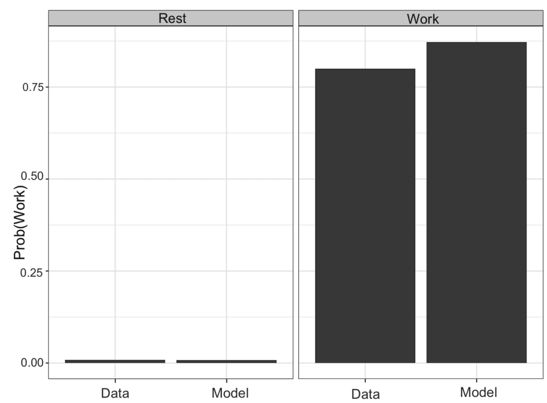


Figure 2.14: Model Fit of Probability of Working by Lagged Work Decisions

Notes: In this figure, the "Rest" panel shows the probability of working conditional on drivers not working in the previous hour. The "Work" panel shows the probability of working conditional on drivers working in the previous hour.

Estimation Results. In Table 2.5, we present the parameter estimates. The estimate of γ, which captures the drivers' sensitivity to wage changes, are positive across the four subgroups. However, the magnitudes vary across the groups. Male drivers are more responsive to exogenous wage changes than female drivers. Consider, for instance, young male drivers. All else being equal, a 1% increase in market wages induces young male drivers to increase their probability of working by around 0.75%.

Our estimation shows considerable dispersion in the value of time $\beta_{h(t)}$ across hours of the week for each of the four subgroups. Recall that a large $\beta_{h(t)}$ in absolute value corresponds to a high cost of driving at $h(t)$, and consequently a low probability of working, all else being equal. On average, the absolute value of the estimated $\beta_{h(t)}$ at the 90th percentile is twice as large as the absolute value of the estimated $\beta_{h(t)}$ at the 10th percentile.

We also find significant adjustment costs in our model as captured by the estimate of μ. For example: Among young male drivers at 8:00 a.m. on Monday, those who worked at 7:00 a.m. have a predicted probability of working around 0.86, while the probability is only 0.07 for those who did not work at 7:00 a.m. This state-dependency emphasizes the importance of incorporating the dynamic component in the driver's decision problem.

Our model allows for three unobserved types among drivers. The type parameter, η, shifts the cutoff in the work choice equation. We normalize η to 0 at all hours of the week for drivers of the baseline type. Thus, high η shifts up the value of work and increases the probability of working relative to the baseline type.

Our estimates of η's suggest three types of drivers. One type is likely to drive in the evening. Another type drives frequently both during the day and at night. The third type, the baseline type, drives infrequently. Evening drivers ($\eta_{1,Night}, \eta_{1,Day}$) have a much higher cutoff of work at night than frequent drivers ($\eta_{2,Night}, \eta_{2,Day}$). Across all the four demographic groups of drivers, the estimated $\eta_{1,Night}$ is almost twice as large as the estimate of $\eta_{2,Night}$. Consider, for example, young male drivers. Conditional on not working in the previous period, the model predicts evening drivers have a probability of working as high as 0.08 at midnight 12:00 a.m., whereas frequent drivers' work probability is only 0.01 and infrequent drivers' work probability is close to zero.

			Weighted Average	Old Male	Young Male	Old Female	Young Female
Preference for Wage	γ		0.036	0.040	0.042	0.024	0.021
Time Preferences	β	$E[\beta_{h(t)}]$	-1.544	-1.612	-1.760	-1.257	-0.977
		$Sd(\beta_{h(t)})$	0.513	0.516	0.462	0.597	0.570
		Median($\beta_{h(t)}$)	-1.328	-1.375	-1.598	-0.972	-0.734
		$q_{10}(\beta_{h(t)})$	-2.295	-2.428	-2.351	-2.164	-1.861
		$q_{90}(\beta_{h(t)})$	-1.055	-1.151	-1.301	-0.653	-0.432
Adjustment Cost	μ		-6.377	-6.191	-6.269	-6.771	-6.864
Unobserved Types	η						
		$\eta_{(1,Night)}$	1.857	1.847	1.901	2.039	1.571
		$\eta_{(1,Day)}$	0.601	0.589	0.669	0.547	0.483
		$\eta_{(2,Night)}$	1.178	1.120	1.237	1.288	1.067
		$\eta_{(2,Day)}$	0.794	0.773	0.802	0.838	0.788
Selection Term	$\frac{\rho_{u\xi}}{\sigma}$		-0.039	-0.043	-0.044	-0.028	-0.024

Table 2.5: Estimates of Model Parameters

Notes: "Weighted average" is calculated by averaging the estimates of the four demographic groups weighted by the share of the drivers. "Young" is defined as those whose ages are less than or equal to the median age.

Table 2.5 also reveals that the estimates of the correction term $\frac{\rho_{u\xi}}{\sigma}$ are negative in all four demographic groups. Recall that $\rho_{u\xi}$ is the correlation coefficient between the preference shock ξ_{it} and the wage component u_{it}. Our estimation results indicate that the costs of working tend to co-move with the market wages. As shown in Appendix 2.C, it is important to take this endogeneity into account to obtain reliable estimates of the preference parameters.

2.5 Insights from the Model

We now use the estimated model to compute the labor supply elasticities, the reservation wages as well as to perform counterfactual analyses. These counterfactuals allow us to infer the drivers' willingness to pay for the ability to customize and adjust their work schedule.[10]

Labor Supply Elasticities

To interpret the magnitude of the preference parameters, we use our estimated model to calculate two types of labor supply elasticities. The first is the Frisch elasticity for the labor supply decision of whether to drive in a given hour of the week. This extensive margin Frisch elasticity can be defined in our model as the percent change in the probability of working in a given hour of the week for an anticipated and temporary one percent exogenous increase in the hourly wage. Formally, we follow [24] and define the Frisch labor supply elasticity on the extensive margin per hour as:

$$
\begin{aligned}
\delta^F \equiv& \frac{\log(\partial Pr(a_{it}=1|s_{it},w_{it}))}{\partial \log(w_{it})} \\
=& \frac{\partial(\Delta V(s_{it},w_{it}) - (\nu_{it}-\epsilon_{0it}))}{\partial w_{it}} f(\Delta V(s_{it},w_{it}) - (\nu_{it}-\epsilon_{0it})) \frac{w_{it}}{Pr(a_{it}=1|s_{it},w_{it})} \\
=& \gamma f(\Delta V(s_{it},w_{it}) - (\nu_{it}-\epsilon_{0it})) \frac{w_{it}}{Pr(a_{it}=1|s_{it},w_{it})}
\end{aligned}
\tag{2.3}
$$

where $\Delta V(s_{it},w_{it}) \equiv V(a_{it}=1,s_{it},w_{it}) - V(a_{it}=0,s_{it},w_{it})$ is the difference between the values of work and rest, $V(\cdot)$ is the value function, s_{it} is the vector of state variables excluding w_{it}, ν_{it} and ϵ_{0it} are idiosyncratic shocks, and f is the probability density function of the choice which follows the logistic distribution. To calculate the elasticity, we first use the

10. Throughout the counterfactual analyses, we abstract from how the market wages may be affected by changes in the labor supply of the drivers. Taking into account such effects would require data on and a model of the demand side of the market.

estimated model parameters to recover the value function V. With the value functions, we then compute the elasticities evaluated for every possible realization of the state variables, using equation (2.3). We average the resulting elasticities weighted by the share of the states in the data and report them in Table 2.6.

Our Frisch labor supply elasticities range from 0.36 to 0.83, with a weighted average of 0.65. On average, male drivers have higher labor supply elasticities than female drivers. Even conditional on observables, there is substantial variation in the elasticity: Infrequent and frequent drivers have the highest elasticities, evening drivers the lowest. By way of comparison, our model implies much smaller differences in the labor supply elasticity over time than across drivers.

		Weighted Average	Daytime	Evening	Weekday	Weekend
				Frisch elasticity		
Observed heterogeneity	Old male	0.71	0.67	0.78	0.68	0.76
	Young male	0.76	0.72	0.83	0.74	0.82
	Old female	0.43	0.41	0.48	0.42	0.47
	Young female	0.38	0.36	0.41	0.36	0.41
Unobserved heterogeneity	Frequent	0.65	0.61	0.74	0.63	0.71
	Evening	0.52	0.54	0.49	0.52	0.54
	Infrequent	0.65	0.62	0.70	0.63	0.70
				IES		
Observed heterogeneity	Old male	0.47	0.42	0.56	0.45	0.50
	Young male	0.52	0.47	0.60	0.50	0.56
	Old female	0.30	0.27	0.37	0.29	0.33
	Young female	0.26	0.23	0.32	0.25	0.28
Unobserved heterogeneity	Frequent	0.35	0.32	0.43	0.34	0.38
	Evening	0.28	0.31	0.22	0.28	0.29
	Infrequent	0.48	0.44	0.58	0.47	0.52

Table 2.6: Model Implied Extensive Margin Labor Supply Elasticities

Notes: "Weighted Average" is calculated by averaging the estimates of the four demographic groups weighted by the share of the drivers. Young is defined as age less than or equal to the median age.

The other elasticity we compute is the intertemporal elasticity of substitution (IES). The IES between two periods, t and $\tilde{t}$, can be defined as follows:

$$\delta^{IES}_{t,\tilde{t}} \equiv \frac{\partial \log(\frac{Pr(a_{it}=1|s_{it},w_{it})}{Pr(a_{i\tilde{t}}=1|s_{it},w_{it})})}{\partial \log(\frac{w_{it}}{w_{i\tilde{t}}})} \qquad \text{for } \tilde{t} > t$$

and measures how much a driver is willing to substitute work between t and $\tilde{t}$ for changes in the relative wage $\frac{w_{it}}{w_{i\tilde{t}}}$. As shown in Appendix 2.E.2, the following expression shows the close link between the IES and the Frisch elasticity:

$$\delta^{IES}_{t,\tilde{t}} = \delta^F - \int \delta^F_{\tilde{w}} g(w_{i\tilde{t}} = \tilde{w}, s_{i\tilde{t}} = \tilde{s}, s_{it}, w_{it}; \mu) d(\tilde{w}, \tilde{s}) \qquad \text{for } \tilde{t} > t \tag{2.4}$$

where $\delta^F_{\tilde{w}} \equiv \frac{\partial \log(P(a_{i\tilde{t}-1}=1|w_{i\tilde{t}}=\tilde{w}, s_{i\tilde{t}}=\tilde{s}, s_{it}, w_{it}))}{\partial \log(w_{it})}$ is the Frisch elasticity of labor supply in $\tilde{t}-1$ if drivers had perfect foresight of all states in $\tilde{t}-1$, and $g(\cdot;\mu)$ is a known increasing function in the adjustment cost μ. The magnitude of the gap between IES and the Frisch elasticity is governed by the adjustment cost μ. When $\mu = 0$, the function $g(\cdot;\mu)$ becomes 0, and hence the Frisch elasticity and the IES coincide.

To calculate IES, we use the estimated model parameters to recover δ^F and $g(\cdot;\mu)$, and then calculate the IES using equation (2.4). The results are presented in Table 2.6. Our IES estimates range from 0.22 to 0.60, with a weighted average of 0.45. The IES estimate is about 30% smaller than the Frisch elasticities, suggesting a considerable size of the adjustment cost.

Value of Time and Reservation Wages

A key objective of the model is to recover the reservation wages and study how they vary over time and across people. We start with presenting the value of time, $\beta_{h(t)}$, which is an important component of the reservation wages.

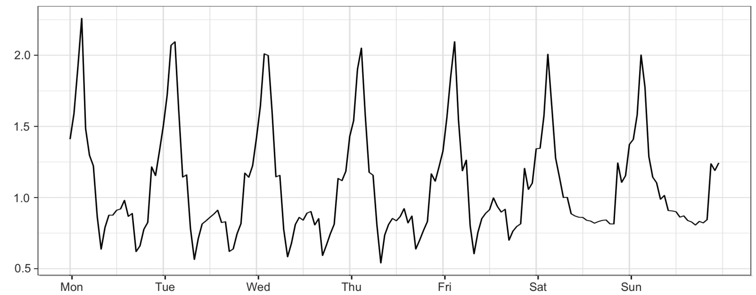


Figure 2.15: Variation in $\beta_{h(t)}$ Relative to Saturday 8 a.m. over Time

Notes: We compute for each hour of the week the weighted average of $\beta_{h(t)}/\beta_{Saturday8a.m.}$ using population shares of each demographic group.

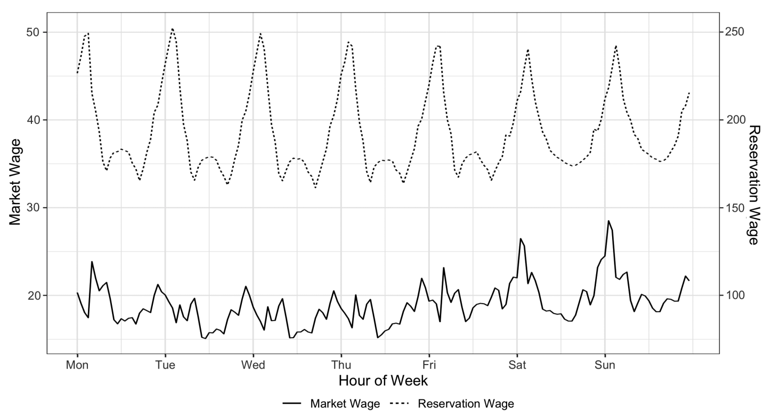


Figure 2.16: Comparison of Observed Market Wages and Reservation Wages

Notes: In this figure, we compute the reservation wage as the minimal wage needed to work given mean preference shocks and averaged across the state variables. The y-axis on the left represents the scale of the market wages, while the y-axis on the right represents the scale of the reservation wages.

Figure 2.15 plots the weighted average of $\beta_{h(t)}/\beta_{\text{Saturday8a.m.}}$ across the four demographic subgroups. Our findings reveal that the value of time varies systematically during a typical week, with the value peaking at late nights around 4 a.m. As shown in Appendix 2.E.3, the value of time varies a lot across hours within a day, whereas there is little variation in the value of time across weekdays. In Figure 2.32 in the Appendix, we find that gender is the key dimension of observable heterogeneity when it comes to the value of time.

Knowledge of the value of time $\beta_{h(t)}$ is necessary but not sufficient to draw inference about reservation wages. We also need to take into account the unobserved preference component ξ_{it}, which may correlate with market wages. In Figure 2.16, we compute and plot the reservation wages against the expected market wages for an average driver during a typical week. The reservation wages stay low during the day, start to increase in the late evening, peak around 4 a.m., and then gradually decline until 9 a.m. The reservation wages are slightly higher during weekends compared to weekdays. While the market wage and the reservation wage tend to co-move across hours, the levels differ significantly. In particular, the average reservation wage is always considerably higher than the average market wage. This finding explains why no more than 4 percent of the drivers are choosing to work in an average hour during the week.

Panel A: Observed Heterogeneity

	Daytime	Evening	Weekday	Weekend
Old male drivers	1.00	1.28	1.09	1.11
Young male drivers	1.08	1.28	1.14	1.16
Old female drivers	1.93	2.42	2.08	2.11
Young female drivers	2.35	2.88	2.51	2.57

Panel B: Unobserved Types

	Daytime	Evening	Weekday	Weekend
Frequent drivers	1.00	1.54	1.17	1.20
Evening drivers	1.26	1.12	1.24	1.15
Infrequent drivers	2.33	2.73	2.45	2.49

Table 2.7: Reservation Wages by Driver Types and Time

Notes: In Panel A, the reservation wages are normalized by the mean reservation wage of old male drivers during the daytime. In Panel B, the reservation wages are normalized by the mean reservation wage of frequent drivers during the daytime. The daytime is defined as 6 a.m.-9 p.m., and the evening is defined as 10 p.m.-5 a.m.

The reservation wages vary not only over time but also across people. Panel A of Table 2.7 presents the reservation wages across the four demographic groups and Panel B presents the reservation wages across the three types of drivers. We normalize the reservation wages by the daytime reservation wage of older male drivers. On average, the female drivers' reservation wages are twice as high as the male drivers' reservation wages. Furthermore, young drivers tend to have slightly higher reservation wages than older drivers. There is also a lot of heterogeneity across drivers conditional on age and gender. The frequent drivers have half as large reservation wages as compared to the infrequent drivers, and the evening drivers have lower reservation wages in the evenings as compared to the daytime.

Value of the Ability to Set Customized Work Schedules

Another key objective of the model is to infer the value of job flexibility. One important source of job flexibility is the driver's ability to set a customized work schedule, so she can plan when to work based on her expectations about reservation wages relative to market wages. We quantify the value of this job flexibility by restricting the driver's ability to set the work schedule. In particular, we remove certain hours of a day or days of a week from the choice set of a driver, and then solve the driver's problem given this restricted choice set.

For computational reasons, it is useful to consider a situation where each person drives a given number of consecutive hours per week. In our counterfactual analyses, we let each driver work 5 consecutive hours per week. This choice matches the average number of work hours among the drivers in our estimation sample. At the beginning of a week, each driver is required to choose one 5-hour block to work. We consider two scenarios: The benchmark case and the restricted case. In the benchmark case, drivers can choose among all possible 5-hour blocks. In the restricted case, we remove certain blocks from the choice set. The only difference between the benchmark case and the restricted case is the restriction on the drivers' choice set. Thus, by comparing a driver's utility in these two cases, we can calculate the wage multiplier that she would need to accept a restricted choice set.

In Figure 2.17, we plot the average wage multipliers that the drivers would demand to accept various restrictions on the choice set. In the left panel, we remove the drivers' preferred hours from their choice set. In the calculations behind the first bar, we remove the favorite 5-hour block. The resulting wage multiplier is 1.05. In other words, the average worker would require 5 percent higher wages to accept such a restriction on the choice set. In the second bar, we remove the entire day containing the favorite 5 hours from the driver's choice set. The wage multiplier barely increases. In the third bar, we remove the preferred 5 hours from each day of the week. The wage multiplier now exceeds 1.1.

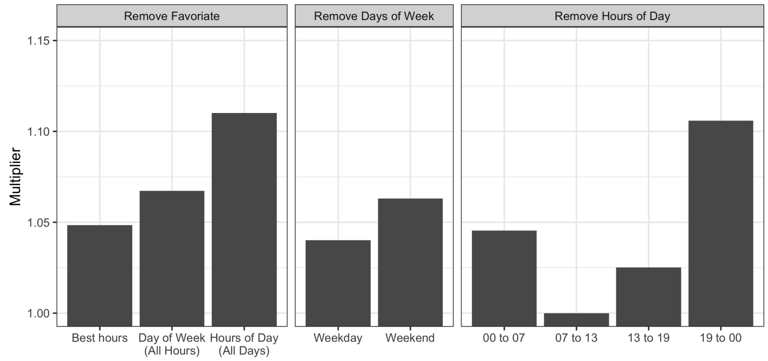


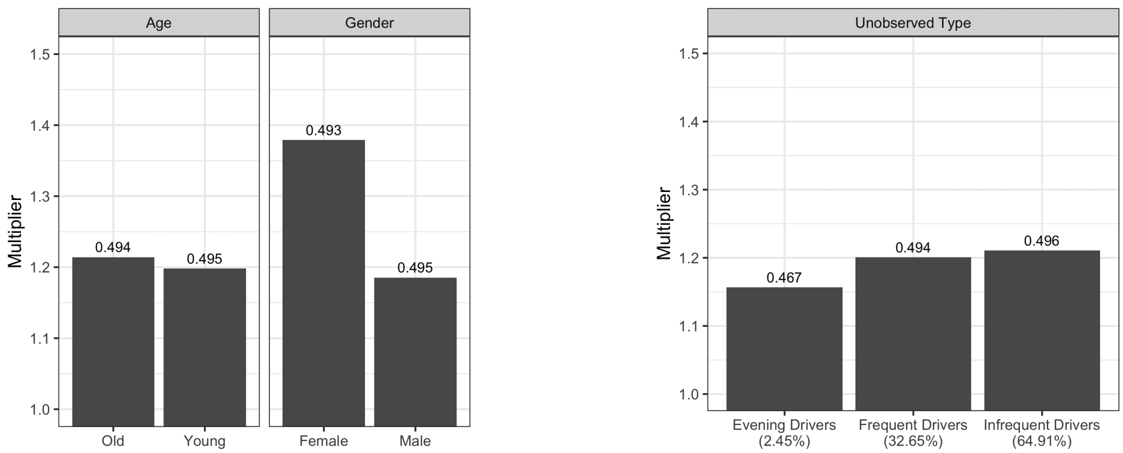
Figure 2.17: Wage Multipliers to Accept Restrictions on the Choice Set

Notes: The left panel removes each drivers' favorite 5-hour block from their choice sets. The "Best hours" indicates the best 5-hour block of the week for the drivers. The middle panel removes weekdays or weekends from drivers' choice sets. The right panel removes certain 5-hour blocks across the entire week from drivers' choice sets.

In the second and the third panel, we restrict the choice set of all drivers to certain days of a week or certain hours of a day. We find that removing the morning block (7 a.m. to 1 p.m.) results in a small wage multiplier. This is because relatively few drivers prefer to work during this time of the day. In contrast, the drivers would demand a large wage multiplier if the evening block (7 p.m. to midnight) would be removed from the choice set. In Figure 2.34 in the Appendix, we show that young drivers and female drivers require relatively high multipliers. This suggests that the ability to set customized work schedules is more important for these groups. We also show in the same figure that frequent drivers and especially evening drivers would require large multipliers to accept restrictions on the possibility of driving in the evening and at night.

Value of the Ability to Adjust Work Schedules

Another source of job flexibility comes from the drivers' ability to adjust work schedules from day to day or even hour to hour in response to unanticipated changes to market wages or the costs of driving. We quantify the value of this job flexibility by forcing the driver to commit to a work schedule before observing the realizations of the innovations to wages and preferences. In particular, we compare the expected values of the two distinct types of work arrangements. The first is a commitment scheme where each driver commits to working the 5-hour block that gives her the highest expected utility. The alternative is a flexible scheme in which each driver is allowed to adjust their choice of 5-hour block once she observes the realizations of the innovations to wages and preferences. By comparing a driver's utility in these two cases, we can calculate the wage multiplier that she would need to accept the commitment scheme instead of the flexible scheme.



(a) Demographic Groups

(b) Unobserved Types

Figure 2.18: Wage Multipliers to Accept the Commitment Instead of the Flexible Scheme

Notes: In Figure (a), we compute the weighted average of the wage multipliers needed for drivers in each demographic group to be indifferent between the flexible and the commitment scheme. In Figure (b), we compute the wage multipliers needed for drivers of each unobserved type to be indifferent between the flexible and the commitment scheme. The estimates above the bars show the fraction of drivers that will switch to the second-best 5-hour block.

In Figure 2.18, we report the average wage multipliers that the drivers would demand to accept the commitment scheme. For now, each driver is only allowed to adjust her work schedule once per week in the flexible scheme. On average, about half of the drivers would use this flexibility and change their work schedule due to preference or wage shocks. The wage multiplier needed for drivers to accept the commitment scheme is 1.21. In other words, the average worker would require 21 percent higher wages to prefer the commitment scheme over the flexible scheme. Female drivers place a higher value on the flexibility of adjustment to shocks than male drivers; they need a multiplier of 1.38, while male drivers only require 1.19.

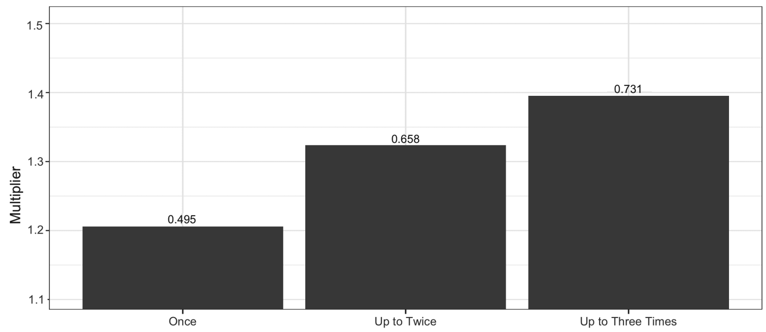


Figure 2.19: Wage Multipliers to Accept the Commitment Scheme Instead of the Flexible Schemes Where Drivers Can Adjust Work Schedules Once, Twice or Three Times

Notes: We compute the weighted average of the wage multipliers needed for drivers to be indifferent between the commitment scheme and the three types of flexible schemes where drivers are allowed to adjust to shocks once, twice, or three times. The estimates above the bars show the fraction of drivers that will switch away from the first best window.

Next, we examine how the wage multipliers change as we increase the number of times drivers are allowed to adjust to shocks. In this analysis, we compare the commitment scheme to flexible schemes where drivers are allowed to adjust once, or twice, or three times within the next week. The results are presented in Figure 2.19. The wage multiplier approaches 1.4 when we compare the commitment scheme to a flexible scheme where each driver can make three adjustments to her weekly work schedule as she observes the realizations of her wage and preference shocks.

Ability to Both Customize and Adjust Work Schedules

In Figure 2.35 and 2.36, we quantify the value of the ability to both customize and adjust work schedules. To this end, we consider three types of work schedules: (1) a fully fixed work schedule where drivers are assigned a particular work schedule and cannot adjust, (2) a commitment scheme where workers can customize but not adjust work schedules, and (3) a flexible scheme where workers can both customize and adjust work schedules. Comparing (3) to (1) allows us to infer the total value of the ability to both customize and adjust work schedules, while the comparison between (3) and (2) isolates the value of the ability to adjust work schedules.

In Figure 2.35, we show that the total value of the ability to both customize and adjust work schedules corresponds to a 43 percent increase in wages. About half of this wage increase can be attributed to the value of the ability to adjust work schedules. Figure 2.36 reveals that female and evening drivers place a particularly high value on work flexibility. In Figure 2.37, we further examine how the wage multipliers change as we increase the number of times drivers are allowed to adjust in (3). The wage multiplier exceeds 1.45 when we compare the commitment scheme to a flexible scheme where each driver can make many adjustments to her weekly work schedule as she observes the realizations of her wage and preference shocks.

Substitution Possibility to Lyft

In the period and cities we consider, Uber has a large majority of the U.S. consumer ride-sharing market. Nevertheless, one might be worried that our estimates are affected by substitution between Lyft and Uber.

To assess this, we perform two sets of analyses. First, we examine if our findings differ systematically in cities or in periods in which Lyft has relatively high or low market share. Second, we compare the results for all drivers to those we obtain from a subsample of drivers who are ineligible to drive for Lyft. The results are reported in Appendix 2.D. In each set of analyses, we find that neither the experimental estimates nor the estimated model parameters depend materially on the possibility of substitution between Lyft and Uber.

2.6 Conclusion

Over the years labor markets have varied dramatically in both their flexibility for where and when the agent works. Just considering the U.S. in the past 75 years, World War II pushed more than 9 million women who previously worked at home to enter factory and other office positions. Work life was a fixed daily and hourly routine completing inflexible tasks. The past decade has witnessed a shift back, with the Gig Economy providing much higher levels of flexibility, autonomy, and task variety. We leverage a natural field experiment at Uber to quantify how reservation wages and labor supply elasticities vary between people and over time, and to infer workers' valuation of flexibility in their choice of work hours. Economists and policymakers are keenly interested in these quantities, especially lately with the growth in jobs that offer flexible work schedules. Combining the experiment with high frequency panel data on wages and work decisions, we documented how labor supply responds to exogenous changes in offered wages in a setting with no restrictions on hours choices stemming from the demand side of the market. We found evidence of systematic heterogeneity in labor supply responses between people and over time, significant fixed costs to starting to drive, and high demand when it is costly for drivers to work.

These experimental findings motivated a dynamic model of labor supply with flexible heterogeneity in preferences over work schedules, start up costs, and the correlation between offered wages and costs of driving in a given period. The primitives of the model were recovered from a combination of the experimental estimates and other data moments. We used the estimated model to compute how labor supply elasticities and reservation wages vary between people and over time, and to perform counterfactual analyses. These analyses allowed us to infer drivers' willingness to pay for the ability to customize and adjust their work schedule.

2.A Wage Prediction Model

We now describe how we construct the expected market wages. As described in Section 2.3, a driver's expected hourly wage is the product of two components: a wage multiplier and the pre-multiplier wage. In the subsections below, we describe each component in detail.

2.A.1 Wage Multipliers

The wage multiplier augments a driver's baseline compensation for driving. It is the maximum of the surge level and the GSL experimental wage multiplier, i.e. max(GSL, Surge). However, since both GSL and surge levels vary over time and across places, this complicates how we construct the wage multipliers.

We illustrate how GSLs vary at the time×place level and how surge levels could limit the experimental variation in GSLs with Table 2.8. The table provides a simple example of a GSL menu for drivers in Chicago which was divided into five regions during most of our sample period: North, West, Loop, South, and Others.

Across all five regions, wherever the GSL experiment is switched on, the treated drivers receive 0.1 higher GSL multipliers than the control group drivers. However, the treatment effect on the wage multiplier is not always 0.1 for two reasons. First, the wage multiplier takes on the maximum value of GSL and the surge level for a given region and time. In regions where the surge levels are higher than the treatment group's GSL level, all drivers in the region receive the same wage multiplier (i.e., the surge level) regardless of their treatment status. Second, there is spatial variation in GSLs and surge levels so GSLs may be higher than surge levels in some regions but not in the others. As a result, we compute the average of the maximum of GSL and surge levels for each driver, weight them by the driver's fraction of time driving in each region in the past, and use this weighted average to predict the driver-specific wage multiplier at every hour.

	Region in Chicago				
	South	North	Loop	West	Others
Treatment	1.5×	1.3×	1.0×	1.1×	1.0×
Control	1.4×	1.2×	1.0×	1.0×	1.0×

Table 2.8: GSL Example in Chicago Across Regions

Notes: Within each region, the treatment group has a GSL multiplier 0.1× higher than the control group.

2.A.2 Pre-multiplier Wages

We calculate the pre-multiplier wages as the observed hourly wages divided by the calculated wage multipliers. We then fit the following regression model to the panel data on the pre-multiplier wages:

$$\tilde{W}_{it} = \alpha_i + \kappa_{h(t)} + \epsilon_{it}$$

where t is an hour, $h(t)$ is the hour of the week at t, i denotes a driver, α_i and $\kappa_{h(t)}$ are driver and hour-of-week fixed effects, and $\tilde{W}$ is the pre-multiplier hourly wage. For each hour t, we fit the model with the panel data up to $t-1$, and use the estimated fixed effects to compute a predicted value for $\tilde{W}$ for every worker at every hour. The predicted hourly wage, $\hat{w}_{it}$, is then constructed as the product of the wage multiplier and the predicted pre-multiplier wage.

2.A.3 Other Prediction Model and Cross-Validation

To assess how well our prediction model performs, we compare it to alternative approaches using a cross-validation procedure. This procedure repeatedly divides samples into a training sample and a testing sample. For each approach, we use the training samples to estimate a prediction model, and then we use the estimated model to form a prediction and calculate the out-of-sample mean squared errors (MSE) on the testing samples. Formally, MSE is defined as the squared distance between the predicted hourly wage rate, $\hat{w}_{it}$, and the observed hourly wage rate w_{it}:

$$MSE = \frac{1}{N_1} \sum_{t=1}^{T} \sum_{i}^{N} (\hat{w}_{it} - w_{it})^2$$

where N_1 is the number of observed hourly wage w_{it}. The smaller the MSE is, the better the prediction model performs.

In addition to our current prediction model, we consider several alternative approaches with matching models. In these models, we consider two sets of observables for matching. The first set of observables consists of demographic variables, including gender and age. The second set of observables includes drivers' past driving histories.

For each set of observables in the matching models, we consider two ways to form the prediction. The first way uses the mean of the observed hourly wages in the previous week. For example, we predict young male drivers' hourly wages at 9 a.m. on a Monday by their mean hourly wages on the Monday 9 a.m. in the previous week. The second way of prediction uses the rolling average of the observed hourly wages. For example, we take average of young male drivers' observed hourly wages at 9 a.m. on all past Mondays to form the wage prediction.

One challenge in this procedure is that a driver's driving history is high-dimensional. Therefore, we conduct dimensionality reduction on driving histories. We consider two types of drivers' work histories, the histories on observed wages and the histories on minutes worked. We first obtain a 168-dimensional vector for each driver by averaging observed wages (or minutes worked across) the past weeks. Each element in the 168-dimensional vector corresponds to a driver's average wages (or minutes worked) in an hour of the week. Next, we cluster drivers based on these 168-dimensional vectors by the K-means clustering algorithm. Formally, let $a_{it} = \frac{1}{t} \sum_{s=1}^{t} a_{is}$ be a vector representing the average observed wages (or minutes worked) for each hour of the week up to t for driver i. Driver groups are

Specification		MSE
Fixed Effect Model		
Fixed Effects		0.067
Matching Model		
Var. Matched On	*Prediction From*	
Age × Gender	Last Week	0.077
Age × Gender × History of Wage	Up to Last Week	0.088
Age × Gender × History of Wage	Last Week	0.067
Age × Gender × History of Minutes Worked	Up to Last Week	0.077
Age × Gender × History of Minutes Worked	Last Week	0.088

Table 2.9: Cross-Validation

Notes: MSE refers to the mean squared error of each prediction model. "Prediction From" indicates the periods used to form wage predictions.

obtained by solving the following:

$$arg\min_{k_1,...,k_N}\sum_{i=1}^{N}||a_{it-1}-\bar{a}_{kt-1}||_2$$

where $i \in \{1, ..., N\}$ denotes a driver, $k_i \in \{1, ..., K\}$ denotes driver i's group, and $\bar{a}_{k_i t-1}$ denotes the mean of a_{it-1} of the group k_i.

We present the results in Table 2.9. The demographic groups we consider include age and female dummies. The driving history includes drivers' histories on observed hourly wages and histories on minutes worked. Our fixed effect wage prediction model outperforms most matching models in terms of out-of-sample mean squared errors. Given its performance and simplicity, we adopt the fixed effect model as our preferred wage prediction model.

2.B Model Identification and Estimation

2.B.1 Model Identification

By Assumption 5, we can rewrite ξ_{it} as

$$\xi_{it} = \frac{\rho_{u\xi}}{\sigma} u_{it} + \psi_{it} = \frac{\rho_{u\xi}}{\sigma}(w_{it} - \delta_{h(t)} - \delta_0 m_t - \delta_1 m_t \times z_{it}) + \psi_{it}$$

where $\psi_{it} \sim \mathcal{N}(0, 1 - \rho_{u\xi}^2)$ and $\psi_{it} \perp u_{it}$ by construction. We define a new state variable $\phi_{it} \equiv w_{it} - \delta_{h(t)} - \delta_0 m_t - \delta_1 m_t \times z_{it}$. Thus, the flow payoffs of the problem become

$$U_{it} = \begin{Bmatrix} \gamma w_{it} + \beta_{h(t)} + \mu \mathbf{1}\{a_{it-1} = 0\} + \eta_{j(i),h(t)} + \frac{\rho_{u\xi}}{\sigma}\phi_{it} + \nu_{it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{Bmatrix}$$

where $\nu_{it} = \psi_{it} + \epsilon_{1it}$. By further substituting w_{it} in the flow payoff with the wage equation, we have

$$U_{it} = \begin{Bmatrix} \gamma\delta_0 m_t + \gamma\delta_1 m_t \times z_{it} + (\beta_{h(t)} + \gamma\delta_{h(t)}) + \mu \mathbf{1}\{a_{it-1} = 0\} & , \text{ if } a_{it} = 1 \\ \qquad + \eta_{j(i),h(t)} + (\frac{\rho_{u\xi}}{\sigma} + \gamma) u_{it} + \nu_{it} & \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{Bmatrix}$$

It is useful to observe that the state variables can be either expressed as $(w_t, h(t), a_{it-1}, \phi_{it}, j(i))$, or equivalently, $(m_t, z_{it}, h(t), a_{it-1}, u_{it}, j(i))$. Note also that we can recover u_{it} from the wage equation, so u_{it} becomes an observed state in the driver's decision problem. Under Assumption 6, the only unobservables in the flow payoffs are ν_{it}, ϵ_{0it}, and $\eta_{j(i),h(t)}$. In the case where unobserved types $j(i)$ were observed, we can identify $(\gamma\delta_0, \beta_{h(t)} + \gamma\delta_{h(t)}, \mu, \frac{\rho_{u\xi}}{\sigma} + \gamma)$ given the observed state $(m_t, z_{it}, h(t), a_{it-1}, u_{it}, j(i))$. Since $\delta_{h(t)}$, δ_0 and δ_1 are identified from the wage equation, the structural parameters $(\gamma, \beta_{h(t)}, \mu, \frac{\rho_{u\xi}}{\sigma})$ are thus identified. In the case where $j(i)$ is unobserved, we apply (REACH) and (NTYPE) of Assumption 6 as in [44] to jointly identify $\eta_{j(i),h(t)}$ and $(\gamma, \beta_{h(t)}, \mu, \frac{\rho_{u\xi}}{\sigma})$.

2.B.2 Model Estimation

Wage Equation Estimation. We start with estimating the following wage equation using OLS to obtain $\hat{u}_{it}$ and the estimates of $\delta_{h(t)}, \delta_0, \delta_1$:

$$w_{it} = \delta_{h(t)} + \delta_0 m_t + \delta_1 m_t \times z_{it} + u_{it}.$$

Estimation of State Transition. Next, we estimate the joint distribution of the state variables. Recall that we assume the state transition follows a first order Markov process, where the state vector contains hours of the week $h(t)$, treatment status z_{it}, wage component u_{it}, experiment hours m_t, lagged action a_{it-1}, and unobserved type $j(i)$. We also assume that drivers with different unobserved types share the same state transition probability.

Ideally, we can discretize the state space[11] and nonparametrically estimate the transition probabilities. However, the experiment menus vary across blocks and cities, resulting in around 260 distinct experiment menus. Including all of the experiment menus in the estimation of our model greatly enlarges the dimension of the state space. For computational feasibility, we group together similar experiment menus for each city by the K-means clustering method. Specifically, for each experiment menu, we construct a vector of binary indicators for whether an hour has the GSL experiment switched on. We then apply the K-means clustering algorithm on the indicator vectors. We arrive at 5 average experiment menus for each block type and each city. These average experiment menus are taken as inputs in the estimation of the model.

11. Since u_{it} is the only continuous component, we discretize the distribution of $\hat{u}_{it}$ into deciles.

Full Model Estimation. We use the EM algorithm to obtain the maximum likelihood estimates based on equation (2.2), where $s_{it} = (u_{it}, z_{it}, a_{it-1}, h(t), m_t)$ is the vector of the observed states, $\hat{p}$ is a vector of empirical conditional choice probabilities, $\pi(s_{i1})$ is the population probability of type j conditional on the initial state s_{i1}. The number of unobserved types, J, is assumed to be known and equal to 3. We assume that $\hat{\pi}_j(s_1) = \hat{\pi}_j(w_1, a_0)$ so that a driver's unobserved type depends on her initial wage and lagged actions observed in the data.

To obtain the starting values in the EM algorithm, we start with initializing ($\eta_{1,Day}$, $\eta_{1,Night}$, $\eta_{2,Day}$, $\eta_{2,Night}$). The initialization of the remaining structural parameters is set to the estimates of a structural model without any unobserved type. We set the tolerance level to be $1e-7$.

2.C Comparison with Alternative Model Specifications

We illustrate the potential bias that may arise from more restrictive model specifications. We progressively build up from the simplest possible discrete choice model to highlight how several of our modeling choices – including adjustment costs, permanent observed and unobserved heterogeneity, and the field experiment to address wage endogeneity – are key not only to match the data but also for the estimates of the reservation wages and for the results from the counterfactual analyses. The estimates of all the model specifications discussed are reported in Table 2.14.

2.C.1 Static: Constant Time Preference and Exogenous Wage

Setup. At the beginning of each period, a driver observes the wage w_{it} and the idiosyncratic preference shocks $(\epsilon_{0it}, \epsilon_{1it})$, and she chooses between work $(a_{it} = 1)$ and rest $(a_{it} = 0)$:

$$U_{it} = \begin{cases} \gamma w_{it} + \beta + \epsilon_{1it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{cases}$$

The driver's labor supply decision depends only on the wage, idiosyncratic shocks, and a constant utility of work β. Once the driver observes the realization of the wage and the preference shocks, she chooses to work if and only if $\gamma w_{it} + \beta \geq \epsilon_{0it} - \epsilon_{1it}$. We maintain the assumptions (IID) and (EXOG) as in Section 2.4. We further assume that ϵ_{0it} and ϵ_{1it} are T1EV and independent: $\epsilon_{1it} \perp \epsilon_{0it}$. In this model, we specify the wage equation as $w_{it} = \delta_{h(t)} + u_{it}$. We also show the results of a nonparametric wage process as a robustness check in the later section.

Figure 2.20: Model Fit of Probability of Working Across Hours of the Week

Notes: The solid line plots the data and the dashed line plots the prediction from the model.

Model Fit. Figure 2.20 compares the model implied probability of working across hours of a week against the data counterparts. While the data reveals a clear picture of systematic variation in the probabilities of working over time, the simple static model does not predict enough variation. Since the preference shocks are IID across time and drivers in our model, the variability in the probability of working comes only from the variability in wages. Therefore, we conclude that the variation in wages is not enough to generate the large variation in the probabilities of working observed in the data.

2.C.2 Static: Varying Time Preference and Exogenous Wage

Since the variation in wages is not sufficient to generate the observed variation in the probability of working over time, we allow the utility of work to vary across hours of the week in the model.

Setup. At the beginning of each period, a driver observes w_{it} and the idiosyncratic shocks $(\epsilon_{0it}, \epsilon_{1it})$, and she chooses a_{it}:

$$U_{it} = \left\{ \begin{array}{ll} \gamma w_{it} + \beta_{h(t)} + \epsilon_{1it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{array} \right\}$$

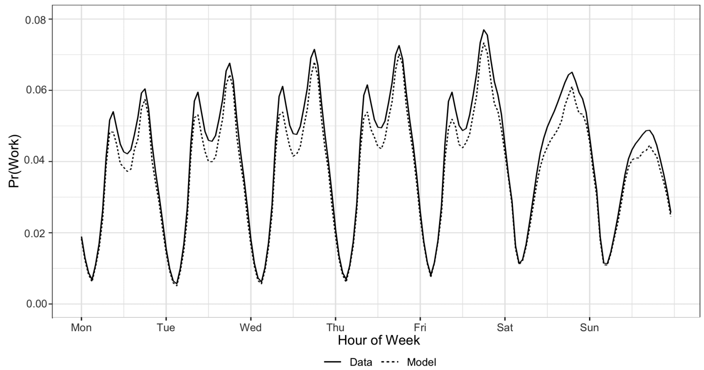
where $h(t)$ is the hour of a week at t, and $\beta_{h(t)}$ is the time-specific shifter of the cost of driving at a given hour of the week. We maintain the above assumptions and the parametric specification on the wage process.

Model Fit. Figure 2.21a compares the predicted probability of working across hours of a week against the data counterparts. With time-varying time preferences, the model now successfully captures drivers' working patterns across hours of a week.

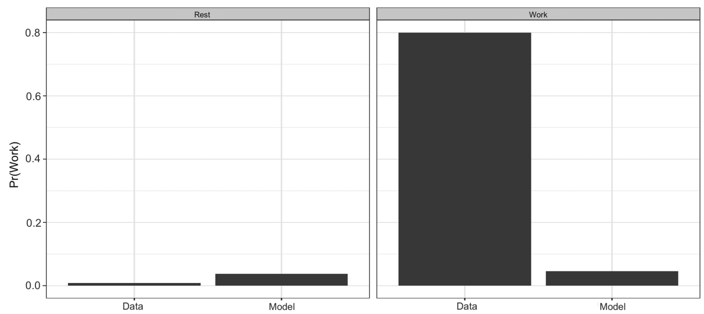
2.C.3 Dynamic:Time-Varying Preference and Exogenous Wage

Although the static model specification in the previous section fits the working patterns across hours of a week excellently, it fails to capture the dynamics of work that we observe in the data. In Figure 2.21b, we compare the probabilities of working conditional on lagged work choices. The static model predicts no difference in the probabilities of working between those who worked in the previous period and those who rested. Yet we observe in the data a much higher probability of working if one worked in the previous period. This state-dependency emphasizes the importance of incorporating the dynamic component in the driver's decision problem to capture the connection between the decision to drive in the current period and future utility. We now consider a dynamic model with adjustments costs and time-varying preferences.

Setup. At the beginning of each period, a forward-looking driver observes w_{it} and the idiosyncratic preference shocks $(\epsilon_{0it}, \epsilon_{1it})$, and forms expectations on future wages and pref-



(a) Hours of the Week



(b) Lagged Choices

Figure 2.21: Model Fit of Probability of Working

Notes: In Figure (a), the solid line plots the data and the dashed line plots the prediction from the model. In Figure (b), the "Rest" panel shows the probability of working conditional on drivers not working in the previous hour. The "Work" panel shows the probability of working conditional on drivers working in the previous hour.

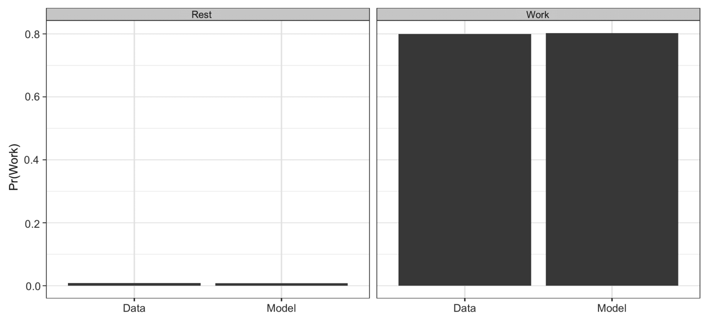
erence shocks. The driver then chooses a_{it} based on the following flow payoffs, taking into account her current choice will affect the expected payoffs in the future:

$$U_{it} = \left\{ \begin{array}{ll} \gamma w_{it} + \beta_{h(t)} + \mu \mathbf{1}\{a_{it-1} = 0\} + \epsilon_{1it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{array} \right\}$$

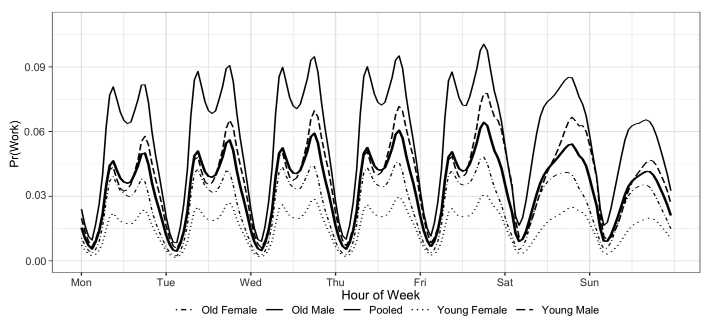
Note that the flow payoff from work depends on whether the driver worked in the previous period; if the driver did not work at $t-1$, she needs to pay an adjustment cost μ to start to work. Both the adjustment cost and the wage transition are linking together periods, since a driver who has paid the adjustment cost in t has a higher continuation value in $t+1$. With fixed costs, the model now becomes a dynamic discrete choice problem. We maintain the above assumptions and the parametric specification of the wage process. In addition, we assume (CI-X) and (DISCOUNT) with the following modification:

(CI-X') State transition probability F satisfies

$$F(w_{it+1}|a_{it}, a_{it-1}, w_{it}, h(t), \epsilon_{1it}) = F(w_{it+1}|a_{it}, a_{it-1}, w_{it}, h(t))$$



(a) Lagged Choices



(b) Demographic Groups

Figure 2.22: Model Fit of Probability of Working

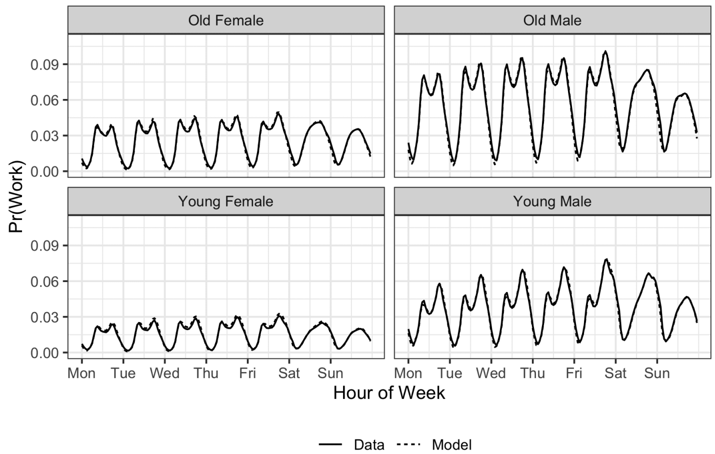
Notes: In Figure (a), the "Rest" panel shows the probability of working conditional on drivers not working in the previous hour. The "Work" panel shows the probability of working conditional on drivers working in the previous hour. In Figure (b), we estimate the probability of working across hours of the week for each demographic group.

Model Fit. Figure 2.22a plots the actual and model implied probabilities of working conditional on the lagged work choice. The dynamic model with the adjustment cost fits the probabilities of working conditional on the lagged choice very well. We, therefore, conclude that (some form of) adjustment costs need to be included to meaningfully represent the driver's decision problem.

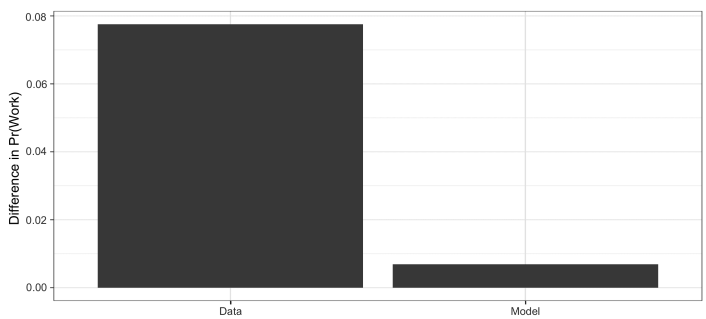
2.C.4 Observed Heterogeneity

So far, we have assumed no systematic preference heterogeneity among drivers. Conditional on wages, the hour of a week, and the lagged choice, the only source that generates differences in drivers' choices is the idiosyncratic preference shocks. Figure 2.22b plots the probability of working across the four subgroups of drivers partitioned by gender and age. This figure shows substantial heterogeneity across the four demographic groups in the probabilities of working. On average, male drivers and older drivers work more than other drivers. This large difference in the probabilities of working across the demographic groups suggests that it is key to let preferences of the drivers to vary by observables such as gender and age.

Estimation and Model Fit. In order to capture the observed heterogeneity along the demographic (gender $\times$ age) dimension, we estimate the dynamic model subgroup by subgroup. The dynamic model fits excellently each demographic group's working patterns. Figure 2.23a plots the conditional choice probabilities for each subgroup against the data counterparts.



(a) Hours of the Week Across Demographic Groups



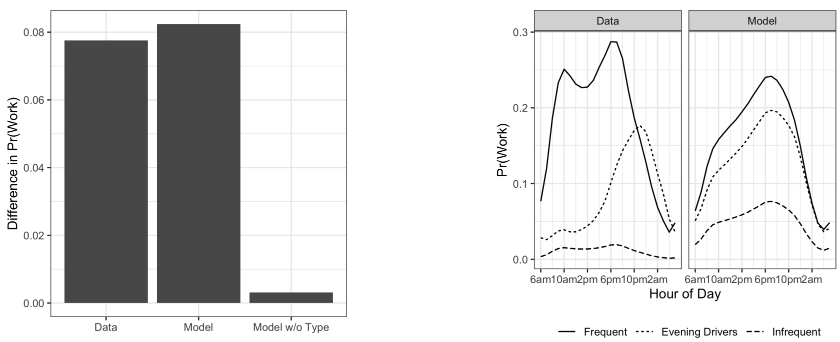
(b) Bottom 50 Percent v.s. Top 50 Percent

Figure 2.23: Model Fit of Probability of Working

Notes: In Figure (a), the solid line plots the data and the dashed line plots the prediction from the model. We estimate the probability of working across hours of a week for each demographic group in the data and compare it with the probability predicted by the model. In Figure (b), we estimate the difference in the average probability of working between drivers in the top 50 percent and those in the bottom 50 percent, where drivers are ranked by the average hours worked per week.

2.C.5 Time-Invariant Unobserved Heterogeneity

We have shown that there exists a lot of observed heterogeneity across drivers. However, there may also be important unobserved heterogeneity across drivers in the costs or non-wage benefits of working. To illustrate this, Figure 2.23b shows the difference in the average probabilities of working of what we refer to as full-time drivers and infrequent drivers. Full-time drivers have average hours worked above the median. Infrequent drivers have average hours worked below the median. The rank of average hours worked is calculated by pooling all drivers from the four subgroups. As is shown in Figure 2.23b, there is a large difference in the average work probabilities between the top 50 percent and the bottom 50 percent of the drivers, pointing out that there is a great deal of heterogeneity even conditional on observables. We now introduce persistence in the costs or non-wage benefits of working to the model by including unobserved driver types in the flow payoff of work.



(a) Bottom v.s. Top 50 Percent (b) Hours of the Day by Types

Figure 2.24: Model Fit of Probability of Working

Notes: In Figure (a), we estimate the difference in the average probability of working between drivers above and below the median in the distribution of the average hours worked per week. We conduct the estimation on the data, the simulated data from the model without unobserved types, and the simulated data from the model with two time-invariant unobserved types. In Figure (b), we plot the average probabilities of working within a day of three groups of drivers: (1) infrequent drivers whose average hours worked in the daytime and in the nighttime both rank in the bottom 80 percent, (2) evening drivers whose average hours worked in the daytime rank in the bottom 80 percentile but hours worked in the nighttime rank in the top 20 percent, and (3) the remaining frequent drivers. The ranks are calculated by pooling all four observed subgroups. The daytime is defined to be 6 a.m.-9 p.m., and the nighttime to be 10 p.m.-5 a.m. next day.

Setup of Time-Invariant Unobserved Types. At the beginning of each period, a forward-looking driver observes w_{it} and the idiosyncratic shocks $(\epsilon_{0it}, \epsilon_{1it})$, and forms expectations on future wages and shocks. She then chooses a_{it} based on the following:

$$U_{it} = \left\{ \begin{array}{ll} \gamma w_{it} + \beta_{h(t)} + \mu \mathbf{1}\{a_{it-1} = 0\} + \eta_{j(i)} + \epsilon_{1it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{array} \right\}$$

where $j(i)$ is driver i's unobserved type, and $\eta_{j(i)}$ is a shifter of the cost of driving, e.g., full-time drivers have higher $\eta_{j(i)}$ and are thus more likely to drive. We maintain all the assumptions above, use the parametric specification of the wage equation, and invoke the assumption (REACH) and (NTYPE). For now, we assume there are two unobserved types of drivers, full-time drivers and infrequent drivers.

Model Fit. Figure 2.24a plots the difference in the average probabilities of working between drivers whose average hours worked rank above the median and those below the median. There is a significant improvement in the model fit once we include the two unobserved types to the model. Capturing such unobserved persistence in driving is important to model the labor supply of Uber drivers.

2.C.6 Time-Varying Unobserved Types

The assumption of time-invariant unobserved types implies that observationally equivalent workers may systematically differ in how much they work, but not in when they work a lot. In the Uber setting, however, there is likely to be a subset of drivers who work mostly in the evenings due to daytime jobs other than driving for Uber. To illustrate this, Figure 2.24b plots the average probabilities of working within a day of three groups of drivers: (1) infrequent drivers whose average hours worked in the daytime and in the nighttime both rank among the bottom 80 percent,[12] (2) evening drivers whose average hours worked in the daytime rank among the bottom 80 percent but hours worked in the nighttime rank among the top 20 percent, and (3) the remaining frequent drivers. Figure 2.24b reveals a large

12. We define the daytime to be 6 a.m.-9 p.m. and the nighttime to be 10 p.m.-5 a.m. next day.

amount of heterogeneity in the persistence of working within a day across the three types of drivers. By comparison, the model with time-invariant unobserved types is not able to generate such time-specific persistence in driving, even conditional on observables. Motivated by this finding, we introduce time-varying unobserved types to the previous model.

Setup of Time-Varying Unobserved Types. At the beginning of each period, a forward-looking driver observes w_{it} and the idiosyncratic shocks $(\epsilon_{0it}, \epsilon_{1it})$, and forms expectations on future wages and shocks. She then chooses a_{it} based on the following:

$$U_{it} = \left\{ \begin{array}{ll} \gamma w_{it} + \beta_{h(t)} + \mu \mathbf{1}\{a_{it-1} = 0\} + \eta_{j(i),h(t)} + \epsilon_{1it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{array} \right\}$$

where $\eta_{j(i),h(t)}$ is a shifter in the cost of driving for driver i with unobserved type j at the hour of week $h(t)$. For example, evening drivers have lower $\eta_{j(i),h(t)}$ for $t \in$ Morning but higher $\eta_{j(i),h(t)}$ for $t \in$ Evening, so they are more likely to drive only in the evenings. We maintain all the assumptions above and assume there are 3 unobserved driver types.

Model Fit. Figure 2.25 shows the average probabilities of working of the three groups of drivers (infrequent drivers, full-time drivers, and evening drivers) in the data and in the simulated data from the model. The model captures reasonably well the three unobserved types of drivers after taking into account the observed heterogeneity. The patterns of work vary distinctly over time across the three types of drivers. This is true both in the actual and the simulated data.

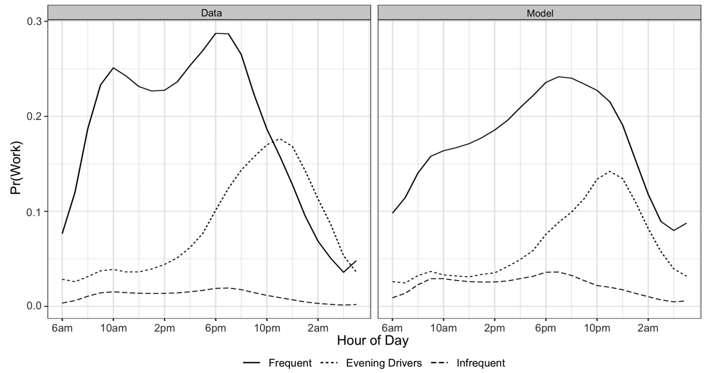


Figure 2.25: Model Fit:Work Probability Across Hours of the Day by Driver Types

Notes: We plot the average probabilities of working within a day of three groups of drivers: (1) infrequent drivers whose average hours worked in the daytime and in the nighttime both rank in the bottom 80 percent, (2) evening drivers whose average hours worked in the daytime rankin the bottom 80 percent but hours worked in the nighttime rank in the top 20 percent, and (3) the remaining frequent drivers. The ranks are calculated by pooling all four observed subgroups. The daytime is defined to be 6 a.m.-9 p.m., and the nighttime to be 10 p.m.-5 a.m. next day.

2.C.7 Probability of Working by Cumulated Hours of Work

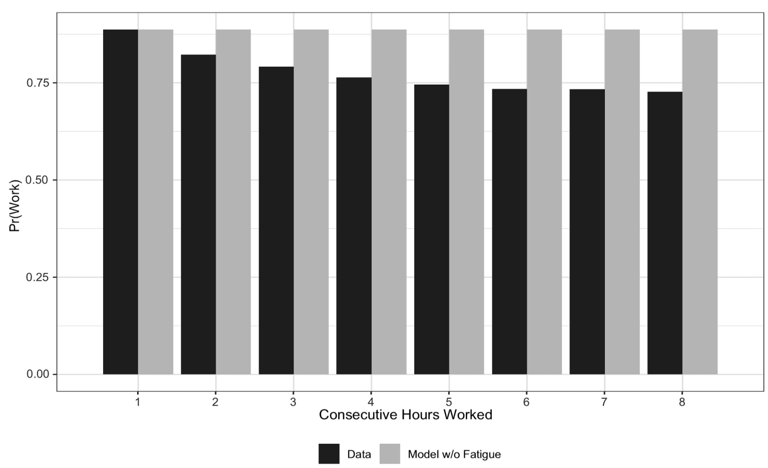


Figure 2.26: Model Fit of Probability of Working by Consecutive Hours Worked

Notes: We define consecutive hours worked as the total number of hours a driver works in a row. The consecutive hour resets to zero whenever a driver stops driving in an hour.

2.C.8 Model of Wages

Before allowing for endogeneity of wages, we show that our estimation results from the previous model do not materially change if we maintain the assumption of exogenous wages but relax the restriction on the process for how wages evolve over time. We relax the wage equation by allowing for a nonparametric evolution of wages as follows. The wage transition is assumed to follow a first-order Markov process. The wage distribution nonparametrically depends on lagged wages, lagged choices, and the hour of the week.[13] We compare the parameter estimates from the models with and without the parametric assumptions on wages in Table 2.10. The estimates of the structural parameters are very similar between the types of models.

Until now, we have assumed exogeneity of wages, ruling out the possibility that market wages may co-move with the cost of driving. As we point out in Section 1.3, the downward bias in the OLS estimates of labor supply responses to wage changes is consistent with demand being high when it is costly or difficult to drive. The inclusion of fixed effects of workers and hours of a week helps reduce but not eliminate the bias, suggesting that market wages and the cost of driving may co-move due to idiosyncratic factors, such as weather conditions or entertainment events. This motivates the use of the experiment to address concerns about wage endogeneity.

13. We first discretize the wage distribution into decile grids, and then nonparametrically estimate the transition probability of the wage process conditional on lagged wages, lagged choices, and the hour of the week.

			Non-parametric Wage Model				Parametric Wage Model			
			Old Male	**Young Male**	**Old Female**	**Young Female**	**Old Male**	**Young Male**	**Old Female**	**Young Female**
Preference for Wage	γ		0.004	0.007	-0.009	-0.006	0.004	0.007	-0.009	-0.005
Time Preferences	β	$E[\beta_{h(t)}]$	-0.666	-0.757	-0.643	-0.724	-0.673	-0.761	-0.645	-0.751
		$Sd(\beta_{h(t)})$	0.615	0.506	0.797	0.694	0.614	0.506	0.794	0.697
		Median$(\beta_{h(t)})$	-0.400	-0.619	-0.258	-0.427	-0.410	-0.616	-0.282	-0.453
		$q_{10}(\beta_{h(t)})$	-1.642	-1.417	-1.825	-1.793	-1.641	-1.429	-1.840	-1.822
		$q_{90}(\beta_{h(t)})$	-0.093	-0.205	0.052	-0.066	-0.100	-0.208	0.042	-0.091
Adjustment Cost	μ		-5.515	-5.651	-6.303	-6.256	-5.517	-5.653	-6.306	-6.212
Unobserved Types	η	$\eta_{(1,Night)}$	1.313	1.130	1.715	1.513	1.312	1.131	1.713	1.529
		$\eta_{(1,Day)}$	-0.074	-0.009	-0.047	0.007	-0.074	-0.010	-0.500	0.006
		$\eta_{(2,Night)}$	1.213	1.106	1.683	1.575	1.213	1.106	1.678	1.581
		$\eta_{(2,Day)}$	0.339	0.407	0.457	0.498	0.339	0.406	0.458	0.502

Table 2.10: Estimates of Non-Parametric and Parametric Wage Models

Notes: ”Young” is defined as those whose ages are less than or equal to the median age.

			Time-Variant Unobs. Types					Full Model with Control Function				
			Weighted Average	**Old Male**	**Young Male**	**Old Female**	**Young Female**	**Weighted Average**	**Old Male**	**Young Male**	**Old Female**	**Young Female**
Preference for Wage	γ		0.002	0.004	0.007	-0.009	-0.006	**0.036**	**0.040**	**0.042**	**0.024**	**0.021**
Time Preferences	β	$E[\beta_{h(t)}]$	-0.704	-0.666	-0.757	-0.643	-0.724	-1.544	-1.612	-1.760	-1.257	-0.977
		$Sd(\beta_{h(t)})$	0.607	0.615	0.506	0.797	0.694	0.513	0.516	0.462	0.597	0.570
		Median$(\beta_{h(t)})$	-0.468	-0.400	-0.619	-0.258	-0.427	-1.328	-1.375	-1.598	-0.972	-0.734
		$q_{10}(\beta_{h(t)})$	-1.599	-1.642	-1.417	-1.825	-1.793	-2.295	-2.428	-2.351	-2.164	-1.861
		$q_{90}(\beta_{h(t)})$	-0.113	-0.093	-0.205	0.052	-0.066	-1.055	-1.151	-1.301	-0.653	-0.432
Adjustment Cost	μ		-5.757	-5.515	-5.651	-6.303	-6.256	-6.377	-6.191	-6.269	-6.771	-6.864
Unobserved Types	η											
		$\eta_{(1,Night)}$	1.320	1.313	1.130	1.715	1.513	1.857	1.847	1.901	2.039	1.571
		$\eta_{(1,Day)}$	-0.036	-0.074	-0.009	-0.047	0.007	0.601	0.589	0.669	0.547	0.483
		$\eta_{(2,Night)}$	1.277	1.213	1.106	1.683	1.575	1.178	1.120	1.237	1.288	1.067
		$\eta_{(2,Day)}$	0.399	0.339	0.407	0.457	0.498	0.794	0.773	0.802	0.838	0.788
Selection Term	$\frac{\rho_{u\xi}}{\sigma}$							**-0.039**	**-0.043**	**-0.044**	**-0.028**	**-0.024**

Table 2.11: Estimates of the Exogenous Wage Model and the Full Model

Notes: "Weighted average" is calculated by averaging the estimates of the four demographic groups weighted by the share of the drivers. "Young" is defined as those whose ages are less than or equal to the median age.

Table 2.11 compares the point estimates of the model assuming exogenous wages and the one allowing for endogenous wages. Notably, the point estimates of the preference for wage parameter, γ, which captures drivers' sensitivity to wage changes, increase substantially for all of the four subgroups. In the case of female drivers, the sign of γ estimates flips and becomes positive. The estimate of the correction term $\frac{\rho_{u\xi}}{\sigma}$ is negative in all four subgroups. Recall that $\rho_{u\xi}$ is the correlation coefficient between the preference shock ξ_{it} and the wage component u_{it}. Our estimation results indicate that the costs of working tend to co-move with the market wages. Therefore, it is important to take this endogeneity into account to obtain reliable estimates of the preference parameters.

2.C.9 Economic Implications of the Alternative Modeling Choices

We have demonstrated that the fit of the models improves as we gradually build up from the simple static model up to the full specification. The inclusion of the key components in our model also affects the estimates of the key parameters of interest. We now examine how these parameter estimates, such as preference heterogeneity and adjustment costs, influence the implied reservation wages.

	Reservation Wage			
	Daytime	Evening	Weekday	Weekend
Constant Time Preference	1.80	1.80	1.80	1.80
Varying Time Preference	0.81	0.99	0.87	0.88
Adjustment Cost	6.10	7.99	6.70	6.81
Obs. Heterogeneity	3.16	4.18	3.48	3.56
Time-Invariant Unobs. Types	4.87	6.25	5.31	5.38
Obs. & Unobs. Heterogeneity	3.27	4.64	3.70	3.79
Full Model	1.34	1.65	1.43	1.46

Table 2.12: Comparison of Counterfactual Analyses by Model Specifications

Notes: "Reservation Wage" is calculated as the average minimal wages required to work. We normalize the reservation wages to the old male drivers' reservation wage in the daytime in the full model.

In Column 1-4 in Table 2.12, we report the model implied reservation wages (normalized by old males' reservation wages in the daytime). Clearly, the model specification matters for the estimates of reservation wages. Under the static model with a constant preference for time, the model implied expected reservation wages do not vary over time. Once we allow for time-varying preferences, the expected reservation wages exhibits variation over time. Compared with the full model, the exogenous wage models have much higher reservation wages.

2.D Robustness Checks: Substitution to Lyft

We now present the results from two sets of robustness checks. First, we examine if our findings differ systematically in cities or in periods in which Lyft has relatively high or low market shares. Second, we compare the results for all drivers to those we obtain from a subsample of drivers who are ineligible to drive for Lyft.

2.D.1 Variation in Lyft Share Across Markets

Experimental Estimates by Market. Our sample consists of 3 cities (Boston, Chicago, and San Fransisco), each of which experienced differential growth in the market share of Lyft during the sample period. We estimate wage elasticities for each city-month pair, using the GSL experimental variation in wages, and present the elasticities against Lyft market share in Figure 2.27. As the figure suggests, there is no evidence for correlation between wage elasticities and Lyft market shares.

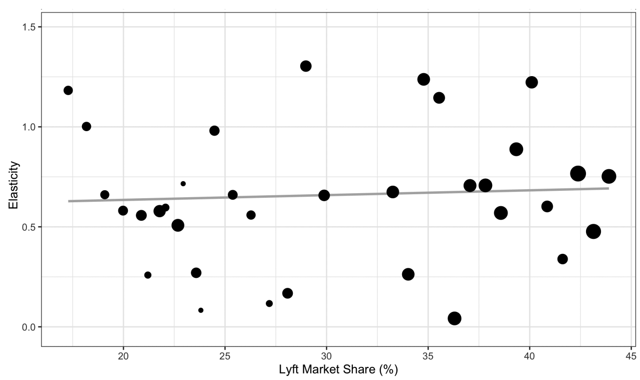


Figure 2.27: Wage Elasticities by Lyft Market Share

Notes: Each dot represents a city $\times$ month wage elasticity estimate. The x-axis denotes the Lyft market share. The size of each dot indicates the number of observations in each city $\times$ month. The solid line plots the linear regression fit weighted by the number of observations in city $\times$ month.

Structural Model Estimates by Market. We now show how much the estimated model parameters vary with the Lyft market share over time and across cities. For each city$\times$time pair,[14] we estimate the model and present the estimates of the value of time parameter, $\beta_{h(t)}$, in Figure 2.28.[15] We see limited changes in $\beta_{h(t)}$ estimates across markets for each demographic subgroup.

2.D.2 Variation in Lyft Eligibility

Experimental Estimates with Lyft Ineligible Drivers. Drivers whose vehicle's model year falls between 2001 and 2003 can only drive for Uber but not for Lyft. In our sample, 3 percent of drivers have vehicles in this range. We define these drivers as Lyft ineligible drivers and estimate the labor supply elasticities on this Lyft ineligible sample. Similar to the labor supply elasticity estimate 0.53 from the sample of all drivers, the Lyft ineligible drivers' elasticity is estimated to be 0.56.

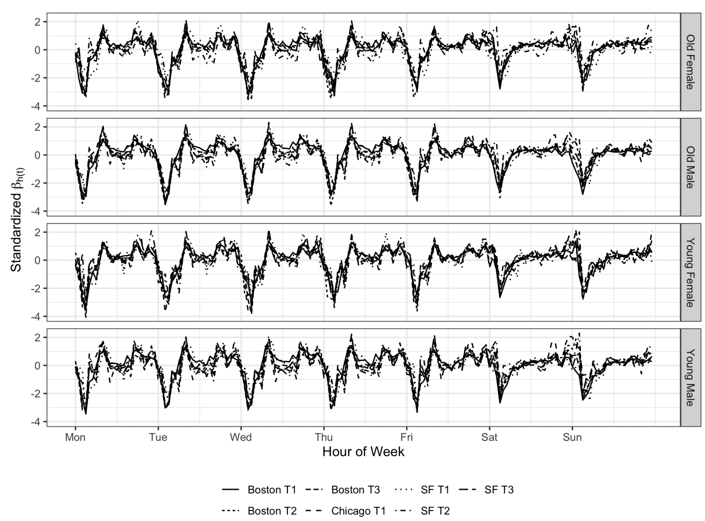


Figure 2.28: Standardized $\beta_{h(t)}$ by Markets and Demographic Groups

Notes: We define a market as a 5-month $\times$city pair. We estimate the model for each market. T1 represents the period from 2016/10/21 to 2017/03/31. T2 represents the period from 2017/04/01 to 2017/09/31. T3 represents the period from 2017/10/01 to 2018/03/01. We standardize $\beta_{h(t)}$ by de-meaning the raw estimates and dividing them by the standard deviation.

14. We define time as a 5-month period.

15. To avoid empty cells due to sample size restrictions, we estimate a model with endogenous wage but no unobserved heterogeneity.

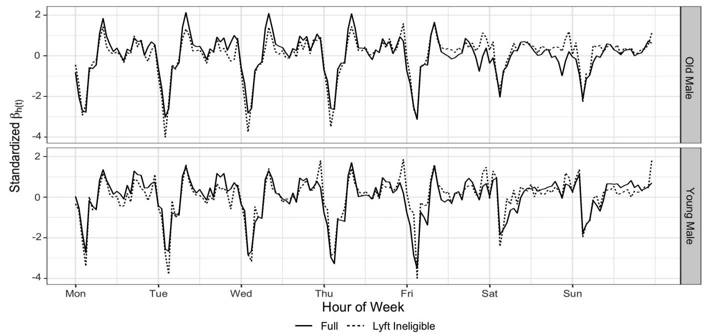


Figure 2.29: Standardized $\beta_{h(t)}$ by Lyft Eligibility

Notes: We standardize $\beta_{h(t)}$ by de-meaning the raw estimates and dividing them by the standard deviation.

Structural Model Estimates with Lyft Ineligible Drivers. Taking advantage of the variation in the Lyft driving eligibility, we separately estimate the structural model on the full sample and the Lyft ineligible sample for male drivers. We allow for heterogeneity by age but no unobserved types, due to the much smaller sample size of Lyft ineligible drivers. In Figure 2.29, we compare the parameter estimates of $\beta_{h(t)}$ for all drivers to those we obtain from the subsample of drivers who are ineligible to drive for Lyft. The estimated value of time parameters, $\beta_{h(t)}$, for Lyft ineligible drivers exhibit a very similar pattern as compared to all drivers. We report the structural parameter estimates from the full sample and the Lyft ineligible sample in Table 2.13. Reassuringly, the structural estimates from the two samples mirror each other.

			Full		Lyft Ineligible	
			Old Male	**Young Male**	**Old Male**	**Young Male**
Preference for Wage	γ		0.042	0.039	0.044	0.043
Time Preferences	β	$E[\beta_{h(t)}]$	-1.152	-0.986	-1.140	-1.082
		$Sd(\beta_{h(t)})$	0.418	0.409	0.592	0.596
		Median$(\beta_{h(t)})$	-1.068	-0.872	-0.995	-0.976
		$q_{10}(\beta_{h(t)})$	-1.641	-1.515	-1.776	-1.691
		$q_{90}(\beta_{h(t)})$	-0.754	-0.607	-0.633	-0.521
Adjustment Cost	μ		-6.479	-7.115	-7.331	-8.176
Selection Term	$\frac{\rho_{u\xi}}{\sigma}$		-0.034	-0.030	-0.037	-0.035

Table 2.13: Estimates by Lyft Eligibility

Notes: "Full" refers to the full sample.

2.E Additional Results on Model

2.E.1 Additional Model Fit Results

Figure 2.30 shows that the working pattern by treatment status across the four demographic groups is well captured by our model, and Figure 2.31 shows that the three unobserved types of drivers are captured reasonably well after taking into account the observed heterogeneity. We use the observed data and the simulated data to plot the average work probabilities within a day of three groups of drivers: (1) infrequent drivers whose average hours worked during the day and at night both rank in the bottom 80 percent,[16] (2) evening drivers whose average hours worked in daytime rank in bottom 80 percent but the hours worked in the

16. We define the daytime to be 6 a.m.-9 p.m., and the nighttime to be 10 p.m.-5 a.m. next day.

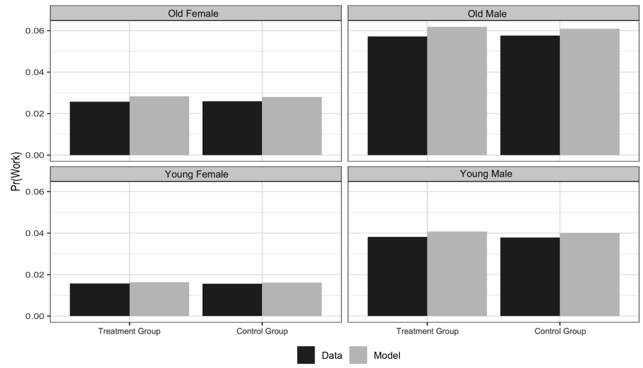


Figure 2.30: Model Fit of Probability of Working by the Treatment Status Across Demographic Groups

Notes: The black bar plots the data, and the grey bar plots the prediction from the model.

nighttime rank in the top 20 percent, and (3) the remaining drivers, who we refer to as frequent drivers because they tend to work a considerable amount both during the day and at night. The patterns of work vary distinctly over time across the three types of drivers. This is true both in the actual and the simulated data.

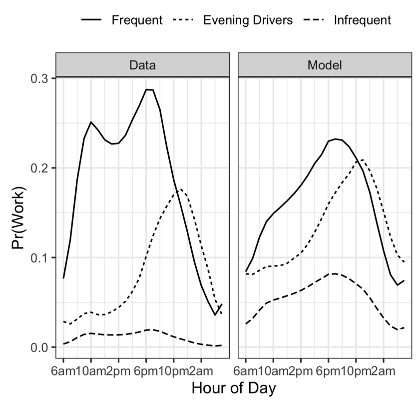


Figure 2.31: Model Fit of Probability of Working by Worker Types Across Hours of the Day

Notes: In this figure, we plot the average work probabilities within a day of three groups of drivers: (1) infrequent drivers whose average hours worked in the daytime and in the nighttime both rank in the bottom 80 percent, (2) evening drivers whose average hours worked in the daytime rank in the bottom 80 percent, but hours worked in the nighttime rank in the top 20 percent, and (3) the remaining frequent drivers. The ranks are calculated by pooling drivers from all four observed subgroups. The daytime is defined to be 6 a.m.-9 p.m., and the nighttime to be 10 p.m.-5 a.m. next day.

2.E.2 Derivation of IES

To derive the desired expression of the IES, we first denote $Pr(a_{it} = 1|s_{it} = s, w_{it} = w)$ by P and $Pr(a_{i\tilde{t}} = 1|s_{it} = s, w_{it} = w)$ by $\tilde{P}$. Denote $\frac{\partial P}{\partial w_{it}}$ by P' and $\frac{\partial \tilde{P}}{\partial w_{it}}$ by $\tilde{P}'$. We set $\tilde{t}=$ t+1 and consider the labor supply response by changing w_{it} around $w_{it} = w$ but holding the distribution of future wage $w_{i\tilde{t}} = \tilde{w}$ fixed. Now, we rewrite δ^{IES} as follows:

$$\delta^{EIS}_{t,\tilde{t}} = \left(\frac{P}{\tilde{P}}\right)' w\frac{\tilde{P}}{P} = \frac{P'}{P}w - \frac{\tilde{P}'}{\tilde{P}}w = \delta^F - \frac{\tilde{P}'}{\tilde{P}}w$$

i.e., the IES is the Frisch elasticity minus the elasticity of future labor supply w.r.t. changes in the current wage. Note that $\tilde{P}$ can be expressed as

$$\begin{aligned}
\tilde{P} &= Pr(a_{i\tilde{t}} = 1|s_{it} = s, w_{it} = w) \\
&= \int\int Pr(a_{i\tilde{t}} = 1|s_{i\tilde{t}} = \tilde{s}, w_{i\tilde{t}} = \tilde{w}, s_{it} = s, w_{it} = w) \\
&\qquad\qquad\qquad\qquad Pr(s_{i\tilde{t}} = \tilde{s}, w_{i\tilde{t}} = \tilde{w}|s_{it} = s, w_{it} = w)d\tilde{s}d\tilde{w} \\
&= \int\int Pr(a_{i\tilde{t}} = 1|s_{i\tilde{t}} = \tilde{s}, w_{i\tilde{t}} = \tilde{w})Pr(s_{i\tilde{t}} = \tilde{s}, w_{i\tilde{t}} = \tilde{w}|s_{it} = s, w_{it} = w)d\tilde{s}d\tilde{w} \\
&= \int\int Pr(a_{i\tilde{t}} = 1|s_{i\tilde{t}} = \tilde{s}, w_{i\tilde{t}} = \tilde{w})Pr(s_{i\tilde{t}} = \tilde{s}|w_{i\tilde{t}} = \tilde{w}, s_{it} = s, w_{it} = w) \\
&\qquad\qquad\qquad\qquad Pr(w_{i\tilde{t}} = \tilde{w}|s_{it} = s, w_{it} = w)d\tilde{s}d\tilde{w}
\end{aligned}$$

The third equality is a result of applying (CI-X). We can then simplify $\frac{\tilde{P}'}{\tilde{P}}w$ as

$$\begin{aligned}
\frac{\tilde{P}'}{\tilde{P}}w = \int\int & Pr(w_{i\tilde{t}} = \tilde{w}, a_{i\tilde{t}-1} = 1|s_{it} = s, w_{it} = w) \times \delta^F_{\tilde{w}} \\
& \times \frac{(Pr(a_{i\tilde{t}} = 1|a_{i\tilde{t}-1} = 1, w_{i\tilde{t}} = \tilde{w}) - Pr(a_{i\tilde{t}} = 1|a_{i\tilde{t}-1} = 0, w_{i\tilde{t}} = \tilde{w}))}{Pr(a_{i\tilde{t}} = 1|s_{it} = s, w_{it} = w)} d\tilde{w}d\tilde{s'}
\end{aligned}$$

where $\tilde{s'}$ represents states other than wage in period $\tilde{t}$, and

$$\delta^F_{\tilde{w}} = \frac{w\frac{\partial}{\partial w} Pr(a_{i\tilde{t}-1} = 1 | w_{i\tilde{t}} = \tilde{w}, s_{it} = s, w_{it} = w)}{Pr(s_{i\tilde{t}} = 1 | w_{i\tilde{t}} = \tilde{w}, s_{it} = s, w_{it} = w)},$$

and we can define

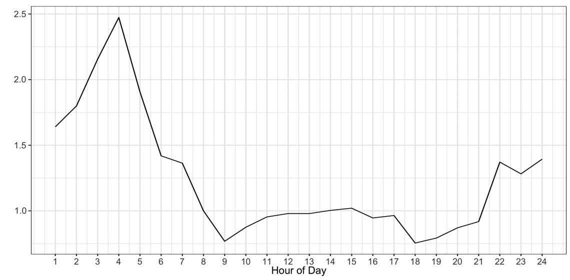
$$g(s, w) \equiv Pr(w_{i\tilde{t}} = \tilde{w}, a_{i\tilde{t}-1} = 1 | s_{it} = s, w_{it} = w)$$

$$\times \frac{(Pr(a_{i\tilde{t}} = 1 | a_{i\tilde{t}-1} = 1, w_{i\tilde{t}} = \tilde{w}) - Pr(a_{i\tilde{t}} = 1 | a_{i\tilde{t}-1} = 0, w_{i\tilde{t}} = \tilde{w}))}{Pr(a_{i\tilde{t}} = 1 | s_{it} = s, w_{it} = w)}.$$

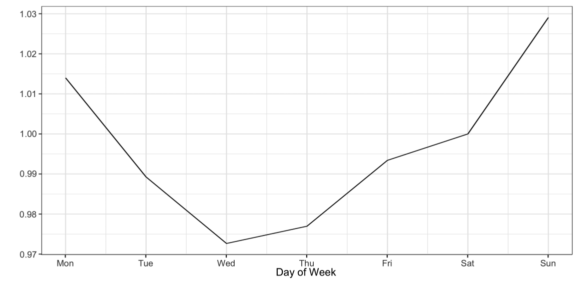
2.E.3 Counterfactual: Value of Time and Reservation Wages

To examine whether the variation mostly comes from hours of a day or days of a week, Figure 2.32a shows the average $\beta_{h(t)}$ per hour of the day (relative to 8 a.m.), while Figure 2.32b reports the average $\beta_{h(t)}$ per day of the week (relative to Saturday). The value of time varies a lot across hours within a day. By way of comparison, there is little variation in the value of time across weekdays. The value of time tends to be higher during weekends than weekdays, but the differences are rather small.

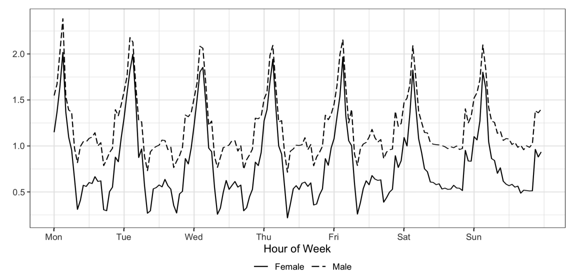
In Figure 2.32c and 2.32d , we compare the value of time across the demographic groups. For each group, we normalize the estimates of $\beta_{h(t)}$ by the value of time for older males on Saturdays at 8 a.m.



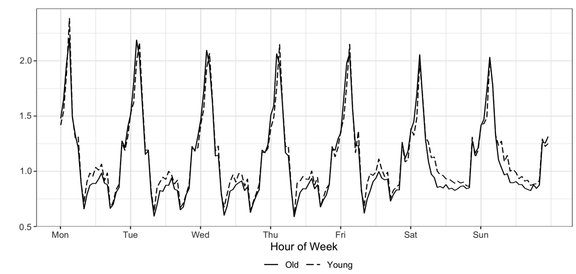
(a) $\beta_{h(t)}$ Relative to 8 a.m.



(b) $\beta_{h(t)}$ Relative to Saturday



(c) Across Gender



(d) By Age

Figure 2.32: Variation in the Value of Time, $\beta_{h(t)}$,

Notes: In (a), we compute the average $\beta_{h(t)}$ per hour of the day (relative to 8 a.m.) weighted by the shares of each demographic group. In (b), we compute the average $\beta_{h(t)}$ per day of the week (relative to Saturday) weighted by the population shares of each demographic group. In (c) and (d), we compute the weighted average of $\beta_{h(t)}$ relative to old males' value of time on Saturday 8 a.m. using population shares as weights.

2.E.4 Counterfactual: Value of the Ability to Set Customized Work Schedules

Figure 2.33 and 2.34 plot the wage multipliers by gender, age, and driver types.

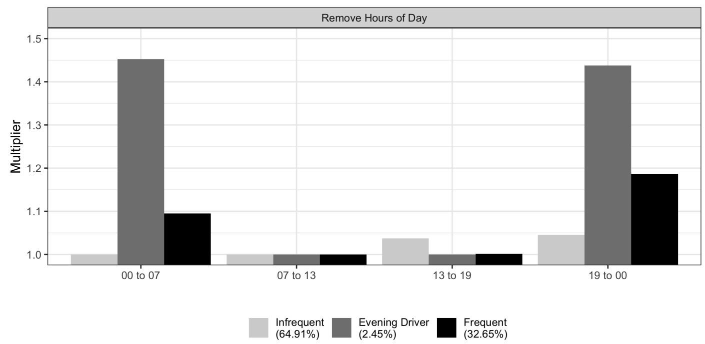
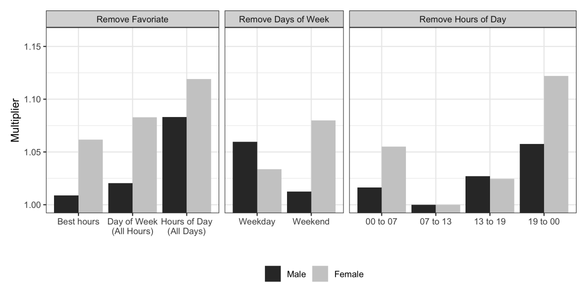
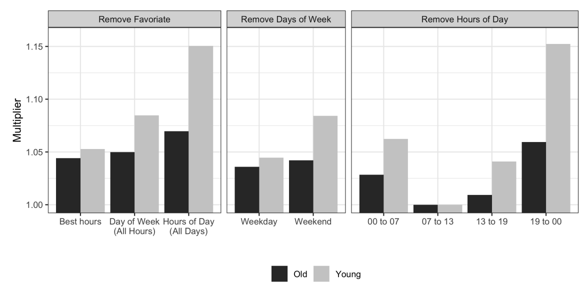


Figure 2.33: Unobserved Types

Notes: We remove certain hour blocks across the entire week from drivers' choice sets. The numbers in the parentheses indicate the fraction of each type of drivers.



(a) Gender



(b) Age

Figure 2.34: Wage Multipliers to Accept Restrictions on the Choice Set

Notes: In Figure (a) and (b), the left panel removes each drivers' favorite 5-hour block from their choice sets. The "Best hours" indicates the favorite 5-hour block of the week for the drivers. The middle panel removes weekdays or weekends from drivers' choice sets. The right panel removes certain hour blocks across the entire week from drivers' choice sets.

2.E.5 Counterfactual: Value of the Ability to Both Customize and Adjust Work Schedules

To infer the value of the ability to both customize and adjust work schedules, we let every driver work 5 consecutive hours per week. For now, we only let the drivers make one adjustment per week in scenario (2) and (3). In scenario (1), we fix the work schedule of each driver to be Friday evening, 7:00 -11:59 p.m.

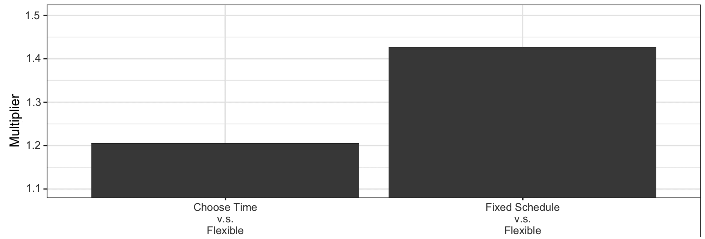
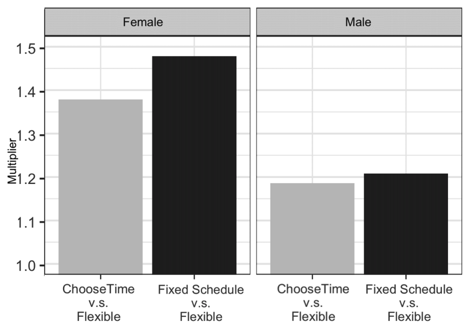
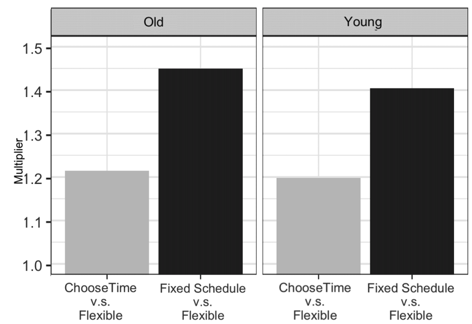


Figure 2.35: Wage Multipliers to Accept Restrictions on the Choice Set and Work Schedule Adjustments

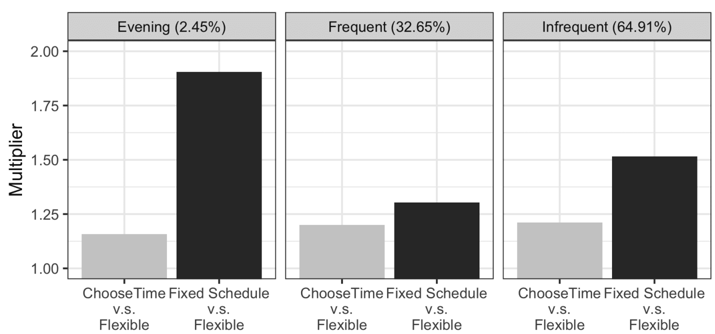
Notes: In this figure, we compute the weighted average of the wage multipliers needed for drivers to be indifferent between the different types of work schedules. The fixed work schedule starts from 7:00 p.m. on Friday and ends at 11:59 p.m. on Friday.



(a) Gender



(b) Age



(c) Unobserved Types

Figure 2.36: Heterogeneity in Wage Multipliers to Accept Restrictions on the Choice Set and Work Schedule Adjustments

Notes: In these figures, we compute the wage multipliers needed for drivers to be indifferent between the different types of work schedules. The fixed work schedule starts from 7:00 p.m. on Friday and ends at 11:59 p.m. on Friday.

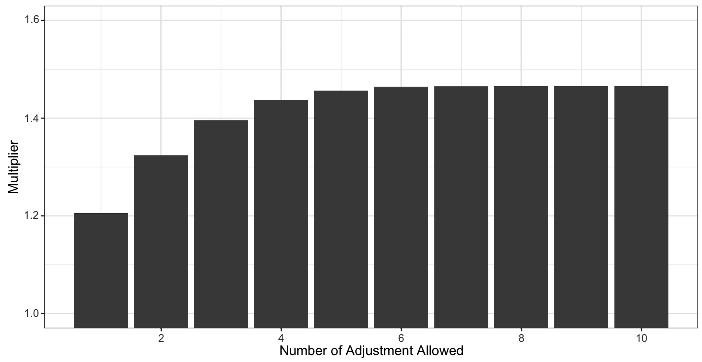


Figure 2.37: Wage Multiplier with Increasing Number of Work Schedule Adjustments

Notes: In this figure, we compute the wage multipliers needed for drivers to be indifferent between the commitment scheme and the flexible scheme with increasing numbers of adjustments.

Specification		Preference for Wage	Time Preferences					Adjustment Cost	Unobserved Types					Selection Term
		γ	β					μ	η					$\frac{\rho_{u\xi}}{\sigma}$
			$E[\beta_{h(t)}]$	$Sd(\beta_{h(t)})$	Median$(\beta_{h(t)})$	$q_{10}(\beta_{h(t)})$	$q_{90}(\beta_{h(t)})$		η	$\eta_{(1,Night)}$	$\eta_{(1,Day)}$	$\eta_{(2,Night)}$	$\eta_{(2,Day)}$	
Constant Time Preference		0.015	-3.692											
Varying Time Preference		0.034	-4.059	0.672	-3.778	-5.259	-3.441							
Adjustment Cost		0.005	-0.376	0.401	-0.283	-0.812	0.021	-6.397						
Obs. Heterogeneity	Weighted Average	0.002	-0.322	0.414	-0.222	-0.790	0.084	-6.453						
	Old Male	0.004	-0.358	0.401	-0.251	-0.881	0.025	-6.018						
	Young Male	0.007	-0.394	0.386	-0.313	-0.751	0.009	-6.470						
	Old Female	-0.008	-0.187	0.489	-0.060	-0.755	0.250	-6.735						
	Young Female	-0.007	-0.129	0.467	-0.024	-0.670	0.320	-7.426						
Time-Invariant Unobs. Types	Weighted Average	0.001	-0.640	0.440	-0.523	-1.161	-0.220	-6.517	0.615					
	Old Male	0.003	-0.660	0.425	-0.540	-1.219	-0.259	-6.148	0.567					
	Young Male	0.007	-0.741	0.409	-0.642	-1.161	-0.320	-6.490	0.612					
	Old Female	-0.013	-0.475	0.527	-0.325	-1.104	-0.057	-6.900	0.719					
	Young Female	-0.009	-0.439	0.493	-0.314	-1.041	0.034	-7.325	0.662					
Time-Variant Unobs. Types	Weighted Average	0.002	-0.712	0.606	-0.477	-1.609	-0.122	-5.753		1.321	-0.037	1.277	0.399	
	Old Male	0.004	-0.673	0.614	-0.410	-1.641	-0.100	-5.517		1.312	-0.074	1.213	0.339	
	Young Male	0.007	-0.761	0.506	-0.616	-1.429	-0.208	-5.653		1.131	-0.010	1.106	0.406	
	Old Female	-0.009	-0.645	0.794	-0.282	-1.840	0.042	-6.306		1.713	-0.050	1.678	0.458	
	Young Female	-0.005	-0.751	0.697	-0.453	-1.822	-0.091	-6.212		1.529	0.006	1.581	0.502	
Time-Variant Unobs. Types (Nonparam)	Weighted Average	0.002	-0.704	0.607	-0.468	-1.599	-0.113	-5.757		1.320	-0.036	1.277	0.399	
	Old Male	0.004	-0.666	0.615	-0.400	-1.642	-0.093	-5.515		1.313	-0.074	1.213	0.339	
	Young Male	0.007	-0.757	0.506	-0.619	-1.417	-0.205	-5.651		1.130	-0.009	1.106	0.407	
	Old Female	-0.009	-0.643	0.797	-0.258	-1.825	0.052	-6.303		1.715	-0.047	1.683	0.457	
	Young Female	-0.006	-0.724	0.694	-0.427	-1.793	-0.066	-6.256		1.513	0.007	1.575	0.498	
Full Model with Control Function	Weighted Average	0.036	-1.544	0.513	-1.328	-2.295	-1.055	-6.377		1.857	0.601	1.178	0.794	-0.039
	Old Male	0.040	-1.612	0.516	-1.375	-2.428	-1.151	-6.191		1.847	0.589	1.120	0.773	-0.043
	Young Male	0.042	-1.760	0.462	-1.598	-2.351	-1.301	-6.269		1.901	0.669	1.237	0.802	-0.044
	Old Female	0.024	-1.257	0.597	-0.972	-2.164	-0.653	-6.771		2.039	0.547	1.288	0.838	-0.028
	Young Female	0.021	-0.977	0.570	-0.734	-1.861	-0.432	-6.864		1.571	0.483	1.067	0.788	-0.024

Table 2.14: Estimates of All Specifications

REFERENCES

[1] Jaap H. Abbring and Øystein Daljord. Identifying the discount factor in dynamic discrete choice models. *Quantitative Economics*, 11(2):471–501, 2020.

[2] Victor Aguirregabiria and Pedro Mira. Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, 156(1):38–67, 2010.

[3] Hyungtaik Ahn and James L. Powell. Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58(1):3 – 29, 1993.

[4] Joshua D Angrist, Sydnee Caldwell, and Jonathan V Hall. Uber vs. taxi: A driver's eye view. *American Economic Journal: Applied Economics*, forthcoming.

[5] Peter Arcidiacono and Robert A. Miller. Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity. *Econometrica*, 79(6):1823–1867, 2011.

[6] Daniel Barczyk and Matthias Kredler. Evaluating long-term-care policy options, taking the family seriously. *Review of Economic Studies*, 85(2):766–809, 2018.

[7] Natalie Bau. Can policy change culture? government pension plans and traditional kinship practices. 2019.

[8] Jan Michael Bauer and Alfonso Sousa-Poza. Impacts of Informal Caregiving on Caregiver Employment, Health, and Family. *Journal of Population Ageing*, 8(3):113–145, 2015.

[9] Magali Beffy, Richard Blundell, Antoine Bozio, Guy Laroque, and Maxime To. Labour supply and taxation with restricted choices. *Journal of Econometrics*, 211(1):16–46, 2019.

[10] Richard Blundell and Thomas MaCurdy. Labor supply: A review of alternative approaches. In *Handbook of labor economics*, volume 3, pages 1559–1695. Elsevier, 1999.

[11] George J Borjas. *Immigration economics*. Harvard University Press, 2014.

[12] David Card and Giovanni Peri. Immigration economics by george j. borjas: a review essay. *Journal of Economic Literature*, 54(4):1333–49, 2016.

[13] Chen-Fen Chen. The development of long-term care labor force in taiwan. Technical report, 6 2014.

[14] M. Keith Chen, Judith A. Chevalier, Peter E. Rossi, and Emily Oehlsen. The value of flexible work: Evidence from uber drivers. *Journal of Political Economy*, 2019.

[15] Victor Chernozhukov, Iván Fernández-Val, and Ye Luo. The sorted effects method: discovering heterogeneous effects beyond their averages. *Econometrica*, 86(6):1911–1938, 2018.

[16] Raj Chetty, Adam Guren, Day Manoli, and Andrea Weber. Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *American Economic Review*, 101(3):471–75, 2011.

[17] CY Cyrus Chu and Ruoh-Rong Yu. *Understanding Chinese families: A comparative study of Taiwan and Southeast China*. Oxford University Press, 2010.

[18] H Chuang and Eric S Lin. The evolution of the empirical study of taiwan's female labor supply. *Taiwan Economic Review*, 34(2):119, 2006.

[19] Peter Cohen, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe. Using big data to estimate consumer surplus: The case of uber. Technical report, National Bureau of Economic Research, 2016.

[20] Francesca Colombo, Ana Llena-Nozal, Jérôme Mercier, and Frits Tjadens. *Help Wanted?*, volume 41 of *OECD Health Policy Studies*. OECD, may 2011.

[21] Cody Cook, Rebecca Diamond, Jonathan V Hall, John A List, and Paul Oyer. The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. 2020.

[22] Patricia Cortés and Jessica Pan. Outsourcing household production: Foreign domestic workers and native labor supply in Hong Kong. *Journal of Labor Economics*, 31(2):327–371, 2013.

[23] Patricia Cortés and Jessica Pan. When time binds: Substitutes for household production, returns to working long hours, and the skilled gender wage gap. *Journal of Labor Economics*, 37(2):351–398, 2019.

[24] John K Dagsvik. Marginal compensated effects and the slutsky equation for discrete choice models. 2020.

[25] Stefano DellaVigna, John A. List, and Ulrike Malmendier. Testing for Altruism and Social Pressure in Charitable Giving *. *The Quarterly Journal of Economics*, 127(1):1–56, 01 2012.

[26] Stefano DellaVigna, John A List, Ulrike Malmendier, and Gautam Rao. Estimating social preferences and gift exchange at work. Working Paper 22043, National Bureau of Economic Research, February 2016.

[27] Department of Health. Survey of long-term care needs, stage ii. Technical report, 2013.

[28] Directorate-General of Budget, Accounting and Statistics, Executive Yuan. Manpower utilization survey. Technical report, 1989-2011.

[29] Avraham Y Ebenstein. The causal effect of fertility on female labor supply: evidence from taiwanese son preference. *Available at SSRN 988852*, 2007.

[30] Tor Eriksson and Nicolai Kristensen. Wages or fringes? some evidence on trade-offs and sorting. *Journal of Labor Economics*, 32(4):899–928, 2014.

[31] Itzik Fadlon and Torben Heien Nielsen. Family labor supply responses to severe health shocks. *Nber Wp 21352*, 2017.

[32] Henry S Farber. Is tomorrow another day? the labor supply of new york city cabdrivers. *Journal of political Economy*, 113(1):46–82, 2005.

[33] Henry S Farber. Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. *The Quarterly Journal of Economics*, 130(4):1975–2026, 2015.

[34] Wolfgang Frimmel, Martin Halla, Jörg Paetzold, and Julia Schmieder. Health of elderly parents, their children's labor supply, and the role of migrant care workers. 2020.

[35] Yang-Chih Fu. 2016 taiwan social change survey (round 7, year 2): Family (c00321$_1$)$. Technical report, 2017.$

[36] Christian Gourieroux, Alain Monfort, and Eric Renault. Indirect inference. *Journal of applied econometrics*, 8(S1):S85–S118, 1993.

[37] Jonathan V Hall and Alan B Krueger. An analysis of the labor market for uber's driver-partners in the united states. *ILR Review*, 71(3):705–732, 2018.

[38] Daniel S Hamermesh. Changing inequality in markets for workplace amenities. *The Quarterly Journal of Economics*, 114(4):1085–1123, 1999.

[39] Glenn W Harrison and John A List. Field experiments. *Journal of Economic literature*, 42(4):1009–1055, 2004.

[40] Haoran He, David Neumark, and Qian Weng. Do workers value flexible jobs? a field experiment on compensating differentials. Technical report, National Bureau of Economic Research, 2019.

[41] V Joseph Hotz and Robert A Miller. Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3):497–529, 1993.

[42] Hui-Chuan Hsu and Chen-Fen Chen. Ltc 2.0: The 2017 reform of home-and community-based long-term care in taiwan. *Health Policy*, 123(10):912–916, 2019.

[43] Hiroyuki Kasahara and Katsumi Shimotsu. Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices. *Econometrica*, 77(1):135–175, 2009.

[44] Hiroyuki Kasahara and Katsumi Shimotsu. Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica*, 77(1):135–175, 2009.

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

[46] Michael P Keane and Kenneth I Wolpin. The career decisions of young men. *Journal of political Economy*, 105(3):473–522, 1997.

[47] Mark R Killingsworth and James J Heckman. Female labor supply: A survey. *Handbook of labor economics*, 1:103–204, 1986.

[48] David Laibson and John A. List. Principles of (behavioral) economics. *American Economic Review*, 105(5):385–90, May 2015.

[49] Thibaut Lamadon, Magne Mogstad, and Bradley Setzler. Imperfect competition, compensating differentials and rent sharing in the us labor market. Technical report, National Bureau of Economic Research, 2019.

[50] Legislative Yuan. Gazette of legislative yuan. Technical Report 13, 2019.

[51] Legislative Yuan. Gazette of legislative yuan. Technical Report 61, 2019.

[52] Legislative Yuan. Gazette of legislative yuan. Technical Report 53, 2019.

[53] Katrine V. Løken, Julie Riise, and Shelly Lundberg. Lifting the burden: Formal care of the elderly and labor supply of adult children. *Journal of Human Resources*, 52(1):247–271, 2017.

[54] Hamish Low and Luigi Pistaferri. Disability insurance and the dynamics of the incentive insurance trade-off. *American Economic Review*, 105(10):2986–3029, 2015.

[55] Nicole Maestas, Kathleen J Mullen, David Powell, Till Von Wachter, and Jeffrey B Wenger. The value of working conditions in the united states and implications for the structure of wages. Technical report, National Bureau of Economic Research, 2018.

[56] Thierry Magnac and David Thesmar. Identifying dynamic discrete decision processes. *Econometrica*, 70(2):801–816, 2002.

[57] Thierry Magnac and David Thesmar. Identifying dynamic discrete decision processes. *Econometrica*, 70(2):801–816, 2002.

[58] Isabel Z Martinez, Emmanuel Saez, and Michael Siegenthaler. Intertemporal labor supply substitution? evidence from the swiss income tax holidays. *American Economic Review*, 111(2):506–46, 2021.

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

[60] Ministry of Labor. Foreign workers statistics. Technical report, 2020.

[61] Corina Mommaerts. Long-Term Care Insurance and the Family. *Working Paper*, (2020):1–70, 2020.

[62] E.C. Norton. Health and Long-Term Care. *Handbook of the Economics of Population Aging*, 1:951–989, jan 2016.

[63] Edward C. Norton. Chapter 17 Long-term care. *Handbook of Health Economics*, 1(PART B):955–994, 2000.

[64] John Pencavel. Labor supply of men: a survey. *Handbook of labor economics*, 1:3–102, 1986.

[65] Brooks Pierce. Compensation inequality. *The Quarterly Journal of Economics*, 116(4):1493–1525, 2001.

[66] Sherwin Rosen. The theory of equalizing differences. *Handbook of labor economics*, 1:641–692, 1986.

[67] John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, pages 999–1033, 1987.

[68] Meghan M. Skira. Dynamic wage and employment effects of elder parent care. *International Economic Review*, 56(1):63–93, 2015.

[69] Matthew Wiswall and Basit Zafar. Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, 133(1):457–507, 2017.