Essays in Labor and Development Economics

Kirk Bennett Doran

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF

ECONOMICS

Adviser: Henry S. Farber

April 2008

UMI Number: 3299832

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Do employers substitute adults for children, or do they treat them as complements? Using a Mexican schooling experiment, I find that a decrease in child field work participation is accompanied by an increase in adult labor demand. This increase was not directly caused by treatment money reaching employers: there were no significant effects on food prices, hectares of land used, or harvest size. Furthermore, the wages of healthy non-treated adults living around children who stopped working also increased. This finding thus supports Basu and Van's Substitution Axiom, raising the possibilities of multiple equilibria and a welfare-improving ban on child labor.

Standard neoclassical theory says that daily consumption of goods and leisure determines daily utility; together with other realistic assumptions this implies that for most workers transitory increases in one day's income should not decrease that day's labor supply. While Farber (2005) verifies that the labor supply of taxi drivers provides no evidence against this prediction on average, I ask whether individual drivers may behave differently. Using a new panel of New York City Taxi drivers who can choose their own labor supply every day, I find that half show no significant impact of daily income on the decision of whether to stop after a given trip, but roughly half show a large and positive impact. The size of a driver's income coefficient is negatively related to the standard deviation in daily income across days. Furthermore, the average income coefficient among the latter drivers is such that if $20 were dropped in their taxi before their first trip, they would end their day early with only $6 extra. Together, these results suggest that almost half of the drivers may have two non-standard features in their utility: their utility may depend on daily income as well as daily consumption, and their utility may have a kink at a particular value of income. Such reference-dependent

behavior could be unrelated to income expectations, or it could be determined by them (Rabin & Koszegi 2006). The fact that these drivers neither work fewer hours nor earn the same income after an exogenous hourly wage increase suggests that their reference points increased when the fare increased, thus supporting the expectations-based theory of Rabin & Koszegi.

A growing empirical literature reports that the physically beautiful are more likely to succeed in many areas. I propose that, if this beauty premium exists, it can be fully accounted for by premiums on other dimensions of attractiveness, such as personality and grooming. Using transcripts tied to a nationally representative survey of American middle and high school students, I find that beautiful students receive higher grades, but also that this beauty premium disappears when attractiveness of personality and grooming are controlled for. Indeed, the remaining marginal effects of physical beauty include significantly lower GPAs and slower course advancement. This marginal "beauty deficit" can be explained by two factors. First, I find a negative marginal effect of beauty on academic effort. Second, I find that physical beauty is associated with much higher social relationship activity. This evidence together suggests that beautiful students may substitute towards social activities and away from academic ones, lowering their academic achievement.

I dedicate this dissertation to my beautiful wife Maggie.

Acknowledgements

This work was initiated, carried out, and brought to completion through the patience, sacrifice, and talent of many people. My adviser, Hank Farber, must be the most dedicated and self-giving adviser in the world. If there isn't already a lifetime achievement award for taking care of your students, they should make one, and they should name it after him. And he should be the first recipient. Orley Ashenfelter sparked my interest in Labor Economics through his graduate Labor Class, guided the first steps of my research on child labor, and provided more useful comments to my work than I can remember (although I did respond to them in this draft, I promise!). Cecilia Rouse taught me what a good economics paper is. And Christina Paxson opened my eyes to the techniques and importance of the economics of child health and development.

I never would have completed graduate school without the guidance of all of these mentors, and without the friendship of so many others. The IR Section students and visitors kept me going when I felt like it was too hard, especially Marie Connolly, Molly Fifer, Jane Fortson, Radha Iyengar, Giovanni Mastrobuoni, Analia Schlosser, Courtney Stoddard, Elod Takats, and Susan Yeh. Linda Belifield, Kathleen DeGennaro, Joyce Howell, and Thu Vu always supported me no matter how many times I messed up. Ling Ling Ang provided invaluable research assistance and an awesome attitude. Awo Addo, Howard Yu, and so many other research assistants made it possible for me to gather and use new data, and never flagged in their effort to complete the job.

Without so many of my graduate student friends I never would have finished this work: Vince West, Lee Escandon, Adam Hincks, Charles Roddie, and many others. It was my family – Mom, Dad, Chuck, Brent, and Connemara – who made graduate school possible for me, and

who never tired of listening to my research ideas, encouraging me when I was down, and rejoicing with me at the smallest victory. Most of all this work is about my wife, Maggie. I would not have amounted to anything without her, and to her this is lovingly dedicated.

Table of Contents

Introduction

Key determinants of the wages of a typical worker are the demand for labor, and the supply of labor. The demand for labor depends on, among other variables, the supply of other inputs to production. In the first chapter of this dissertation, I demonstrate that schooling experiments may be good natural experiments to study how the demand for one type of labor (namely, adult farm labor) depends on the supply of another type of labor (child farm labor). The supply of labor depends on, among other variables, the preferences of individual workers. Thus, the supply of labor may be heterogeneous across individual workers. Using a new data set of New York City Taxi Drivers, I find that daily income dependence in daily labor supply varies widely across drivers. The results are consistent with – but are not proof of – some drivers having daily income reference dependence. Finally, the demand for labor also depends on the productivity of the worker, which in our economy is often measured at the beginning of a person's career by his or her academic achievement. In the third chapter of this dissertation, I show how the marginal effect of physical appearance on academic achievement may actually be negative when the channels of personality, grooming, and other ascriptive characteristics are controlled for.

Chapter 1. Are Adults in Demand when Children Leave the Land? Evidence from Rural Mexico

1.1. Introduction

What happens to adult labor market outcomes when children are removed from the labor force? The empirical evidence regarding this question is scant, while the policy implications are far-reaching (Galli 2001). According to the International Labor Organization's recent estimates, there are 186.3 million child laborers worldwide (Basu & Tzannatos 2003). If we wish to propose government interventions to reduce child labor and encourage education, then the optimal manner of intervention depends on whether or not children and adults are labor substitutes. Where employers substitute adults for children, an increase in adult wages and/or hours will accompany a decrease in child labor, partially offsetting the short-term welfare loss that families face when some of their children are no longer working. In particular, the work of Basu and Van (1998) shows that in this case there is the possibility of multiple labor market equilibria, with a ban on child labor possibly resulting in a higher welfare equilibrium. But where adults complement children – i.e., where adult wages and/or hours decrease when children leave the workforce – interventions to reduce child labor can seriously harm household welfare, and thus such interventions may need to be accompanied by extensive government programs to make up for this loss.

Indeed, the possibility that in developing countries adult labor complements that of children is not necessarily remote: popular wisdom famously cites the supposed "nimble fingers" of children as a reason why children and adults may be complements in industries such as carpet weaving in India (ILO 1996). In addition, from their empirical work on aggregate production functions, Diamond and Fayed conclude that children and adult men are complements in

Egyptian industry (Diamond and Fayed 1998). Finally, the 2001 survey by Rosalind Galli cites (apparently phenomenological, task-based) evidence that suggests that in household production and agriculture, children complement adults (Galli 2001). However, Galli herself concludes that there is not yet enough good empirical evidence to support either complementarity or substitutability, and she cites this issue as a main gap in the empirical literature on child labor.

Despite the mixed evidence and lack of good empirical studies, governments and international organizations have argued that child labor is a major determinant of adult unemployment, i.e. that children and adults are substitutes. Thus, there is a pressing need for empirical work to address the goals and assumptions of policy makers. Galli states:

> The. . .Child Labor Deterrent Act introduced in the United States in 1993 argued that a worldwide ban on trading goods produced by child labour would benefit the exporting countries practicing child labour through reduced adult unemployment.. . .This idea is not exclusive to the Act, and has been often stated by researchers and by the ILO itself in the book 'Combating Child Labour', where it is asserted that "…child labour is a cause of, and may even contribute to, adult unemployment and low wages …" (ILO 1988: 90). Notwithstanding its popularity, there are very few theoretical and applied studies examining the child labour impact on [the] adult labour market.

In this paper, I address this empirical gap. I suggest, establish – and make use of – the fact that schooling experiments can reduce child labor supply without directly affecting adult labor demand, thus obtaining experimental evidence on the effect of child labor supply shifts on adult labor market outcomes.

I apply this new strategy to Mexico's PROGRESA experiment. I find that a decrease in child field work participation is accompanied by an increase in adult labor demand. This increase was not directly caused by treatment money reaching employers: there were no significant effects on food prices, hectares of land used, or harvest size. Furthermore, the wages of healthy non-treated adults living around children who stopped working also increased, suggesting that treatment-related health increases were not responsible for the wage change.

Thus, employers substituted adults for children. What are the theoretical and practical implications of this result? The theoretical work of Basu and Van (1998) suggests that in the presence of two axioms (the *Luxury Axiom* and the *Substitution Axiom*), the labor market for adults and children may have multiple equilibria. In particular, this suggests that a ban on child labor may result in a higher-welfare equilibrium, where adults' wages are high enough to support their families, and children can afford to leave the workplace for the classroom. The work of Eric Edmonds (Edmonds 2003) shows that in the agricultural setting of Vietnam the Luxury Axiom seems to hold. My results suggest that in this agricultural area of Mexico the Substitution Axiom seems to hold. But how applicable are these results to the typical settings in which children work around the world? As Udry (2004) points out: "Child labor is overwhelmingly a rural and agricultural phenomenon. For example, in Pakistan, 70% of working children are employed in agriculture." Thus, together with (Basu and Van 1998) and (Edmonds 2003), my results suggest the possibility of multiple equilibria in the types of labor markets that most children work in throughout the world.

In Section VIII and Section IX I consider the potential implications for policy makers in more detail. In the next sections, I explain the related empirical literature, introduce my identification strategy, and explain my results.

1.2. Literature Review

There are very few studies of child labor demand, or of employers' elasticity of substitution between the labor of children and that of other age groups. Parameters of labor demand functions are in general difficult to measure: establishment data is rare, and it is not easy to gather it consistently across multiple establishments. This leaves aggregate data or household surveys; but estimates based on aggregate data suffer from simultaneous equations bias, and household surveys measure the decisions of workers, so in either case one needs a reliable exogenous shift in labor supply or wages. With child labor, these difficulties are compounded because of the problems in identifying the employers, the parents, or the children themselves, and because even when identified they may be unwilling to share information about their employment, especially where child work is illegal.

Perhaps because of these obstacles, the literature on the parameters of labor demand interactions across age groups is sparse and permits few generalizations. But a survey by Hammermesh (1993) concludes that the (then) current results suggested that most elasticities of substitution "are quite small, implying that changes in the relative [labor] supply of one group will not greatly affect wages received by workers in other groups." Brown, Deardorff and Stern (2002) report the results of Diamond and Fayed (1998), who estimate aggregate production functions from Egyptian household survey data to conclude that "the elasticity of substitution between children and adult females is. . .quite a high figure," but that "adult male and child labor are complementary." Finally, Ray (2000) claims to test Basu and Van's substitution axiom via household surveys in Peru and Pakistan, but only finds evidence of substitution in the case of adult males and children in Peru[3]. Galli (2001) interprets the existing empirical evidence to conclude: "Whether children actually do substitute [for] adult workers creating adult unemployment and/or reducing adult wage rates remains an open question. . . Further

[3] Ray did *not* test Basu & Van's Substitution Axiom of labor demand, b/c he measured the household's decision to *supply* labor.

qualitative and scattered evidence suggests that in household-based production activities and in agriculture the complementarities between children and adults are stronger."

However, each study in this small set uses either aggregate data (producing estimates that suffer from simultaneous equations bias), or household surveys (which, in the absence of some exogenous shift in labor supply, simply produce estimates of parameters of labor supply).

I circumvent these difficulties by using data from PROGRESA, a randomized controlled experiment performed in about 500 villages in rural Mexico, which exogenously reduced the supply of child labor in treatment villages. I exploit this exogenous shock to child labor in order to estimate the effect of a decrease in the supply of child labor on the demand for adult labor.

1.3. Theory and Identification

In general, a profit-maximizing firm may treat the labor of adults as either a substitute for or a complement to the labor of children (Varian 1999). Empirical evidence can answer which it is by determining the effect of a treatment that changes child labor supply on the demand for adult labor. In this section, I demonstrate in Proposition 1 that if a treatment has increased both the price and quantity of adult labor, then this is sufficient to show that it increased the demand for adult labor – even if the treatment may also have affected the supply of adult labor. The main assumption is that the partial derivative of the demand for adult labor with respect to its price is everywhere less than zero. Finally, I state that if the positive effect of the treatment on the demand for adult labor came only through the pathway of a decrease in child labor supply, then the latter must have caused the former, and hence adults and children must be substitutes. By demonstrating that it is not necessary to find a treatment that had no effect on adult labor supply, I am able to use the PROGRESA treatment to measure the effect of decreases in child labor supply on adult labor demand.

Formally, I define substitution and complementarity as follows:

***Definition**: Substitutability and Complementarity*

Let w_C^i = the wage paid to children in period *i*, and let w_A^i = the wage paid to adults in period *i*, where $i = 1, 2$.

Let $D_A^*(w_A, w_C)$ = a firm's profit-maximizing demand for the labor of adults as a function of the wages paid to adults and the wages paid to children.

Adults *substitute* for children if:

$\forall\ w_C^1$, and $\forall\ w_C^2 > w_C^1$, it is true that $D_A^*(w_A, w_C^1) > D_A^*(w_A, w_C^2)$, $\forall\ w_A$

Adults *complement* children if:

$\forall\ w_C^1$, and $\forall\ w_C^2 > w_C^1$, it is true that $D_A^*(w_A, w_C^1) < D_A^*(w_A, w_C^2)$, $\forall\ w_A$

Intuitively, if the labor supply of children decreases, thus increasing their wages, then the children's employers could either increase demand for adult labor (in which case adults and children are substitutes) or decrease it (in which case they are complements). If there are other inputs in the firm's production function – such as capital – then in fact adults may be substitutes for children while children complement adults. Thus, since the exogenous variation in this paper is over child labor supply, here I will only study whether adults substitute for or complement children, not the converse question.

I will show below that where the following two circumstances are jointly satisfied, it is possible that empirical evidence can help answer this question:

(1) a treatment causes a shock to child labor supply

(2) that treatment has no effect on adult labor demand except through the change in child labor supply.

In order to determine the effect of a treatment that varies child labor on the demand for adult labor, I first must determine the effect of that treatment on the equilibrium price and quantity of adult labor. I assume that the treatment effect on the equilibrium price and equilibrium quantity of adult labor is equivalent to the treatment effect on the average price and average quantity of adult labor. Thus, I am assuming that the average price and quantity of adult labor in treated areas will be at the intersection of the adult labor demand and labor supply curves in the treated areas, while the average price and quantity of adult labor in the control areas will be at the intersection of the adult labor demand and labor supply curves in the control areas. Using the effect of the treatment on the observed price and quantity of adult labor, I can infer with few assumptions the effect of the treatment on adult labor demand. Assume that market demand for adult labor may always be represented by some function of its price, $D(p)$. Let the control group demand for adult labor be $D_{Control}(p)$, and suppose that this function describes the pre-treatment demand in the treatment group as well.

Proposition 1: If, $\forall p$, $\partial D_{Control}(p) / \partial p < 0$, and if $p_{Treated} > p_{Control}$ and $q_{Treated} \geq q_{Control}$, then: $\exists$ a new demand function $D_{Treated}(p)$ such that $q_{Treated} = D_{Treated}(p_{Treated})$.

Proof: Intuition: the demand curve at any point in time is restricted to be nowhere increasing with respect to price. Thus, since the observed equilibrium price and quantity have *both* increased, the post-treatment equilibrium cannot lie on the pre-treatment demand curve. Therefore, the treatment must have *shifted* the pre-treatment demand curve to a new post-treatment demand curve.

Formal Proof: At the control group competitive equilibrium market price, $p_{Control}$, the equilibrium quantity will be $q_{Control} = D_{Control}(p_{Control})$.

We are given that $p_{Treated} > p_{Control}$ and $q_{Treated} \geq q_{Control} = D_{Control}(p_{Control})$.

Then, because $\forall p$, $\partial D_{Control}(p) / \partial p < 0$, it is clear that $p_{Treated} > p_{Control}$ ➔

$D_{Control}(p_{Treated}) < D_{Control}(p_{Control})$.

Therefore, $D_{Control}(p_{Treated}) < D_{Control}(p_{Control}) \leq q_{Treated}$, so $D_{Control}(p_{Treated}) \neq q_{Treated}$.

Thus, there must $\exists$ a new demand function $D_{Treated}(p)$, such that

$q_{Treated} = D_{Treated}(p_{Treated})$. QED.

While it is thus clear that the demand for labor has shifted to a new function, it is not necessarily the case that $D_{Treated}(p) > D_{Control}(p)$, $\forall p$. Thus, such a shift in equilibrium price and quantity proves that demand has shifted in some way, and that in particular it has *increased* for at least one price level. More complete evidence in support of the hypothesis that demand has *increased everywhere* (i.e. for all p, as in the definition of substitutability), would of course consist of other observations of price and quantity along the same new demand curve. If this increase in adult labor demand is caused by a treatment whose only effect on adult labor demand is through a change in child labor supply, then this increase in adult labor demand is evidence that adults and children are substitutes, by the definition of substitutability above.

In the following sections, I describe my empirical strategy to determine whether the PROGRESA experiment: (1) decreased child labor supply; (2) increased the price and quantity of adult labor; and (3) affected adult labor demand only through changes in child labor supply, and not through treatment benefits (direct or indirect) to adults or through treatment money reaching the farms who hired adult labor. Based on the empirical results that verify these three points, and on the proposition proved above, I conclude that the PROGRESA experiment provided evidence that adults and children are substitutes.

1.4. Data

Mexico's Program in Educaciòn, Salud y Alimentaciòn (ProgrESA) or "The Program in Education, Health and Nutrition", was the first large scale schooling experiment in Latin America. PROGRESA was designed to promote education and health in poor rural areas of Mexico. It began with an experimental phase, one of whose primary aims was to determine whether, if payments were made to families conditional on their children's school attendance, school attendance would increase in the treatment group. Census and administrative data identified 506 villages in rural Mexico as "poor" (Skoufias & Parker 2001). Of these villages, 320 were randomly selected to form the treatment group. The remaining 186 villages formed the randomized control group[4].

Five surveys were conducted over households in all 506 villages at the following times: October 1997, March 1998, October 1998, May 1999 and November 1999. In the Spring of 1998, the Mexican government announced that it would give benefits (conditional on the children's school attendance and family participation in health and nutrition programs) to the eligible families of the treatment group. The first payments were made in May 1998. Thus, the first two surveys are pre-treatment, and the latter three surveys are during the treatment. After the experimental phase was complete, eligible families in the control group began receiving benefits as well.

PROGRESA administrators used the results of the October 1997 census to determine, based on variables associated with household welfare, the families that were relatively poor. It assigned these families to the eligible group, assigning relatively well-off families to the non-eligible group (Skoufias, Davis, Behrman 1999). This assignment was conducted for families in both control and treatment villages. Eligible families in the treatment group of villages received

[4] See Behrman & Todd 1999 for a discussion of the randomization procedure.

conditional benefits targeted towards improving education and health[5]. If a child under 18 missed less than 15 percent of the school days in a particular month, then PROGRESA provided a cash award that month to the mother of the child. Cash awards increased to keep pace with inflation, increased with the grade of the child, and were higher for girls than boys. These monthly grants ranged from about 80 pesos for third graders to 280 pesos for ninth grade boys and 305 pesos for ninth grade girls. As a comparison, in 1997 the average monthly salary income of an adult jornalero was about 600 pesos, and that of a child jornalero was about 500 pesos. The program also provided basic health care for all family members and a fixed monetary transfer for nutritional supplements (Skoufias & Parker 2001). Table 1.1 shows a summary of benefits.

I make use of data from this experimental phase of PROGRESA. I obtained the data from the Opportunidades office, which is the new name for the agency that currently runs PROGRESA. The same raw data set that I used to construct my own data set can be downloaded from evaloportunidades.insp.mx/en. I make use of three surveys that were conducted at the same time in the agricultural cycle (October/November): the pre-treatment survey in 1997 and two post-treatment surveys in 1998 and 1999. The 506 villages in the experiment were located in seven Mexican states, shown shaded in Figure 1.1. According to Table 1.2, these village economies were primarily agricultural, and the primary crop in these villages was corn. The primary corn harvest in Mexico lasts from October 1st to the end of December.[6] Thus, I interpret my results as information about production technology and labor demand during the corn harvest. It is of course possible that production technology and labor

[5] The eligibility status was revised in 1998, and according to my data the number of eligible families was higher in 1998 than in 1997 and higher still in 1999.
[6] According to www.tradefutures.cc/education/Corn/worldcornsd.htm

demand are different for corn planting or for the planting or harvesting of other crops in other regions.

Table 1.3a shows some summary statistics across both treatment and control villages for the three years in my sample. In the data sets from all three surveys there is information regarding whether individuals were eligible for the program, whether they were working for a salary, what their job title was, measures of their income, and measures of the amount of time they worked. Table 1.3b shows the distribution of adults and children across job categories listed in the main job category variable that is available each year. There are two job title categories for which workers consistently report salary information: *jornaleros* (field workers), and *obreros* (non-agricultural workers) – those in other categories typically do not report earning a salary. This paper analyzes the jornalero workforce, which has nearly three times as many observations as the obrero workforce (see Figure 1.3) and – given the corn-heavy nature of agriculture in this sample – is presumably more homogenous than the obrero workforce (which seems to potentially include *all* regularly paid non-agricultural jobs). Figure 1.2 shows the age frequency histogram of jornaleros earning a salary. The first shaded area shows the jornaleros I classify as children (ages 16 and under), and the second larger shaded area shows the jornaleros I classify as adults (ages 17 to 59). In 1997, children made up 8.78 percent of the total jornalero workforce, while adults made up an additional 80.22 percent.

Everyone who reports income reports it in *one* of the following measures: pesos per day, pesos per week, pesos per two weeks, pesos per month, or pesos per year. The measures of the amount of time worked are hours per day and days per week, and most people who report income report the amount of time they worked using both of these measures. About 90 percent of the income observations are in pesos per day or pesos per week. For people who report daily salaries, I impute hourly wages by dividing the daily salary by the number of hours worked per day. For people who report weekly earnings, I impute hourly wages by dividing by the number of days worked per week multiplied by the number of hours worked per day. For the remaining 10 percent of income observations, I assume that bi-weekly reporters work both weeks, that monthly reporters work four weeks per month, and that yearly reporters work fifty weeks per year.

The resulting hourly wages range from .0002857 pesos per hour to 7506.25 pesos per hour. With bounds these extreme, it is likely that the very high and very low hourly wages suffer from measurement error. Mean regressions of wages are thus likely to be biased by the incorrect measurements at the top of the distribution, and mean regressions of log wages are likely to be biased by the incorrect measurements at the bottom of the distribution. Thus, in later sections I will perform two tests that do not depend only on means in order to establish the existence and direction of any treatment effect on the distribution of wages: a kolmogorov smirnov test of first-order stochastic dominance; and estimation of quantile regressions by decile. But, once the existence and direction of the treatment effect have been established by the above tests, in order to get one number for the size of the treatment effect, I do run mean regressions as well, attempting to eliminate the bias caused by the incorrect measurements at the top and the bottom of the distribution by dropping observations with wages in the top and bottom five percent for each of the six comparison groups (control vs treatment, 1997 vs. 1998 vs. 1999)[7].

In asking about workers' hours, the surveys asked workers how many hours a day they tended to work last week, or simply how many hours a day they worked. Thus, if workers worked a different number of hours each day, the estimate of the hours per week will be noisy unless the workers correctly averaged their hours when responding to this question. Because of

[7] This cropping is carried out relative to the sample used in each regression (usually, this is all adult jornaleros, but sometimes it is a subsample designed, e.g., to determine the impact of living in a treatment village without directly receiving treatment money).

this, in the analysis below I replicate all mean regressions of hourly variables using daily variables – i.e., with daily income instead of hourly wages, and days worked per week instead of hours per week. This also helps ensure that measurement error in hours is not driving the results.

1.5. Did the Experiment Reduce Child Labor Participation?

In the first few months of the program, as measured by the 1998 survey, it is unclear whether the experiment has yet reduced child participation in the jornalero workforce. But by 1999, 18 months after the program started, the treatment has clearly caused child participation in the jornalero workforce to decline. These results are demonstrated in the difference-in-difference estimates of the treatment effect described below.

1.5(A) Empirical Strategy

My usual empirical strategy in this section and the next is to estimate reduced form equations of the treatment effects on labor market outcomes such as the work participation rate, hourly wages, etc. My unit of observation is an individual at a point in time. As Table 1.3 shows, some characteristics of treatment villages and control villages differed in small but significant ways before the treatment even started, so it is important to use a difference-in-differences approach[8]. This entails a treatment village dummy variable, a post-treatment dummy and an interaction dummy – with the interaction coefficient being the difference-in-difference estimate of the treatment effect. In addition to differencing out the pre-program differences between the control and experimental group, I also control for the effect of composition

[8] Furthermore, in a key PROGRESA paper (Schultz 2004), Schultz argues, "even if the randomization of program placement is not challenged, . . ., the difference in difference estimators are preferred to the post-program differences, because they remove persistent sources of regional variation. . . that might exist."

differences between the two groups by including controls for important personal characteristics[9]. Finally, to ensure that I control for village-specific components of the variance of the error term, I include clustering at the village level in most specifications. Thus, in summary, the difference-in-difference equations are of the following pattern:

$$Y_{i,t} = a \cdot Treated_{i,t} + b \cdot TreatmentVillage_{i,t} + c \cdot Post_{i,t} + d \cdot PersonalChars_{i,t} + \varepsilon_{i,t}$$

where i indexes people and t indexes time.

The *Treated* dummy variable is 1 when the observation is from a treatment village and is also from a post-treatment survey. The *Treatment Village* dummy is 1 whenever the observation is from a treatment village. The *Post* dummy is 1 whenever the observation is from a post-treatment survey. I include in *personal characteristics* dummies for gender, age, schooling, language abilities and marriage status. I run this specification separately for the 1997 vs. 1998 comparison and the 1997 vs. 1999 comparison.

1.5(B) The Decline in Child Jornalero Work Participation

First, I add to the previous studies of this experiment ((Schultz 2004) [10] and (Skoufias & Parker 2001)[11]) that have estimated significant decreases in work participation for children, by estimating specifically the treatment effect on child participation in the jornalero workforce. I

[9] Schultz (2004) explains the logic of this: "It may still be useful to add additional explicit control variables and estimate their marginal effects jointly with those of the program on the enrollment of poor children, because this should increase the statistical power of the model estimated at the level of the individual child to isolate significant effects attributable to the program treatment, if there are any." This is also a justification for making the unit of observation as small as possible in my specifications (usually it is at the level of the individual).

[10] Based on differences between means, Schultz (2004) concludes: "All of the differences in child work between treatment and control populations are negative, as expected, and they are statistically significant at least at the 10% level for the probability of paid work for primary school females and males and for secondary school males, for household and market work for secondary school females, for paid work for secondary school males, for the OLS hours for primary school boys, and for the Tobit hours for primary school females and males and secondary school males" (I deleted references to Schultz's tables in this sentence). He goes on to use more sophisticated IV estimates to further conclude that the program had statistically significant negative effects on child work.

[11] Based on a difference in differences estimate, (Skoufias and Parker 2001) conclude: "The results. . . show that PROGRESA has had a clear negative impact on children's work."

create a dependent variable dummy for working as a jornalero by assigning the dummy the value 1 if the person worked as a jornalero in the last week and 0 if they did not work or worked in a different job category. I regress the dummy for working as a jornalero on my independent variables as outlined in Equation 1. The OLS results are reported in Table 1.4a and summarized in Table 1.4b. I find that by 1998, there was no significant effect on child jornalero field work participation. However, by 1999, child jornalero work participation saw a large and significant decrease due to the treatment. This corresponds with Skoufias and Parker's result that 12 to 17-year-old males (51 percent of whom are jornaleros if they work at all, and who make up 87 percent of the child jornalero workforce) only saw a significant decrease in child work participation by 1999. I also run probit specifications of the same difference-in-difference equations, and report the summarized results in Table 1.4c. According to these results, there is a significant decrease in child work participation by 1998 that grows through 1999, but only the 1999 increase is robust to clustering at the village level.

Thus, while the initial 1998 treatment effects on child labor participation are inconclusive, it is clear that by 1999 child labor participation in the jornalero workforce has significantly decreased.

In the next section, I estimate the treatment effects on the quantity and price of adult labor. I then look for additional evidence to determine whether the decline in child work participation in the fields that I observed in this section was responsible for the change in the demand for adult labor that I observe in the next section.

1.6. Did the Reduction in Child Labor Cause an Increase in the Demand for Adult Labor?

The results in the previous section showed that there was a decrease in child work participation in the jornalero workforce by 1999. Thus I need to check whether the demand for adult labor increased by 1999. According to Proposition 1, if a treatment has increased the price of adult jornalero labor without decreasing its quantity, then this is sufficient to show that it increased the demand for the labor of adult jornaleros. Thus, I check whether by 1999 there was an increase in the price of adult jornalero labor without an accompanying decrease in the quantity. First, I consider the treatment effect on the price of adult labor, and secondly the treatment effect on the quantity of adult labor.

1.6(A) Treatment effects on the price of adult labor:

I estimate treatment effects on adult hourly wages and daily income. As explained in Section IV, I establish the existence and direction of these treatment effects from kolmogorov smirnov tests on the distribution of wages, and from quantile regressions by decile. I then estimate a single number for the size of the treatment effect by following the empirical strategy outlined in the previous section, estimating OLS hourly wage and daily income specifications. These results show that by 1999, there are positive and significant treatment effects on both adult jornalero hourly wages, and daily income.

The kolmogorov smirnov test on the pre-treatment distribution functions show that the pre-treatment distribution of wages in treatment villages is first-order stochastically dominated by that in the control villages[13]. But the kolmogorov smirnov tests clearly show that the post-

[13] The p-value for the null hypothesis that the two distributions are identical – when the alternative hypothesis is that the treatment distribution is stochastically dominated by the control distribution – is 0.02, and is thus rejected. The p-value for the null hypothesis that the two distributions are identical – when the alternative hypothesis is that the control distribution is stochastically dominated by the treatment distribution – is 0.20, and cannot be rejected.

treatment distribution of wages in the treatment villages first-order stochastically dominates that in the control villages[14]. This shift can be seen visually in Figure 1.4, which plots the cumulative distribution functions of the hourly wages of adult jornaleros in 1997 and in 1999. The wage distribution is too lumpy for all deciles to increase, but the quantile regressions by decile reported in Section A2 show that four deciles increased significantly (two below the median and two above) and none decreased significantly.

Thus, it is clear that by 1999 the hourly wages of adult jornaleros have increased due to the treatment. Furthermore, the adult wage increase appears to be real, not only nominal: the 2000 study by Handa, Huerta, Perez & Straffon concludes that the treatment did not produce food price inflation in the treated villages.

What number summarizes the size of this increase? I consider the treatment's effect on mean wages, by estimating OLS regressions on log hourly wages and log daily income according to the difference-in-differences strategy discussed in the previous section, with the effect of the tails diminished via the cropping discussed in Section IV, reporting the results in Table 1.5a. The results suggest an increase in adult jornalero wages of over 6%.

1.6(B) Treatment Effects on the Quantity of Adult Labor:

Having thus established that, by 1999, the treatment increased the price of adult jornalero labor, I turn now to the quantity of labor hired. I estimate treatment effects on mean work outcomes for adult jornaleros between 1997 and 1999. From Table 1.5b, it is clear that the treatment increased both adult hours worked per week and adult days worked per week conditional on working. Table 1.6 shows that it is likely – though not necessarily – true that the

[14] The p-value for the null hypothesis that the two distributions are identical – when the alternative hypothesis is that the control distribution is stochastically dominated by the treatment distribution – is 0.00, and is thus rejected. The p-value for the null hypothesis that the two distributions are identical – when the alternative hypothesis is that the treatment distribution is stochastically dominated by the control distribution – is 0.38, and cannot be rejected.

treatment increased the probability of adult participation in the jornalero workforce as well. I interpret these results to mean that the treatment increased the quantity of adult jornalero labor hired in treatment villages. At the least, these results suggest that it is very unlikely that the quantity of adult labor decreased due to the treatment.[15]

Thus, the treatment increased the price of adult jornalero labor without decreasing its quantity, so by Proposition 1, I conclude that the treatment increased the demand for adult labor. I do not conclude that the treatment had no effect on adult labor supply, but only that any such effects were outweighed by the increase in adult labor demand. For example, if the treatment reduced the labor supply of adults through an income effect[16], then this reduction was clearly outweighed by the increase in demand for adult labor because the quantity of adult labor probably increased. Likewise, the increase in demand for adult labor must have outweighed any increases in adult labor supply, because adult wages increased.

Thus, by November 1999, comparison of treatment and control villages shows a significant decrease in the work participation of child jornaleros, accompanied by a significant increase in the price of adult jornalero labor and no significant decrease in the quantity of adult labor. If the only effect of the treatment on adult labor demand was through the decrease in child labor supply, then these results are sufficient to conclude that adults and children are

[15] All of the specifications in Table 5b show significant increases in adult jornalero hours (conditional on jornalero work participation). Specification (1) of Table 6 shows a significant increase in jornalero work participation as well. But in specification (2) of Table 6, where village fixed effects are replaced with clustering at the village level, the increase in adult jornalero work participation is no longer significant, leaving open the statistical possibility that work participation decreased by a small amount (since the 95% confidence interval of the change in work participation overlaps 0). Thus, if heterscodasticity is being correctly adjusted by village level clustering, and if the true change in adult work participation is on the low end of this confidence interval, and if the large increase in adult hours reported in Table 5b came about *only* because people who would have worked low hours left the workforce, then it is possible that in fact the quantity of adult labor actually decreased due to the treatment. Given the number of conditions that seem to be necessary to conclude that the quantity of adult labor decreased, I believe it is likely that the quantity of adult labor did not decrease.

[16] e.g., labor supply could decrease due to an income effect caused by receipt of treatment money.

substitutes in production: when children became more difficult to hire, employers increased wages for adults, thus increasing both the hours adults worked per week and weekly earnings[18].

1.6(C) Did the decrease in child labor supply cause the increase in adult labor demand?

I now give results which suggest that the only effect of the treatment on the demand for adult jornalero labor was through the decrease in child labor supply. There are three alternative pathways to consider. One is that the treatment families spent their money in a way that would increase the demand for the jornaleros' labor. The second is that the wage increase had something to do with receiving treatment money (e.g. *direct benefits* in income, nutritional consumption, or medical consumption that lead to improved health, leading to better productivity and hence to better wages). The third is that the wage increase had something to do with receiving indirect treatment benefits (e.g. *spillovers* in income, nutritional consumption, or medical consumption that lead to improved health, leading to better productivity and hence to better wages).

The first alternative hypothesis depends on the possibility of treatment money encouraging farm production, causing more adults to be hired. But Table 1.7 and the graphs in Figure 1.5 show that there was no treatment effect on the number of hectares of land used or owned in the treatment villages. Likewise, first differences of the number of tons of corn harvested in treatment households vs. control households in 1998 show that the treatment did not change the size of the harvest (Table 1.8). These results suggest that the treatment did not increase the total amount of field work that employers needed to be done by jornaleros. Eligible families in the treatment villages did buy more of some types of animals, but it is not clear whether the addition of these animals required more labor or less (since horses, e.g., could

[18] I replicate the difference and difference regressions for weekly earnings, finding a large and significant treatment effect on weekly earnings.

substitute for field work) (Angelucci & De Girogi, 2005). Finally, I reported above that there is no evidence of food price inflation due to the treatment[19]. Thus there is no consistent evidence that the treatment money was spent in a way that would increase the demand for jornaleros' labor. This is not surprising, since it seems likely that the markets for basic foodstuffs such as corn are considerably larger in geographic scale (perhaps even international) than those for short-term labor assistance during the corn harvest[20].

The second alternative hypothesis is that something associated with receiving treatment money may have caused the wage increase in treatment villages. An example of a pathway from receiving treatment money to increased price and quantity of labor would be increased nutrition, leading to improved health. Some of the families in both treatment and control villages were not eligible to receive treatment because their wealth was too high, and some did not receive treatment money because of administrative errors. So I can use these families to see if these

[19] Given the agriculture products listed in Table 2, the prices that matter in determining whether the demand for local agricultural goods has increased are (mostly) the price of corn, and (secondarily) the price of beans and coffee. Unfortunately, not every locality reports prices, and (Handa Huerta Perez and Straffon 2000) do not have information on corn (only on corn paste and corn tortillas). Their work shows that the price of beans appears to have increased by similar amounts in both treatment and control villages; that the price of coffee may have decreased in treatment villages and stayed constant in control; that the price of corn paste appears to have increased by similar amounts in both treatment and control villages; and that the price of corn tortillas may have increased by about the same amount in both treatment and control villages, though only the treatment increase was significant. My own regressions show no significant difference between treatment and control prices for corn flour, corn paste, or corn tortillas in the 1999 post-treatment survey used in this paper. This overall evidence is difficult to reconcile with any large positive treatment effect in the price of the crops most local farmers produce. This is not surprising, considering that the above authors believe that government-run Diconsa stores (which are equally distributed across villages) are likely to "maintain a relatively constant supply of basic items at a fixed price," and hypothesize that this should have a stabilizing effect on prices. Furthermore, the authors report that people in outlying communities travel to the municipal centers to receive their benefit checks, and spend money there; thus, people do not always buy goods in the village that they live in.

[20] At the least, the fact that it is more difficult to move people than it is to move corn suggests that distant labor markets would take longer to respond to local wage variation than distant goods markets would for local price variation. Thus, in the short-run, the relevant geographic scale for a labor market should be smaller than that for a corn commodity market, although in the long run international migration shows that labor markets seek to be global as well. This wage increase lasted one season before the experiment was extended to the control group, so one should consider the possible short-term responses, not long term ones.

wage increases only occur for families that receive treatment money, or if other families living in treatment villages also experienced wage increases. If the only families living in treatment villages that experienced wage increases were those that received treatment money, then that would suggest that something associated with receiving treatment money may have caused the wage increase in treatment villages (although it would not necessarily imply this, since the families that received treatment money tended to be poorer).

Thus, in order to rule out that such a pathway is the sole cause of the wage increase, I estimate the same wage regression on a smaller restricted sample. I restrict my sample to all people in the experimental group who were not eligible to receive money in 1997 and did not receive any money by 1999 (this includes people who did not receive money because of administrative error) and a similar sample from the control group (see the Appendix for a description of how these samples were constructed). On this restricted sample, by 1999 there is a 2.2% wage increase due to the treatment, which is significant at the five percent level. Likewise, there is a 2.0% increase in daily income due to the treatment, a 3.6% increase in hours worked per week and a 3.5% increase in days worked per week. The results are reported in Tables A1b and A1c. That the treatment increases the wages on this restricted sample suggests that the results are not dependent on receiving treatment money (e.g. a causal pathway from treatment money to increased nutrition to increased productivity is not responsible for all of the wage increases).

Finally, the above robustness check must itself face a robustness check in the form of the third hypothesis: might treatment spillovers have been responsible for the increase in wages seen in the sample of non-treated adults who were living in treatment villages? To rule out the pathway of treatment *spillovers* leading to health leading to better productivity and wages, I restrict the above sample again by considering in any year only those non-treated adults who report perfect health according to ten criteria.[22] On this restricted sample, I find that the wages paid to healthy adults are again about 2% higher due to the treatment. This suggests health improvements were not necessary for workers to experience the wage increase, only living in a village where child labor decreased.

I therefore conclude that by 1999, a reduction in child jornalero field work participation in the treatment villages of as much as -13% – and hence a reduction in the total jornalero labor force of as much as -1.2% – had a positive and significant hourly wage effect of over 6.0% on adult field workers, which in turn increased adult hours worked per week by 3.5.[23] This result occurs without food price inflation, increased land holdings in treatment villages, or a significant change in the size of the corn harvest, and it is not consistent with shifts in adult labor supply alone. The result does not disappear when I restrict to a much smaller sample that did not receive treatment money, or a subsample of that which includes only the wages of healthy adults. Thus, by Proposition 1, in this region and time period, employers appear to substitute adults for children, not treat them as complements: when child labor supply decreases, the demand for adult labor increases. In the next section, I explore how this exogenous increase in adult jornaleros' wages affected the jornaleros' families

[22] The ten criteria are: days of difficulty performing activities due to bad health in the past month are 0; days of missed activities due to bad health in the past month are 0; days in bed due to bad health in the past month are 0; yes, I can currently perform vigorous activities; yes, I can currently perform moderate activities; yes, I can carry an object of 10kg 500meters with ease; yes, I can easily lift a paper of the floor; yes, I can walk 2 km with ease; yes, I can dress myself with ease; I have had no physical pain in the last month.

[23] The relative magnitudes of these percentage changes depend on the specification.

1.7. How did the Wage Increase Affect the Families of Adult Jornaleros?

Almost half of the families living in treatment villages that did not receive treatment had an adult jornalero in their household. Since the wages of these adult jornaleros increased, any health and nutrition spillovers to non-treated households in treatment villages should be larger for the subset that included adult jornaleros. Indeed, I find in Table 1.9 that consumption spillovers in grains and cereals, and meats and dairy occur only in families that had an adult jornalero. Families that did not have an adult jornalero apparently experienced no significant consumption spillovers in this sample. I also find in Table 1.10 that families with an adult jornalero saw increases in health, although it is not clear if these health spillovers differ from those experienced by families without an adult jornalero. Comparison of the consumption results at least suggests that the increase in the demand for adult jornalero labor had positive consequences for the welfare of their families.

1.8. Interpretation of Results

These results, which confirm the substitutability of adults and children, have theoretical implications that should be of interest to policy makers. For example, in the 1998 model by Basu & Van, child and adult labor are assumed to be substitutes in production. When the labor demand schedule is not too elastic or inelastic, there exist multiple equilibria. In this case, a government intervention such as an implemented ban on child workers or a raise in the minimum wage of children could move the economy from one equilibrium with child labor to a higher-welfare equilibrium without child labor[24] (Basu & Van 1998) (Basu 2000). Thus, the

---

[24] In the presence of multiple equilibria in Basu and Van's model, a minimum wage w' will eliminate child labor if the child market wage < w' < adult market wage, and if child productivity is low enough such that there exists excess demand when only children are working.

observed substitution of adults for children raises the possibility of a welfare-maximizing ban on child labor.

In the absence of multiple equilibria and the specific assumptions of Basu and Van's model, what do these results imply about the welfare consequences of a ban on child labor? Without a precise theoretical framework, any estimates of the size and consequence of these effects will be conjectural at best. In this data set, the average family has 2.5 children under age 17, of which 0.13 are working as jornaleros. If the government implemented a ban on child labor in the jornalero workforce, the lost earnings per week of the children in such a family would be approximately 0.13 * (125 pesos per week) = 16.6 pesos per week. But of the 2.4 adults in an average household, 0.68 tend to work as jornaleros, and (based on the coefficient in Table 1.4b, and the bottom of the confidence interval for the coefficient in Table 1.5a) it seems that these workers might experience an hourly wage increase of 100% * 4.7% /13 % = 36%. Assuming as a lower bound that there was no increase in adult hours worked per week, this would lead to an increase in weekly earnings of 36% * (131 pesos per week) = 47 pesos. This would then entail an increase in adult earnings per week in an average family of (0.68 adult jornaleros per household) * (47 pesos per adult jornalero) = 32 pesos. Thus, 100% of the child earnings lost by the ban would be recovered by the improved adult wages. It is unlikely that the adult wage increase that I observed associated with a 4 % to 13% drop in child labor participation is representative of the new equilibrium wage that would occur in the event of a 100% drop in child labor supply. But this calculation gives some indication that the order of magnitude of the observed wage increase is large enough to potentially counteract much of the welfare loss for poor families due to a ban on paid child labor in the fields.

The distributional consequences of the substitution of adults for children depend on how child workers and adult workers are distributed across families. In families where adults work in industry A and children work in industry B, a ban on child work in industry B will not necessarily lead to higher wages in industry A, and thus the welfare consequences for that family are likely to be negative. Likewise, in a family where children do not work, and adults work in industry B, a ban on child work in industry B will lead to an increase in the adult wage in industry B, improving welfare unambiguously for that family. Thus, even when adults substitute for children in every industry, in order for labor market outcomes of adults to mitigate the welfare losses across all families due to a ban on child labor, it must be the case that either (1) the ban on child labor is successfully implemented across all industries, and/or (2) there is a perfect correlation between the industry of employment of adults and that of children within a family. In the PROGRESA data, there are many households with jornalero adults that are without jornalero children, as well as many jornalero children living in households without jornalero adults, which suggests that the first condition must be kept in mind by policy makers.

1.9. Conclusion

These results demonstrate that when the opportunity wage of not working increased, child workers responded by decreasing their labor participation rates. I rule out alternative pathways to conclude that this reduction in child labor participation is what resulted in an increase in the equilibrium price and quantity of adult labor. Thus, in these areas of rural Mexico during the autumn corn harvest, adult labor substitutes for child labor. The partial elasticity of adult hourly wages with respect to child work participation is clearly negative. Finally, families that were not paid to send their children to school, but that had family members who experienced the wage increase, saw an increase in their consumption of fruits and vegetables as well as meats and dairy products.

The first implications of these results are theoretical. Models such as those of Basu and Van (1998), and Ranjan (2001) – which assume that child and adult labor are substitutes – are reinforced by my result. Indeed, in the context of Basu and Van's 1998 model "The Economics of Child Labor," this paper's update of the previous empirical results – which had showed ambiguous effects of changes in child labor supply on adult wages – is very useful. By providing evidence for their labor demand assumption (the "Substitution Axiom"), the result of my paper reinforces the theoretical possibility of multiple equilibria introduced theroetically by Basu and Van. Since Basu and Van's child labor supply assumption (the "Luxury Axiom") has been supported by recent empirical evidence from another agricultural region, my result helps close a remaining empirical gap (Edmonds 2003).

Furthermore, these results are useful to policy makers, because they suggest that in environments similar to the one observed here (corn-based agriculture), efforts to reduce child labor may have positive impacts on adult wages and employment.

Finally, this paper is the first *experimental* estimate of labor demand parameters across labor input types. The idea of this paper can be easily applied to the many other schooling experiments recently conducted in Latin America and in other nations in the developing world[26], thus showing how these results vary across regions, time, level of industrialization, and cultures. The results here provide a useful estimate of the medium-term effects of child labor reduction on adult labor market outcomes in the agricultural sector, and serve as a guide for policy-makers.

[26] For other experiments see, e.g., Janvry & Sadoulet 2005

Chapter 2. The Existence and Position of Daily Income Reference Points: Implications for Daily Labor Supply

2.1. Introduction

Standard neoclassical theory says that daily consumption of goods and leisure determines daily utility; together with other realistic assumptions this implies that for most workers transitory increases in one day's income should not decrease that day's labor supply. But if some workers have daily utility that depends on daily income directly, and if their marginal utility drops quickly at certain levels of daily income – in other words, if they have daily income reference points – then for these workers transitory increases in one day's income could decrease that day's labor supply.

So I ask: are there daily income reference points in daily labor supply? And if there are, what determines their position? The first of these questions has been met with conflicting empirical evidence, while the second has not (in this author's knowledge) been examined empirically. In this paper, I attempt to provide the best current answer to the first question in order to propose original evidence regarding the second.

Concentrating on taxi drivers, what I find suggests that while (as Farber (2005) found) reference points are not important on average, they are important for almost half of the drivers, even using the more robust techniques Farber introduced. Such reference-dependent behavior could be unrelated to income expectations, or it could be determined by them (Rabin & Koszegi 2006). The fact that these drivers neither work fewer hours nor earn the same income after an exogenous permanent hourly wage increase suggests that their income reference level increased when the fare increased, thus providing the first empirical evidence consistent with the expectations-based theory of Rabin & Koszegi (2006).

I rely on a new panel data set of New York City taxi drivers to estimate their labor supply response to variation in daily income and other variables throughout their shifts. These data offer unique advantages. First, drivers can choose their own hours within every twelve hour shift, making their optimal labor supply responses less likely to be muted by the constraints on working hours common to other occupations. Second, this data spans a permanent exogenous 26% fare increase instituted by the New York City Taxi and Limousine Commission on May 3rd, 2004.

What are the possible ways that daily income could affect the daily labor supply decision of a taxi driver? It is unreasonable to believe that taxi drivers' consumption each day is mechanically tied to what they earn that day: for example, drivers obviously face no physical impediment to smoothing consumption over periods at least as long as a month when their apartment rents are due. Thus, the clear prediction of a simple intertemporal utility model is that as the hours worked during a day increase, the utility from stopping work should increase as well (because the leisure able to be consumed that day has declined), but as the income earned during a day increases, the utility from stopping work should not change very much[1]. On the other hand, Camerer et al. (1997) and Chou (2000) suggest a different reference-dependent model: some drivers may have daily utility that depends on daily income, with a "marginal utility of [daily] income [that] drop[s] substantially sharply around the level of average daily income" (Camerer et al. 1997). The prediction of this reference-dependent model is that as income increases, the probability of stopping work and going home should increase as well. What is the best way to verify which drivers, if any, behave in a way consistent with which model?

[1] This is assuming that the income earned so far during the day tells the driver nothing about the wages that could be earned later that day. If the income earned so far that day is a good predictor of the wages to be earned later that day, then the clear prediction is that the higher the income, the lower the probability of stopping.

There are two approaches to finding the effect of daily income on daily labor supply, each of which makes use of both income and hours variables, albeit in different ways. If the hourly wage is roughly constant throughout a shift, then one could regress the hours worked during a shift on the mean hourly wage (income divided by hours) during that shift, following Camerer et al. (1997). If, however, the average hourly wage changes greatly throughout a shift, then it would be difficult for drivers to base their labor supply decisions on the average wage they have experienced so far that day. Since my data show little autocorrelation in hourly wages within a shift, I follow the technique of Farber (2005), estimating the hazard rate of stopping work for the day as a function of both the hours worked so far that day, the income earned, and other variables that account for effort and earnings opportunities. This estimation strategy allows for income and hours to vary throughout the work day without assuming that the average of income over hours is what drives the decision of the taxi driver at any point in his shift.

When I estimate one function over all the drivers – with the only room for heterogeneity being driver-level fixed effects and driver-level clustering – I find, along with Farber, that daily hours are an important determinant of the probability of stopping work for the day, but that daily income is not. In particular, I find that on average daily income has a very small positive effect on the probability of stopping work. Thus, the average results do not provide any evidence that definitively points towards the existence of income reference points. But, with more data per driver than in previous taxi studies, I am able to estimate sixty-five individual labor supply functions. These estimates demonstrate that roughly half of the drivers have no significant effect of income on stopping probability, but the other half have a large, positive and significant effect.

Two pieces of ancillary empirical evidence are also important. First, drivers with larger estimated income coefficients have smaller inter-day variation in daily income. This suggests that the drivers with larger estimated income coefficients are more likely to restrict their daily income to a small set of values, which is consistent with having a marginal utility of daily income that drops suddenly around a particular value. Second, the average of the positive and significant income coefficients is large enough to imply that if a driver found a $20 bill in his taxi before his first trip, he would end the day early, with only $6 extra. This kind of negative labor supply elasticity – assuming, as I discuss in Section III, that leisure utility is not convex – implies a substantial drop in the marginal utility of income at a particular value of daily income. Thus these drivers may have a utility function that also depends on daily income (as opposed to only on daily consumption), and this utility function may have a substantial "kink" at which the marginal utility of income suddenly decreases. In other words, these drivers may have daily income reference points.

What determines the position of these reference points? It is possible that these income reference levels are unrelated to the prevailing hourly wage rate, and that they remain constant in the face of permanent hourly wage changes. In this case, a permanent wage increase would result in a decrease in the number of hours worked per day, with the income earned per day staying constant. Thus, if I observe that the drivers did not decrease their hours worked per day and did not hold constant their daily income in response to a permanent hourly wage increase, then this implies that the income reference level did not stay constant.

I find that the reference-dependent drivers neither worked fewer hours nor earned the same amount after the exogenous 26% fare increase of May 3rd 2004. In fact, their daily income increased by about the same amount as the fare increase. This suggests that when earnings opportunities increased, the reference income levels increased as well. This provides some evidence consistent with the expectations-based reference dependence theory of Rabin & Koszegi.

In the next sections I discuss the related work on daily labor supply, the conceptual framework for the empirical model of taxi drivers' labor supply, the details of this paper's empirical strategy and data set, the results, and finally their implications. Furthermore, I discuss in more detail the labor supply response of the taxi drivers to this exogenous fare increase, and the implications for future study.

2.2. Background

Is there any reliable empirical evidence for the existence of income reference points in daily labor supply, and if not, what technique and data sources would most reliably find evidence for such behavior if it exists?

An individual worker's daily reference income level can most obviously reveal itself in that worker's *daily* decisions; in the case of labor supply, this would be his daily labor supply. Several data sets of people (bicycle messengers, stadium vendors and taxi drivers) who make such regular labor supply decisions have been gathered recently. Below I review the study of Oettinger on stadium vendors, the study of Fehr & Goette on bicycle messengers, and the studies of Camerer et al. and Chou on taxi drivers.

2.2(A) Stadium Vendors:

Oettinger (1999) makes use of data on all 127 stadium vendors who had the opportunity to work at any of the 81 home games of a baseball team during the 1996 season. He uses this data to estimate the wage elasticity of daily work participation. He estimates both (1) individual labor supply (by first estimating a reduced form probit model of work participation as a function of observables, then estimating log earnings in a way that accounts for selection bias in observed wages, and finally using the latter estimate to predict the potential earnings that vendors face in choosing their participation) and (2) aggregate labor supply through relying on plausible demand shifters. The results demonstrate that vendors are significantly more likely to decide to work a game if the wage that they could earn during that game is high, with estimated wage elasticities in the 0.55 to 0.65 range. Furthermore, he finds that when he ignores the endogeneity between price and quantity in the aggregate model, the elasticity estimates are severely downward biased.

2.2(B) Bicycle Messengers:

Fehr & Goette (2005) report the results of an experiment in which the treatment consisted of informing bicycle messengers that for one month (and one month only) they would earn higher commission rates. They find very clean and clear experimental evidence that the bicycle messengers supplied considerably higher hours of work during the one month commission increase. They also did not accomplish as much per hour during this month. But this decline in effort or performance per hour was far outweighed by the increase in hours. Thus, there was a large increase in the total amount of work done during the month of the commission increase.

While these results are consistent with either a standard neoclassical model with utility non-separable across shifts (days), or with a daily target earnings model (an extreme case of income reference dependence), Fehr and Goette run an additional experiment that suggests the latter explanation. In the second experiment they ask the messengers to choose lotteries in a way that determines their degree of loss aversion. They then compare the two experiments and find that the degree of loss aversion determined from the latter experiment is associated with the degree of negative elasticity in effort per shift in the first experiment. This suggests that the loss averse messengers may have income targets for each shift that they work[2].

[2] An interesting way to further test this hypothesis would be to look for other evidence that might support the alternative conclusion that the time horizon the messengers used for their labor supply decisions was actually much longer than one shift. For instance, one could test if the messengers decreased their labor supply after the future wage increase was announced but before it was put into

2.2(C) Taxi Drivers:

In Camerer et al. (1997), the authors introduce the literature on people who can choose their own hours of work by making use of a data set on New York City taxi cab drivers. They find that regressing the hours worked during a shift on the mean wage during that shift (earnings divided by hours) consistently results in large, negative and significant estimates of the wage elasticity of daily hours. They consider many alternative explanations before concluding that the most likely explanation for this result is that drivers (especially inexperienced ones) are making decisions with respect to a daily time horizon and have kinks in their daily utility as a function of income. Chou (2000) finds similar estimates in the case of taxi drivers in Singapore, making a similar conclusion about their cause.

2.2(D) Comparisons:

In both the stadium vendor study of Oettinger, and the bicycle messenger study of Fehr & Goette, periods of short-term high wages are associated with a positive response in labor supply. However, the taxi driver literature has found some results that point in the opposite direction: in the studies of Camerer, et al.. (1997) and Chou (2000), high daily wages are associated with low daily hours. What can explain this discrepancy?[3]

---

place. To my knowledge, the timing of announcements, measurements, and wage changes precludes such a test using data gathered from the Fehr and Goette experiment.

[3] It is important to note that in the Fehr and Goette (2005) study, the authors found evidence for target earning, but they believed that these income targets were for each *shift* of work. Since the commission increase lasted for a much *longer period than one shift* (namely, one month), in which messengers could vary their number of shifts fairly easily, the authors were thus examining the response of people who may have a very short time horizon when they are faced with a wage increase that lasts much *longer* than their time horizon. Camerer et al. also believed that their drivers may face a time horizon the length of one shift. But they believed that the wage variation these drivers experienced would last only one shift as well. So they were estimating the response of people who may have a very short time horizon when they are faced with a wage increase that lasts the *same length* as their time horizon. Given that the two studies were thus estimating two different types of responses, it is not surprising that the results were different. Indeed, the difference is consistent with a simple target earnings model: Fehr & Goette (2005) explains why a person with an income target for each shift may be more likely to participate in any shift of work during a month in which the wage (or commission rate) is high.

Oettinger suggests that there may be unmeasured common shocks to the supply of taxi drivers that cause the above taxi driver estimates to be inconsistent. Fehr & Goette make the same criticism, adding that there is also the possibility of a selection effect: "higher wages may induce cab drivers to work a few hours on days when they would otherwise not have worked." A more pointed criticism is that of Farber (2005), who suggests specific econometric and interpretive problems with the Camerer et al regressions, and estimates a more robust model on his own newer data set, finding different results from those reported in Camerer et al.

Apart from the econometric issues, Farber specifically points out that when the wage the driver faces later in the day is not strongly related to the wage earned earlier in the day, it makes little sense to estimate hours worked as a function of the average hourly wage, as Camerer et al did. In such a case, it makes more sense to estimate the probability of stopping after each trip as a function of income and hours. Farber uses the data from Camerer et al's paper to show that the autocorrelation in hourly wages in their sample is quite low. Thus, Camerer et al. would have more reliable results if they estimated the hazard model of stopping on their data. Farber (2005) reports the results of such estimation on his own data set.

The results of Farber's analysis are telling: there is strong evidence that daily hours are an important determinant of the stopping decision, but that, on average, daily income is not. Farber concludes that these results are "consistent with there being substantial opportunity for inter-temporal substitution in labor supply and [that they provide] no support for the hypothesis that taxi drivers are target earners." If one believes that the hazard model approach gives more reliable estimates in general than that of Camerer et al., then Farber's results suggest that the previous taxi literature's estimates may be unreliable.

What then, do the results from this literature tell one about the possibility of finding reference-dependent drivers in this study? Farber's results suggest that on average, there is little evidence for reference-dependent behavior. But his results leave open the possibility of a high degree of heterogeneity, such as a situation in which half of the drivers have daily income effects that are insignificant, and the other half have significant income effects. Indeed, among the five drivers for whom Farber has enough data that he can estimate separate stopping functions, two seem to have positive and significant effects of income on stopping probability. Thus, it makes sense to test if a much larger sample shows consistent evidence of heterogeneity in drivers' stopping functions. Furthermore, Fehr and Goette's study and Camerer et al.'s study both suggested that reference-dependent behavior – if indeed it exists – is concentrated in sub-categories of their samples, making heterogeneity in stopping functions a likely possibility in my study.

In the next section I discuss the conceptual framework behind the estimation of taxi drivers' labor supply that I will use to determine whether this potential degree of heterogeneity in fact exists. If I can find a substantial number of drivers who appear to have reference-dependent behavior, then I can test whether an exogenous wage increase causes the hours worked to decrease, as would be predicted if the reference income level was unrelated to wage changes.

3.3. Conceptual Framework

First, I consider sufficient conditions for a model to predict that higher daily income would lead to lower daily labor supply, concluding that daily reference income levels are sufficient. Second, I consider alternative theories about what determines the daily reference income level.

3.3(A) In what Model does Higher Daily Income Lead to Lower Daily Labor Supply?

Based on the treatment in Cahuc & Zylberberg (2004), Chapter 1, Section 1.3, I consider a taxi driver who makes choices over a life cycle, represented by a sequence of days *t*, with labels ranging from $t = 1$ to $t = T$. I assume that the taxi driver's lifetime utility depends on his consumption of goods and consumption of leisure each day throughout his entire life. Because a general utility function over the consumption of goods and leisure over all periods is intractable, I assume that his utility function is temporally separable. I thus write utility as: $\sum_{t=1}^{T} \mathrm{U}(C_t, L_t, t)$, where C_t is the taxi driver's consumption of goods on day *t*, and L_t is his consumption of leisure on day *t*.

This taxi driver can save the money he makes in one period and spend it in another, or can borrow against future earnings. The real interest rate between time *t* and time *t-1* is denoted r_t. Each day, the taxi driver has an endowment of time equal to 1, and spends L_t of this time in leisure, with the remainder spent earning money. Following Cahuc & Zylberberg, I can write the taxi driver's wealth on day t in terms of his assets (A_t), his non-work-related income (B_t), his consumption (C_t), and the work income earned that period (which is a function of the time spent working: $\mathrm{y}_t(1-L_t)$)[4]:

$$A_t = (1+r_t)\cdot A_{t-1} + B_t + \mathrm{y}_t(1-L_t) - C_t \text{ for all } t \geq 1 .$$

When the taxi driver maximizes his lifetime utility, he must do so subject to his lifetime budget constraint. Writing the Lagrangian for the taxi driver's maximization problem, I put the Lagrange multiplier at time *t*, v_t, in front of each term of this budget constraint. Thus, I obtain:

[4] Thus, as in Farber (2005), $\mathrm{y}_t(1-L_t)$ represents the earnings during day *t* as a function of hours of work during day *t*.

$$\mathscr{L} = \sum_{t=1}^{T} \mathrm{U}(C_t, L_t, t) - \sum_{t=1}^{T} v_t \cdot \left[A_t - (1+r_t)A_{t-1} - B_t - \mathrm{y}_t(1-L_t) + C_t\right]$$

I can solve the maximization problem by obtaining the first order conditions:

$$\partial\,\mathrm{U}(C_t, L_t, t) \big/ \partial C \equiv \mathrm{U}_C = v_t \qquad \text{(Equation 1)}$$

$$\partial\,\mathrm{U}(C_t, L_t, t) \big/ \partial L \equiv \mathrm{U}_L = v_t \cdot \mathrm{y}_t'(1-L_t) \qquad \text{(Equation 2)}$$

$$v_t = (1+r_{t+1}) \cdot v_{t+1} \qquad \text{(Equation 3)}$$

What is v_t? v_t represents the addition to utility U obtained from relaxing the budget constraint by one unit. Thus, v_t is the marginal utility of wealth during day *t*.

Combining the first order conditions, it is clear that:

$$\mathrm{y}_t'(1-L_t) = \frac{\mathrm{U}_L}{\mathrm{U}_C} \qquad \text{(Equation 4)}$$

As Cahuc & Zulberberg assert, "limiting ourselves to interior solutions, the optimal consumptions of physical goods and leisure are implicitly written in the following manner."

$$C_t^{\text{optimal}} = \mathrm{C}(\mathrm{y}_{\mathrm{t}}'(1-L_t), vt, t) \qquad \text{(Equation 5)}$$

$$L_t^{\text{optimal}} = \mathrm{L}(\mathrm{y}_{\mathrm{t}}'(1-L_t), vt, t) \qquad \text{(Equation 6)}$$

What can the above conceptual work tell one about how a regression of the probability of stopping work for the day after a given trip will depend on the hours worked so far that day, and the income earned that day? It is reasonable to assume that U_L should be higher when leisure hours L_t are lower, in which case it is clear that the longer a driver works on a given day, the higher U_L is, and hence the greater the probability that $\mathrm{U}_L/\mathrm{U}_C$ will equal or exceed the rate of income generation – thus, the higher the incentive to go home. Thus, the model predicts that the more hours of work a driver has put in, the higher the probability of stopping work and going home.

How will L_t^{optimal} depend on the income earned so far ($\mathrm{y}_t(1-L_t)$)? Equation 6 shows that there are two variables that L_t^{optimal} will depend on: y_t', and v_t, either of which y_t might theoretically have an impact on. If y_t is a good predictor for y_t', then high y_t implies high y_t', which implies that L_t^{optimal} should be lower[5]. Thus, in this case there should be a negative relationship between the income earned so far and the probability of stopping. If, on the other hand, there is no relationship between y_t and y_t', then y_t' does not provide an avenue for any effect of y_t on stopping probability.

It is thus clear that, subject to some empirically verifiable assumptions, y_t might have a negative impact on stopping probability. But is it possible for there to be a positive effect of y_t on stopping probability? I have already considered the pathway of y_t affecting y_t', and, looking at Equation 6 , the only other possible pathway is y_t affecting v_t. How can $\mathrm{y}_t(1-L_t)$ affect v_t, the marginal utility of wealth at time *t*? Following Cahuc & Zylberberg, "successive iterations of the logarithms of . . .[Equation 3]. . . entail:

$$\ln(v_t) = -\sum_{\tau=1}^{t} \ln(1+r_\tau) + \ln(v_0) \qquad \text{(Equation 7)}$$

Cahuc & Zylberberg assert that the value of v_0 should depend on the lifetime earnings of the taxi driver. Thus, a change in earnings on day *t* is unlikely to have a big effect on v_0, and hence is unlikely to have a big effect on v_t.

---

[5] It implies this because of Equation 6 and the assumption that U_L should be higher when leisure hours are lower.

Therefore, there are several clear predictions of this model: (1) The number of hours worked so far in a given day should have a positive impact on U_L, and hence a positive impact on the probability of stopping after any given trip. (2) The income earned so far in a given day can have a negative impact on stopping probability if a high rate of income generation so far implies a high (potential) rate of income generation later that day. (3) The amount of income earned so far in a given day should have no impact on stopping probability if the rate of income generation so far tells the driver nothing about the rate of income generation later that day.

None of the above predictions show a positive impact of daily income on stopping probability. However, in later sections I will show that on average daily income has a small positive impact on the probability of stopping after any trip, and for slightly less than 50% of the drivers, daily income has a large positive impact on the probability of stopping after any trip. Hence, I here consider what modifications to this model would allow for this large positive impact.

It is clear that there are two reasons that the model above does not predict that daily income would positively affect daily stopping probability: that daily utility depends only on daily consumption of goods and consumption of leisure, and that the ability to buy daily consumption depends not on daily income but rather on income earned over a long time horizon. Because of consumption-smoothing over the time horizon, daily consumption in the above model is not tied closely to variation in daily income, so variation in daily income has a difficult time having a strong effect outside of its possible association with a daily wage.

Thus, one modification to the above model would be for daily utility to also depend directly on income earned over some short period of time. How short need it be? A large positive effect of daily income on stopping probability suggests that people will quit early on good days. In fact, later I will present evidence that some drivers will quit so early that they would erase (or almost erase) any extra income dropped in their taxi before the start of their day. Camerer et al. (1997) argues that for such negative elasticities to occur, daily labor supply must be strongly related to *daily* income, not even bi-daily or tri-daily income[6]. A second modification to the above model would be for the marginal utility of daily income to decrease suddenly after a certain level. To see why this helps create negative labor supply elasticities, consider (as do Camerer et al.) a driver facing a wage *w*, who works *h* hours on a given day in order to generate daily income $y = h*w$ and consume daily leisure $L = 24 - h$. Assume that the driver's additively separable daily utility function is $v(y) + u(L)$, with concave $v(\cdot)$ and concave $u(\cdot)$. Then the elasticity equation is $(dh/dw)\cdot(w/h) = (1 - y\cdot r(y))/(y\cdot r(y) + h\cdot r(L))$, with $r(y) = -v''(y)/v'(y)$ and $r(L) = -v''(L)/v'(L)$. As Camerer et al. note: "For $u(L)$ concave ($r(L) > 0$), the elasticity becomes negative for $r(y) > 1/y$ (e.g., more concave than log utility). The elasticity becomes increasingly negative as $r(y)$ gets larger, but does not reach -1 unless $r(y)$ become[s] infinite (corresponding to a kink at the income target reference point . . .). . ." (Camerer et al. 1997).

The authors note that if leisure utility is convex, or if leisure is a strong complement, then it is possible to have highly negative labor supply elasticities without the kink derived above. The most reasonable explanation for why daily leisure may be a strong complement to income in general would be that people may be able to experience better leisure if they can consume more goods while they relax. But given that daily consumption is unlikely to be determined by daily income, it seems unlikely that *daily* income could be a strong complement to daily leisure. Thus, assuming leisure utility is not convex, the most reasonable explanation for a

[6] The Camerer et al. (1997) argument relies on their assertion that even a two-day time horizon would result in drivers working more on "good" days for a wide range of possible specifications.

strong negative relationship between the probability of continuing work and the income earned that day is that daily utility is kinked at some level of daily income.

Thus, modifications to the above model that would allow for large positive impacts of income on stopping probability would be: (a) to make daily utility depend on daily income, and (b) to give drivers kinks in utility as a function of daily income, at which the marginal utility of income suddenly drops.

3.3(B) What Determines the Daily Income Reference Level?

The above section demonstrates that daily income reference levels are sufficient to cause a negative relationship between daily income and daily labor supply. But what determines the reference income level? One possibility is that the reference income level is determined independently of income expectations, and therefore independently of prevailing wages. It is clear that in this case, were the wages to increase permanently, such drivers would work fewer hours each day, earning the same income each day as before. However, it is intuitively appealing that the reference income level would be somehow tied to the income a driver may expect to earn on a given day, and indeed the theory of Rabin & Koszegi (2006) formalizes this intuition. It seems clear that in this case, drivers may work no fewer hours after a permanent wage increase, as their income expectations rise to meet the higher wages; furthermore, they would certainly not hold their income constant.

I have now considered conceptually various pathways through which income and hours could impact a driver's daily labor supply decision, as well as possible explanations of the position of any daily income reference levels. How can I study these issues empirically, using taxi driver trip sheets? Based on this conceptual foundation above, the empirical strategy that I introduce in the next section is relatively simple.

3.4. Empirical Strategy

As I explained in the introduction, because there is little evidence for the existence of a well-defined or predictable "wage" throughout the day (for the evidence, see Section V), I follow the technique of Farber (2005) in viewing the daily labor supply decision of a taxi driver once he has started his shift as the result of a choice after each trip of whether to continue working or to go home. What variables are likely to affect the difference between the utility from going home and the utility from continuing to drive? As in Farber (2005), I assume that for after any trip of any shift, the driver's decision of whether to continue working or to go home is determined by the difference between the expected utilities associated with these two options.

What variables should affect this difference in utilities? Any observables which affect the utility from going home or the utility from continuing to work should be included. The conceptual work above suggests that income and hours worked so far that day should definitely be included. In addition, other things that can affect the difference in utilities include characteristics of particular days of the week or times of the day, as well as potentially the number of trips taken so far (if there is a fixed effort cost per trip).

The conceptual foundation above thus suggests that an accurate empirical model of an individual driver's labor supply decisions would be a probit model in which the dependent variable is the probability of stopping work after any trip, and the independent variables are daily income, daily hours, and the characteristics of the current shift, current day of the week, and current time of the day[7]. In estimating this model, I would be largely following Farber (2005) (and, in part, Farber (2008)), allowing my results to easily be compared with his. Furthermore, I have shown above that different assumptions about the true model of taxi driver behavior can

[7] It would be a probit model assuming that there is also a random component of the difference in utilities between going home and continuing to work, and that this random component is distributed normally about 0.

make different predictions about the effect of daily income on stopping probability, so my estimates may help distinguish between models. Finally, this estimation strategy is simple enough that it allows me to potentially run it separately for each driver in my sample. This allows for a complete analysis of potential heterogeneity in drivers' labor supply functions.

Thus, my empirical strategy is to estimate a probit model of stopping probability as a function of daily income, daily hours, the number of trips completed so far, and characteristics of the current time of day, day of the week and type of shift. I estimate this model separately for each driver for whom I have a sufficient number of observations, and I look in particular for heterogeneity in the effect of daily income on the stopping probability. The different specifications of the probit model of the probability of stopping work after any given trip are summarized by the following equation:

$$\Pr(Stop_{i,j}) = F(\alpha \cdot Income_{i,j} + \beta \cdot TimeWorked_{i,j} + \delta \cdot Controls_{i,j} + \varepsilon_{i,j}) \quad \text{(Eq. 8)}$$

where:

i indexes shifts, and *j* indexes the trips within each shift.

Stop = 1 if the driver stops working after this trip and ends his shift, and *Stop* = 0 otherwise

Income = either the income made so far in this shift in dollars, or a set of fifteen indicator variables for different levels of income

TimeWorked = controls for number of hours worked so far; either in levels, in log form, or flexibly, as indicator variables.

Controls = indicators for the number of trips made so far (or the log of the number of trips), the hour of the day, the day of the week, the month of the year, and whether or not this shift is a day shift.

I also include driver fixed effects in every specification in which I include multiple drivers, and in such specifications I check the robustness to driver-level heteroskedasticity through clustering the standard errors at the driver level. My specifications are similar to those employed in Farber (2005) on a smaller data set, and in Section VI I compare our findings. In the next section I discuss in more detail the data that I used to carry out this strategy.

3.5. Data

3.5(A) Characteristics of Data

The novel data set I introduce in this paper consists of 8,062 taxi driver trip sheets (each covering one shift of work), comprising 175,063 trips. These shifts were undertaken by 114 drivers, 92 of whom report more than 100 trips (or about five shifts of work). A total of 65 drivers report more than 1,000 trips (or about 50 shifts of work), and therefore I have a large number of drivers who report working on many days. Over 95% of the data is from the year 2004, with the remaining from 2003 and 2005. Of the year 2004 data, over 99% is from January through June. I have almost every day of work for some drivers for several months in a row, resulting in a panel data set that makes possible estimation that uses income and work decisions across days. Furthermore, my data set spans an important exogenous event: on May 3rd 2004: a higher fare for taxis was implemented by the New York City Taxi and Limousine Commission.

3.5(B) Are the Numbers Reasonable?

One easy check of the reliability of the data is to make use of the exogenous wage shift instituted by the New York City Taxi and Limousine Commission on May 3rd, 2004. The 2006 New York City Taxi Cab Factbook (Schaller Consulting, 2006) reports that in May 2004, the initial charge (fare drop) increased from $2.00 to $2.50, and the per mile charge increased from $1.50 to $2.00 (the waiting time charge stayed roughly constant, and the per minute charge stayed constant). In addition, in May 2004 a $1 surcharge was added for trips between 4:00pm and 8:00 pm, and during 2004 the flat fare from JFK airport to Manhattan was increased from $35 to $45. Calculating the effect of the May 2004 fare change on a 2.8 mile trip with 4.77 minutes of wait time that does not overlap rush hour (4:00 pm to 8:00 pm), the Factbook concludes that the fare should have increased from $6.85 to $8.65, or by about 26%. This gives a rough estimate of the average fare change I should find in my data.

I find in my data set that the average May fare increase for trips that end before rush hour begins is $2.03, with a 95% confidence interval of $1.91 to $2.16. Given that the average such fare was $7.11 before the fare change, this represents a precisely measured average fare increase of 28.6%. The precision of this estimate and its correspondence with the known change in fare rules are good evidence that the fares that the drivers recorded (and that the research assistants observed and entered into the computer) bear a close connection with reality: it would be very unlikely that the drivers could replicate this fare change if they were merely inventing the fares they wrote down.

3.5(C) Simple Statistics

In Table 2.1a, I report simple statistics at the trip level. It is clear that most trips are short (median time is 10 minutes), with small fares (median fare is $5.90) and that the wait time between trips is often very short (median wait time is 4 minutes). In addition, after any trip the average probability of stopping work for the day is quite low: 4.8%. In Table 2.1b and Figures 2.1 and 2.2 I report simple statistics at the shift level. Drivers typically earn about $150 a shift, after working for about eight hours on average.

3.5(D) Is there a Constant "Wage" Throughout Each Shift?

I follow Farber (2005) in constructing the hourly wage as follows: I break-up each shift into hour-long pieces. For each piece, I calculate the amount of income earned during that hour (assigning 100% of the income of any trips that begin and end in that hour to that hour's earnings, and assigning a proportion equal to the proportion of the trip time that took place during that hour for trips that overlap the boundaries of the hours). For each hour-long piece, I similarly calculate the amount of time spent driving during that hour. I then divide the former by the latter to obtain an estimate of the wage rate during that hour of work. Table 2.2 shows that in this data set there is little autocorrelation between the hourly wages earned in one hour of a shift with the hourly wages earned in an earlier hour.

3.5(E) Drivers without enough observations

Since I wish to estimate separate labor supply functions for each driver, I must restrict myself to drivers with a sufficient number of observations to make this estimation possible. Given the large number of independent variables in each equation, I thus look for drivers with 1,000 or more total trips in the sample. This restricts me to 65 drivers out of the original 114. In the next section, I introduce the results from estimating Equation 8 on all of the drivers, and from estimating a simplified version of equation 1 separately on each of these 65 drivers[8].

[8] When I do not simplify Equation 8 for the individual-level estimation by dropping controls for clock hour, then too many observations are dropped to make the results reliable. Farber (2005) faces a similar problem when estimating individual functions for five drivers. Farber concludes that dropping the clock hour dummies may be affecting the estimates of the effect of daily income and daily hours. When I estimate equation 1 on the entire sample at once and compare the results with an estimate that does not control for clock hours, I find that the effects of daily income and daily hours do not change in terms of sign, orders of magnitude, or statistical significance. The positive point estimate of daily income's effect increases by 75%, but the difference is not significant, and in both cases the effect of daily income is significantly different from 0. The positive point estimate of daily hours' effect increases by 60%, and this difference is significant, but again in both cases the effect of is significantly different from 0. Since the sign, order of magnitude, and statistical significance of the effects of daily income and daily hours do not depend on controlling for clock hour when I estimate variations of one function over all of the drivers, I here assume that these will not be affected by my (forced) choice of specification in estimates of individual drivers.

3.6. Are there any Reference-Dependent Drivers?

The results from estimating individual labor supply functions for each driver are striking. There is a great degree of homogeneity in the effect of daily hours on the probability of stopping, while the effect of daily income on the probability of stopping is heterogeneous. Below I describe first the results of specifications which involve one labor supply function estimated over all drivers in the sample, and then the results from estimating a separate labor supply function for each of the 65 drivers in the sample who have more than 1000 observations. Finally, I carry out additional empirical analysis to explore the heterogeneity that I find when comparing the individual functions.

3.6(A) One Function over all Drivers

The results of estimating one function over all the drivers in the sample are reported in Table 2.4 and in Figure 2.3(a, b, c, and d). They are very similar to the results of Farber (2005). The effect of daily hours on the stopping decision is large, positive, and significant. Although the effect of income is very small, it is clear that income matters for the stopping decision: the income indicators are jointly significant. It is also clear that – except at low incomes below $40 – the stopping probability is increasing with income. In particular, the linear estimate has positive slope, significantly different from 0; and in the flexible estimate the point estimates are increasing, with standard errors small enough that higher income indicators have an effect on stopping probability that is significantly greater than that of earlier income indicators (though not, of course, when comparing adjacent income indicators). These patterns of increasing point estimates and joint significance of the income indicators are robust to changes in the base income comparison group (i.e., the dropped income indicator), and, as Table 2.4 shows, to clustering at the driver level[9].

The main difference with the results of (Farber 2005) seems to be that here income has a small and significant effect on the stopping decision, whereas in (Farber 2005) there is no significant relationship between income and stopping probability. It is possible that sample size explains much of this difference. Indeed, empirical analysis of a random sample of this paper's data set chosen to be the same size as Farber's data set yields income indicators with large standard errors that render their effect on stopping probability insignificant.

Thus, the effect of income on stopping probability is generally positive, and there is no consistent evidence that it is decreasing anywhere. The linear estimate is positive and highly significant. However, in comparison with the effect of daily hours on the stopping decision, it is clear that the effect of daily income is very small. According to Table 2.4, the effect of a $20 increase in daily income (roughly the amount of money a driver can expect to earn from an hour of work) is an increase of approximately 0.0005 in the probability of stopping. Given that the base probability of stopping is 4.8%, this implies that an increase in daily income of $20 should be associated with an over 1% increase in the probability of stopping. The effect of an additional hour of work, on the other hand, tends to be three to seven times greater than this, including in specifications with non-linear effects. Thus, the results from estimating the model across all drivers together are largely the same as in Farber (2005): daily hours are an important determinant of the daily stopping decision, but, at least in comparison, daily income is not.

[9] If the data set is divided into two pieces – before and after the exogenous wage increase of May 4th 2004 – then the above estimation applied to the former period shows that stopping probability generally increases with income for incomes above $30, but the result is only significant without clustering. The above estimation applied to the latter period shows that stopping probability increases with income for incomes above $50 and below $250 (with some ambiguity below and above this range), and this result is significant with clustering at the driver level. Thus, the preponderance of the evidence suggests that stopping probability is increasing with income, even accounting for possible changes in behavior due to the wage increase.

3.6(B) Individual Labor Supply Functions for Each Driver

When I estimate separate labor supply functions for each of the 65 drivers who report more than 1000 trips, the results show homogeneity in the effect of daily hours on stopping probability. Table 2.5 shows that the effect of daily hours on stopping probability is positive and significant for 56 drivers, insignificant for 8 drivers, and negative and significant for one driver. But the results show a great deal of heterogeneity in the effect of daily income on stopping probability, however[10]. Table 2.5 shows that the effect of daily income on stopping probability is positive and significant for 29 drivers, insignificant for 35 drivers, and negative and significant for one driver. Figure 2.4 shows the densities of these income coefficients for all drivers and for the 29 drivers with positive and significant income coefficients

The average size of the effect of daily income for the 29 drivers with positive and significant effects of daily income is 8.7 times as large as the effect of daily income estimated from a single equation over all drivers. This makes the effect of an extra $20 in daily income for these drivers slightly larger than the effect of an additional daily hour as estimated from a single equation over all drivers. Considering the daily hours effects for each individual in this subset, it is clear that for these drivers daily hours are often less important then daily income: of these 29 drivers, 20 of them have either no significant daily hours effect, or a significant daily hours effect whose point estimate is smaller then the effect of an additional $20 in daily income.

---

[10] This degree of heterogeneity disappears if one estimates the same specifications as those in Table 2.5, but leaves out the control for the number of trips taken so far. Without controlling for the number of trips taken so far, daily income has a large, positive and significant effect (of a similar magnitude to that of daily hours) for about 55 drivers out of 65. This is not surprising, however, because drivers who take more trips per hour are likely to expend more effort per hour and thus to have a greater need to go home, conditional on hours worked. Thus, in the absence of a control for the number of trips taken so far, this effort cost per trip will all fall upon income (since drivers who take more trips per hour will likely have higher income per hour). Therefore, I interpret these results without trip controls as being less informative than those from the specification that simultaneously controls for daily hours, daily income, and the number of trips made so far.

Thus, it seems that the vast majority of drivers base their stopping decision strongly on hours. Furthermore, about half of the drivers also strongly consider daily income in their stopping decision, with the other half not making income a significant part of their decision.

3.6(C) What Else is Different about the Drivers Who Care about Income?

In Section III, I suggested that one explanation for a positive effect of daily income on stopping probability was that these drivers have daily income reference points. If the drivers in my sample who have this positive income effect really are more likely to work under the shadow of such kinks in utility than the other drivers are, and if the position of these kinks varies less across days than the typical variation in daily income among people without kinks, then these drivers should have less variation in their daily income at the end of each day than the other drivers do. Thus, I consider how much their daily income varies at the end of the day compared with other drivers. The scatterplot in Figure 2.6 shows how each driver's income coefficient (reported in Table 2.5, and Figure 2.4) compares to the standard deviation in his daily income across days (reported in Table 2.5 and Figure 2.5). It is clear that drivers with high standard deviations in daily income have income coefficients that are close to 0. This correlation is in fact -0.3, and is significant at the 5% level. This suggests that the more drivers care about daily income, the less their income varies across days.

Finally, I develop an intuitive sense for the magnitude of these positive and significant income coefficients by considering the following thought experiment: suppose that the drivers noticed upon entering their cab at the start of their shift that an anonymous donor had gifted them with a $20 bill, left on the driver's seat. On average, how would this windfall affect their behavior?

For this calculation, I initially assume that, ceteris paribus, drivers are equally likely to stop after any trip (this is clearly not the case, since drivers are much more likely to stop after the second or third trip in comparison with much later ones). Looking at the distribution of trips per shift in Table 2.1b, I see that on average drivers make about 20 trips per shift – this number is 20.8 when calculated across shifts. The constant stopping probability per trip that would lead to an expected number of trips of 20.8 is: 1 / (1 + 20.8) = 0.046. The distribution of significant income effects in Table 2.5 and Figure 2.4 has a mean significant income effect of 0.006 on stopping probability for a $20 increase in daily income. Thus, an additional $20 of daily income should increase the stopping probability per trip from 0.046 to 0.046 + 0.006 = 0.052. This would result in a total of 18.18 expected trips; a decrease in expected trips of 2.54 trips. Since Table 2.1a shows that the expected dollar value of each trip is about $7.90, this suggests that the decrease in earnings from this increase in stopping probability will be 2.54 trips times $7.90 per trip = $20.46. Thus, it appears from this calculation that that the drivers in the sub-sample who care about daily income in their stopping decision are daily target earners (they practice extreme reference dependence): starting the day with $20 extra causes them to end the day early, with nothing extra at all.

More precise calculations, in which the probability of stopping after each trip varies as it does in the data, show that the hypothetical average reference-dependent driver would respond to the $20 windfall by ending his day early, with only $6 more than he otherwise would[11]. While not behaving as if he had an income target, his reference dependence is still significant.

[11] I construct an artificial data set in which the probability of stopping after the j^{th} trip is set equal to the average probability in the real data set that the j^{th} trip of any shift is the final trip of that shift. I then add to these probabilities the average effect of a $20 increase in daily income (0.006), starting from the first trip (i.e., assuming that the $20 was dropped into the cab before the first trip). The updated probabilities in this data set imply that, on average, about 1.8 fewer trips will occur per shift. This results in $14 fewer of earned income per shift.

3.7. What Determines the Reference Daily Income Level?

Having established that significant reference dependence may hold for a subset of the drivers, I ask: can I determine empirically whether the evidence supports Rabin & Koszegi's theory of the cause of this reference dependence? Rabin & Koszegi (2006) surmise that a driver's income reference level is determined by his income expectations. An alternative theory is that income reference levels are determined independently of income expectations. Under this alternative theory, reference income levels stay constant during a permanent increase in the hourly wage, and thus reference dependent drivers work fewer hours, earning the same amount per day. Thus, I can rule out this alternative hypothesis by showing that reference-dependent drivers do not work fewer hours after an exogenous wage increase, and do earn more per day.

In Table 2.6, I report the results of such a test. Specifications 1 and 2 show that the reference-dependent drivers do not work fewer hours after the May 3rd wage increase. Thus, their reference points must have increased when the wage increased. In Specification 3 of the same table, I see that their daily incomes in fact increased by about 27%, roughly the same size as the 26% fare increase. It therefore appears that these drivers' income reference levels increased by the same amount as the fare increase[12].

3.8. Conclusion

In this paper I bring empirical evidence to bear for the first time on the question of what causes daily income reference points to be at one position rather than another. Making use of a new panel of New York City Taxi Drivers whose days of work span a permanent exogenous increase in their fares, I find that those drivers who appeared to have daily income reference

[12] Curiously, this implies that the reference income levels seemed to have increased more than the hourly wage increase: because there were fewer fares earned per hour after the fare increase, the hourly wage increase must have been smaller than 26%. Thus, the reference-dependent drivers in fact increased their hours after the wage increase, keeping their income increase equivalent to the fare increase, rather than the hourly wage increase.

points do not decrease their hours of work after the fare increase, but do increase their daily income. This suggests that their daily income reference levels increased when the fare increased. It is intuitively appealing that income reference levels would be tied to the income a driver might expect to earn on a given day, and this intuition was formalized by Rabin & Koszegi (2006).

Future work should look for the existence of reference income levels over longer time periods than one day, as well as the existence of reference leisure levels. Finally, more precise structural estimation of the reference points, as in Farber (2008), should be coupled with the exogenous wage changes considered in this paper to more fully understand the source and significance of the reference points that affect labor supply.

Chapter 3. Is There Less to Beauty Than Meets the Eye? Premiums to Beauty, Grooming, and Personality in American Schools

3.1. Introduction

A wide literature has reported that physical features such as height, weight and beauty are correlated with achievement, both inside and outside the workforce. The sub-literature on beauty has repeatedly found positive correlations between beauty and such variables as earnings (Hamermesh & Biddle 1994) (Biddle & Hamermesh 1998) (French 2002), husband's education (Hamermesh & Biddle 1994), experimental wage outcomes (Mobius and Rosenblat 2006), and college exam grades (Cipriani & Zago 2005). The recent article in the Economist (December 22nd 2007) summarizes: "The ugly are one of the few groups against whom it is still legal to discriminate. Unfortunately for them, there are good reasons why beauty and success go hand in hand."

In fact, as (Mankiw 2006) explains, there are at least three explanations for these ubiquitous beauty premiums:

> "One interpretation is that good looks are themselves a type of innate ability. . . A second interpretation is that reported beauty is an indirect measure of other types of ability. How attractive a person appears depends on more than just heredity. It also depends on dress, hairstyle, personal demeanor, and other attributes that a person can control. . . A third interpretation is that the beauty premium is a type of discrimination. . ." (Mankiw, 2006).

Of these explanations, existing research has emphasized discrimination (Hamermesh & Biddle 1994), a combination of beauty-stereotypes and beauty-related increases in confidence and social skills (Mobius & Rosenblat 2006), or a correlation between beauty and productivity (Cipriani & Zago 2005) (Pfann et al. 2000)[1]. However, to my knowledge my study is the first to

[1] Mobius and Rosenblat find that 60% of the beauty premium can be explained by the higher confidence and social skills of the beautiful. The remaining 40% of their beauty premium in wages can be explained by a "direct stereotype" effect in which – for a given level of employee self-confidence – employers

thoroughly explore the second interpretation (e.g., that dress and personal demeanor may be behind the premium).

Indeed, a recent CNN.com/careerbuilder.com article points out that "many folks who are lovely to look at complain that they lose out on jobs because people assume they are vacuous or lightweights" (Lorenz, 2005). The article goes on to conclude: "Hiring managers say it is the appearance of confidence they find attractive, not the presence of physical beauty." In light of this popular intuition, I propose that the beauty premium may be entirely explained by positive correlations between physical beauty and other (potentially more important) dimensions of attractiveness, such as attractiveness of personality and of grooming. In fact, the psychology literature suggests that beauty is strongly related to perceived attractiveness of personality (Feingold 1992, Eagly, Ashmore, Makhijani, and Longo 2001). If the perceived attractiveness of personality has been unmeasured (or not sufficiently measured) in previous studies, then it is possible that, with good measures of the attractiveness of personality, the beauty premium found in previous studies can be explained by a previously unnoticed premium on attractive personalities. The same argument holds for any other dimension of attractiveness that may be correlated with beauty and not sufficiently measured in previous studies, such as grooming.

For the first time, I control for direct subjective measures of the *attractiveness* of personality and the *attractiveness* of grooming, measured in the same manner that physical attractiveness has been in previous studies. It is important to control for the overall attractiveness of a beautiful person's personality and of their grooming if one believes that it is this attractiveness that has the best chance of affecting professional and academic evaluations, rather than, for example, any individual personality trait. The existing literature has controlled

---

wrongly assume that the physically beautiful are more productive. I propose that the reason that 100% of the beauty premium was not explained by the personalities and grooming of the beautiful is because the attractiveness of these characteristics was never fully measured.

somewhat for the mechanisms of personality and grooming by including various measures of personality traits, and by implementing designs that limit observation of social interaction and grooming, but it has neither measured them nor controlled for them fully[2].

Using data from the first three waves of the National Longitudinal Study of Adolescent Health, I explore for the first time the effects of beauty on academic achievement in middle and high school[3]. My strategy is first to replicate in this new area the beauty premium found in previous work – using the same techniques that others have used – and then to check the robustness of this premium to controls for other ascriptive characteristics such as personality and grooming.

First, I find that secondary school students experience a large beauty premium in the area of academic achievement. Second, this beauty premium is fully accounted for by a premium to attractive personalities and attractive grooming. Third, when controlling for these other dimensions of attractiveness, I find evidence of a beauty deficit: ceteris paribus, more beautiful students' transcripts show slower advancement and lower grades (especially among women). Finally, I find evidence that more beautiful students engage in a variety of relationship activities more frequently.

One possible mechanism for the beauty deficit may be that, controlling for the attractiveness of their other characteristics, the physically beautiful face discrimination. The possibility of discrimination against the beautiful has been discussed by the popular press (CNN

---

[2] In the current literature, the measures of personality traits include one of self-esteem (measured by the degree of agreement with some philosophical statements about life) in Hamermesh and Biddle 1994, and a person's confidence in their own maze-solving ability in Mobius and Rosenblat 2006. Mobius and Rosenblat 2006 limit social interaction in one of their experiments by letting employers only view a picture of their employee. Biddle and Hamermesh 1998 use photographs that concentrate on the face, leaving limited room for observation of grooming.

[3] I add to Cipriani and Zago's work, by providing the first empirical analysis of the effect of beauty on academic outcomes while controlling for any variables other than gender, prior academic success and academic status.

article). This explanation would also be consistent with the experimental work of Andreoni and Petrie 2005, who find that the beauty premium in a repeated public goods game becomes a beauty deficit when individual contributions are revealed: once it becomes clear that beautiful players contribute no more than others, the other players change from over-contributing to beautiful players to under-contributing to them. Likewise, Wilson & Eckel 2006 find that experimental subjects punish beautiful players in the second stage of a game involving "trust and reciprocity" when the beautiful do not "live up to expectations."

Andreoni and Petrie's inference is similar to the results reported in the psychological literature I summarized above: they find that "Players seem to expect beautiful people to be more cooperative." They go on to conclude that "relative to these expectations, . . . [beautiful people] appear more selfish, which in turn results in less cooperation by others" (Andreoni and Petrie). It is thus possible that the beauty deficit I find is driven by beautiful young women who do not have the good personalities people expect them to have, and are punished for their combination of high beauty and average social skills by receiving low marks in the classroom.

Another explanation for the observed beauty deficit is based on the data itself: I find that beautiful young women put forth less (self-reported) effort on their school work, ceteris paribus.[4] To account for this lower effort, I propose that physical beauty may increase the marginal utility from socializing with peers. It might then be optimal for beautiful students to spend more time and effort socializing and less time and effort on school work, helping cause adverse academic outcomes. My empirical results support the hypothesis that beautiful students spend significantly more time in relationships and socializing.

Finally, I examine outcomes that may be affected by low effort in school work and high investments in romantic and social activities. I find – apparently because they are engaging in

[4] In other words, the marginal effect of beauty on self-reported academic effort is negative when the other variables are controlled for, including personality and grooming.

social drinking and entering romantic relationships more frequently – that beautiful students are more likely to engage in sexual activity and to participate in risky activities such as becoming drunk and smoking.

In the following sections, I discuss the theoretical and empirical background of the relationship between beauty and achievement; introduce my empirical strategy; summarize the many outcomes related to beauty, personality and grooming; propose a conceptual explanation of the results; and discuss the implications for future research.

3.2. Existing Evidence for a Beauty Premium

The literature on the correlates of beauty has three threads: (1) a psychological literature, (2) an empirical microeconomics literature, and (3) an experimental economics literature. The first literature has, among other results, shown that there is definitely a relationship between our physical beauty and the way others perceive our personalities, social skills, and intelligence. The second, microeconomics, literature has found positive links between physical beauty and achievement; it has emphasized discrimination, beauty stereotypes, and productivity as important channels. The third, experimental, literature has also found positive links between physical beauty and outcomes, but has suggested that *some* of this premium may be explained by correlations with the perceptions of personality; it has found evidence of beauty deficits when personality expectations are not met.

The literature in experimental and social psychology has long considered beauty's effect on other people's perceptions:

> The main experimental findings are that more beautiful people are viewed superior along several dimensions: personality traits (sociability, dominance, sexual warmth, modesty, character), mental health, intelligence and academic ability, and social skills (Mobius and Rosenblat 2006).

While many reasons could be proposed for the cause of these correlations, in this paper it is enough to note the following: any correlation between beauty and achievement could be easily caused by the correlations between beauty and other attractive characteristics just cited above.

The main strength of the empirical microeconomics literature is the wide variety of real-world areas in which the beauty premium has been found. In addition to the variables mentioned in the introduction (earnings in Hamermesh & Biddle 1994, Biddle & Hamermesh 1998, and French 2002; husband's education in Hamermesh & Biddle 1994, and college exam grades in Cipriani & Zago 2005), this literature has found positive correlations between beauty and election outcomes (in both Hamermesh 2006 and Berggren, Jordahl & Poutvaara 2007), and student perceptions of instructor's teaching productivity (in Hamermesh & Parker 2003). Beauty premiums are clearly not an artifact of the laboratory setting – they matter in the real world. In spite of this strength, the empirical microeconomics literature on beauty has not fully taken into account some of the lessons learned from the experimental and psychological literature: for example, that personality is an important correlate of physical beauty, and that the relationship between achievement, personality, and beauty may be subtle.

For instance, of the papers above, only Hamermesh & Biddle 1994, and Berggren, Jordahl & Poutvaara 2007 attempt to control for aspects of personality. In Hamermesh & Biddle 1994, this control consists of a measure of self-esteem computed through averaging the degree of agreement of responders to six questions, one of which is: "Those who are always trying to get ahead in life will never be happy." It is clearly difficult to use answers to philosophical questions such as this to measure the attractiveness of a person's personality: presumably, those who agree with the philosophical view proposed in the question will find each other more attractive, and those who disagree with it will feel likewise. In Berggren, Jordahl & Poutvaara 2007, the authors asked the internet users who rated the beauty of the photographs in their sample to also guess the competence, likeability, trustworthiness, and intelligence of the people in the photographs. In this case, the fact that any guesses about personality are mediated through the photograph makes it difficult for the answers to be as reliable as a rating based on a face-to-face meeting.

The experimental literature makes a greater effort to consider personality as a mechanism for the beauty premium. Mobius and Rosenblat report the results of a set of experiments in which "employers" had to assign wages to "employees" being hired to solve mazes. Whenever the employers saw photographs of the employees, spoke to them over the phone, or met them in person, the beautiful employees received higher wages, even though the beautiful were actually no better at solving mazes, and even though the employers' payoffs depended on the closeness of their wage assignments to employee productivity. Mobius and Rosenblat run a pooled regression, controlling for the worker's own self-confidence in maze-solving. Based on the sizeable beauty premium that remains, they conclude that 40% of the beauty premium is caused by a beauty stereotype effect: "for a given level of confidence, physically attractive workers are (wrongly) considered more able by employers" (Mobius & Rosenblatt 2006).

The implicit assumption of Hamermesh & Biddle 1994, & Mobius & Rosenblatt 2006, is that various aspects of personality should be directly controlled for if possible. This is why the first paper includes regression controls for an attempted measurement of self-esteem, and why the second paper includes regression controls for employee's own confidence in their maze solving ability. Where I disagree with their interpretation is that they assume that the beauty premium that remains after their limited personality controls is not due to some unmeasured aspects of personality or other dimensions of attractiveness.

My question is: what would have happened to the beauty premiums in Hamermesh & Biddle 1994, and in Mobius & Rosenblatt 2006, if – rather than including the limited personality controls that they did – they had instead controlled for direct subjective measurements of attractiveness of personality and attractiveness of grooming? If the effect on the beauty premiums in my study from adding these controls is any guide, then their marginal effects of beauty might have disappeared completely after adding these controls.

Therefore, one take-away point is that in none of these studies (empirical or experimental) have the authors let physical beauty compete with other dimensions of attractiveness to explain achievement. The other dimensions were almost never directly measured; at best they were either indirectly measured (in the empirical literature) or left-unmeasured but varied slightly through experimental design (in the experimental literature) – usually they were ignored[5]. In light of the psychological literature showing how perceptions of a wide variety of attractive characteristics are related to beauty, it is essential to control for direct measurements of these other dimensions of attractiveness. In Section III, I explain the data set that I use to carry out this analysis.

3.3. Data Set

I make use of the restricted use data set from the National Longitudinal Study of Adolescent Health (Ad Health) (Harris et al., 2003b). This study consisted of several stages. In the first stage, the Ad Health research team took a stratified random sample of 80 high schools from a sample frame consisting of all 26,666 United States high schools that included an 11th grade and had an enrollment of at least 30 students. For each sampled high school, the researchers randomly selected one feeder middle school or junior high school (i.e., one school

[5] An exception is the measurement of beautiful employees' confidence in their own maze-solving ability in Mobius and Rosenblat (2006).

that fed students to that high school and had a 7th grade). Since some schools had no eligible feeders, or were their own feeder, this resulted in a total of 145 participating schools. From these schools, 90,118 students (who all gave active or passive consent) completed a 45 minute questionnaire in school during the period between September 1994 and April 1995.

In the second stage of data collection, during April 1995 through December 1995, researchers administered an in-home survey with more detailed questions to both children and parents. First, they randomly sampled a core group of 12,105 students from the larger sample above, obtaining a group that was representative of United States adolescents in grades seven through twelve during the 1994 to 1995 school year. Second, they added to this core sample several oversamples of special groups such as minorities, children with disabilities, and children from schools where everyone was participating in the survey. Thus, with the correct sample weights, a full sample of 20,745 students can be used for inference about United States adolescents in grades seven through twelve during the 1994 to 1995 school year.

Finally, most of this sample of 20,745 students were reinterviewed both one year and six years later. During the second wave, from April 1996 through August 1996, 14,738 students were interviewed. During the third wave, from July 2001 through April 2002, 15,197 students were interviewed. Some of the reported reasons for the decline in interviews from wave I to wave II are 1) that most students in twelfth grade in wave I were no longer in school during wave II and 2) that the wave I disabled oversample was not reinterviewed. The increase in interviews from wave II to wave III is (likely) due to the fact that "a sample of 1,507 partners of original respondents were also interviewed" (Harris et al., 2003b).

Thus, I am able to infer results about American adolescents who were in seventh to twelfth grades during the 1994 to 1995 school years, and in particular I can see how these students progressed over time. The in-home questionnaires ask numerous questions about the educational, familial, social, health and emotional attributes of the students. I report some simple statistics about important variables in Table 3.1.

The Ad Health data set contains information that reflects the choices students make regarding school work, romance and partying. To find the effect of physical appearance on academic effort and outcomes, I consider the answers the students supplied to the following questions:

- "In general, how hard do you try to do your school work well?"
- "Since school started this year, how often have you had trouble completing homework?"
- "Since school started this year, how often have you had trouble paying attention in school?"
- "Since school started this year, how often have you had trouble getting along with your teachers?"
- The students were asked to report their own grades in Mathematics, Science, Social Studies/History, and English during the first two waves. However, during the third wave the Add Health administrators obtained school transcripts. Thus, I primarily use the school transcripts to find cumulative GPAs and progression in course work, checking the robustness of my results by comparing with the self-reported academic outcomes.

To find the effect of physical attractiveness on romantic activity and sexual behavior, I consider the answers the students supplied to the following questions:

- "Have you had a relationship with anyone in the last 18 months?"
- "How many relationships have you had in the last 18 months?

- "Not counting the people you [may] have described as romantic relationships, [since Wave I] have you [ever] had a sexual relationship with anyone?"

To learn the effect of physical attractiveness on partying, I look at questions about drinking, drunkenness, and smoking:

- "[Since Wave I,] have you had a drink of beer, wine, or liquor—not just a sip or a taste of someone else's drink—more than two or three times [in your life]?"
- "Think of all the times you have had a drink during the past 12 months. How many drinks did you usually have each time? A "drink" is a glass of wine, a can of beer, a wine cooler, a shot glass of liquor, or a mixed drink."
- "Over the past 12 months, on how many days have you gotten drunk or "very, very high" on alcohol?"
- "[Since Wave I,] have you [ever] tried cigarette smoking, even just one or two puffs?"

3.4. Replicating the Empirical Beauty Premium

In this section, I demonstrate that I can replicate the common empirical finding of a beauty premium by finding one in the academic achievement of American high school students. I apply the same empirical strategy as Hamermesh & Biddle 1994, a strategy which was subsequently followed by much of the other work in this area. By demonstrating that in the absence of good controls for the attractiveness of personality and of grooming there exists a beauty premium in academic achievement, I set the stage for my work in Section V, which will show that in the presence of such controls, the beauty premium disappears.

3.4(A) Empirical Strategy

The empirical strategy is to estimate the following equations in a cross-sectional regression on a representative sample of U.S. middle and high school students:

$$Y_{i,t} = \alpha \cdot P_{i,t-1} + \delta \cdot Z_{i,t-1} \quad (1)$$

$$Y_{i,t} = \alpha \cdot P_{i,t-1} + \beta \cdot X_{i,t-1} + \delta \cdot Z_{i,t-1} \quad (2)$$

where $i = 1,..,n$ designates a student, t = Wave II, and $t-1$ = Wave I. The outcome variable Y will be various measures of academic effort, academic achievement, and time and effort spent on relationship and social activities common to teenagers, such as dating and drinking. The variable P denotes a vector of dummy variables for different subjective scores of physical attractiveness (with "average" attractiveness as the omitted category), and X denotes a matrix of vectors of dummy variables for different subjective scores of the other dimensions of attractiveness and ascriptive characteristics: personality, grooming, candor, and degree of physical maturity. Thus, I allow for flexible effects of attractiveness on outcomes, over all measured dimensions of attractiveness. Finally, Z is a set of important controls that are likely to be correlated with physical attractiveness and to affect outcomes: gender, race, age, height, weight, subjective health, school year, family income and parental schooling[6]. All results are replicated with current attributes on the right hand side.

3.4(B) Empirical Results

There is some variation in the size of the effects across specifications, but a consistent picture emerges. In some measures of academic achievement, (particularly in advancement through math courses, advancement through science courses, and cumulative high school GPA), beautiful students show considerably higher achievement then the non-beautiful. As in much of the existing literature (Hamermesh & Biddle 1994), the beauty premium is larger and more significant for men than for women.

[6] I estimate all specifications reported in the body of the paper with both height and weight and BMI. The estimated coefficients are almost identical.

Table 3.2 contains estimates of Equation (1) where the dependent variable is cumulative high school GPA. Specification 1 is for the entire sample, Specification 3 is for males, and Specifications 5 is for females. Specification 1 shows that, compared with a hypothetical student with the same non-beauty attributes but a rating of 3 (average) in beauty, a student with a rating of 5 (very attractive) in beauty would have a cumulative GPA over all courses of 0.10 points higher, an increase of about one eighth of a standard deviation. Specification 3 shows that the beauty premium is larger for young men, while Specification 5 shows that the effect of beauty is insignificant for young women.

Table 3.3 contains estimates of Equation (1), where the dependent variable is the progression in mathematics courses. As before, Specification 1 is for the entire sample, Specification 3 is for males, and Specifications 5 is for females. Specification 1 shows that, compared with a hypothetical student with the same non-beauty attributes but a rating of 3 (average) in beauty, a student with a rating of 4 (attractive) in beauty would be significantly ahead in mathematics course progression. In Table 3.4, I repeat the above specifications with science progression as the dependent variable. Specification 1 of this table shows that the beautiful student above would be significantly ahead in science. In results that I have omitted for the sake of brevity, I find that the faster progression through science courses is accompanied by a significant increase in the total number of science credits earned after 4 years: a student with a rating of attractive would have 12% of a standard deviation more science courses completed than a student with average looks.

In the next section, I show that these beauty premiums become either insignificant effects or significant deficits when other ascriptive characteristics such as personality and grooming are controlled for.

3.5. Premiums on Personality and Grooming Explain the Beauty Premium

In general, where a beauty premium is found in the results above, controlling for other ascriptive characteristics such as personality and grooming fully accounts for the beauty premium, leaving an effect of beauty on academic achievement that is either insignificant or (usually) significant and negative. This beauty deficit is also found in many specifications where no beauty premium exists. Thus, the consistent finding is that any positive effect of beauty on academic achievement – if it exists – is mediated through other ascriptive characteristics, in particular through attractiveness of personality and attractiveness of grooming. Furthermore, the results consistently show that any significant effect of beauty on academic achievement that is independent of these other pathways is negative.

Table 3.2 contains estimates of Equation (2), with the dependent variable cumulative high school GPA. Specification 2 is for the entire sample, Specification 4 is for males, and Specifications 6 is for females. Specification 2 shows that, compared with a hypothetical student with the same non-beauty attributes but a rating of 3 (average) in beauty, a student with a rating of 5 (very attractive) in beauty would earn a 0.15 lower cumulative grade point average in secondary school. This is a decrease of almost one fifth of a standard deviation. This effect of beauty on official school transcript grades is fairly robust to using the (presumably less reliable) self-reported grades from the Wave I and Wave II surveys, especially for self-reported English and Science grades. Specifications 4 and 6 show that this effect is insignificant for men (Specification 4) and is larger and more significant for women (Specification 6).

Table 3.3 contains estimates of Equation 2, with the dependent variable progression in mathematics courses. As above, Specification 2 is for the entire sample, Specification 4 is for males, and Specifications 6 is for females. Specification 2 shows that, compared with a hypothetical student with the same non-beauty attributes but a rating of 3 (average) in beauty, a student with a rating of 5 (very attractive) in beauty would be significantly behind in mathematics. Specifications 4 and 6 show that these effects are not significant for men and that they are similar in magnitude for women.

In Table 3.4, I repeat the above, with the dependent variable as progression in science courses. Specification 2 shows that he would also be behind in science. Specifications 4 and 6 again show that the beauty effects are insiginficant for men, and more important for women. How large are these effects? I find from OLS regressions of the number of math credits completed after 4 years and the number of science credits completed after four years that a student with a rating of 5 in beauty would, ceteris paribus, be 15% of a standard deviation behind in mathematics credits, and 15% of a standard deviation behind in science credits after four years, as compared with a student with average looks.

In other results that I have omitted for the sake of brevity, I find that, compared with a hypothetical student with the same non-beauty attributes but a rating of 3 (average) in beauty, a student with a rating of 5 (very attractive) in beauty would have as much as twelve percent of a standard deviation greater difficulty completing his homework. He would also experience over one fifth of a standard deviation higher frequency of experiencing trouble paying attention in class. These effects are similar or possibly smaller for men, and are likewise similar or possibly larger for women.

The above effects are quite robust to variation in the specifications, including the regression of Wave II outcomes on Wave II attributes, and replacing the outcome variables with related variables: in particular, beautiful students seem to also be more likely to fail math courses, and to report having trouble with their teachers. In addition, as is the case in Hamermesh et al., the effects for regressing wave I outcomes on wave I attributes are similar to those for regressing wave II outcomes on wave I attributes (and indeed for regressing wave II outcomes on wave II attributes).

3.6. Why do Personality and Grooming Make a Difference?

Why might there be a negative effect of beauty on achievement when other ascriptive characteristics are controlled for? One reason may be that when other ascriptive characteristics are controlled for, beauty is associated with putting forth less effort on school work, while when these other characteristics are not controlled for, beauty is not. This is what I find when I look at self-reported effort among the students.

In Table 3.5, I find that, controlling for other ascriptive characteristics, the beautiful report putting forth less effort on their school work. When other ascriptive characteristics are not controlled for, beauty has no significant effect on effort. Breaking up the pooled sample, I show that this effect is strong for women, and is less significant for men. The fact that the lower effort of beautiful students holds mainly in the pooled sample and in the female sample corresponds well with the finding that the beauty deficit in achievement is found mainly in the pooled sample and the female sample[7].

If it is the case that beautiful students put forth less effort on their school work (when their personalities and other characteristics are controlled for), what are they more likely to put time and effort into? I find that beautiful students are spending extra time dating, drinking, and socializing[8].

7 In order to determine whether the lower effort that beautiful students report on school work may be the mechanism behind their low achievement in school, I replicate the regressions in Table 2 (i.e., the regressions of cumulative high school GPA) controlling for self-reported effort. These results suggest that the lower effort of beautiful women explains about one fourth of the beauty deficit in cumulative GPA. These results are hard to interpret, however, because effort is certainly jointly determined with outcomes: people may choose to work extra hard precisely because they are doing poorly in school.

8 I find this regardless of whether I estimate Equation 1 or Equation 2

In Table 3.6, I estimate Equations 1 and 2 with the dependent variable an indicator for having had a romantic relationship in the specified time period. Compared with a hypothetical student with the same non-beauty attributes but a rating of 3 (average) in beauty, a student with a rating of 5 (very attractive) in beauty would enter romantic relationships with a twenty percent higher frequency (see Table 3.6, specification 4). In Table 3.7, I repeat the above for the number of romantic relationships, and in Table 3.8, for the number of months of time spent in total in relationships. I find that this student would spend 1.1 months of additional time in relationships in total – a twenty-two percent increase (see Table 3.8, specification 4).

In Table 3.9, I estimate Equations 1 and 2 with the dependent variable an indicator for having had sexual activity outside of a relationship during the specified time period. I find that compared with a hypothetical student with the same non-beauty attributes but a rating of 3 (average) in beauty, a student with a rating of 5 (very attractive) in beauty would be seventeen percent more likely to engage in non-relationship sexual activity (i.e. on top of any sexual activity already practiced in relationships) (see Table 3.9, Specification 4).

In Table 3.10, I estimate Equations 1 and 2 with the dependent variable an indicator for drinking behavior during the specified time period. Compared with a hypothetical student with the same non-beauty attributes but a rating of 3 (average) in beauty, a student with a rating of 5 (very attractive) in beauty would be up to seventeen percent more likely to drink alcohol (see Table 3.10, Specification 4). Table 3.11 shows that, conditional on drinking, he would also drink an additional 1.2 drinks per drinking bout (see Table 3.11, specification 4), and Table 3.12 shows that he would be up to one fourth of a standard deviation more likely to get drunk (see Table 3.12, specification 4). Since, according to the surveys, most drinking does not occur alone, and since additional estimates show that beautiful students are no more likely to drink alone than other students, this increase in drinking is thus indicative of an increase in social drinking. I thus interpret this as likely evidence for higher rates of partying among beautiful students, ceteris paribus.

These results are robust to variation in the specifications, including the presence or absence of controls for other ascriptive characteristics, linear vs. flexible controls, and the regression of wave II outcomes on wave I characteristics. Furthermore, related outcome variables tell a similar story: beautiful students seem to also be more likely to drink five or more drinks in a row, to smoke cigarettes, and possibly to report "just hanging out" with their friends.

In the next section, I consider conceptually whether the insights of the psychological and experimental literature (that physical beauty may affect our reception of someone's personality, and that personality can affect our assessment of a person's ability) can lead to the empirical results discussed above: (1) the marginal effect of beauty on the amount of time spent in romance is positive, regardless of controls for ascriptive characteristics, (2) the marginal effect of beauty on academic achievement is positive when other ascriptive characteristics are not controlled for, and (3) the marginal effect of beauty on academic effort and academic achievement is negative when other ascriptive characteristics are controlled for.

3.7. Conceptual Explanation

The psychological literature cited in Section I makes it clear that there are correlations between physical beauty and how others perceive our character, intelligence, and social skills. It is possible that these correlations are due to a causal relationship: i.e., beauty may cause people to view us as having a more attractive personality. Furthermore, it is possible that when other people assess our achievement, attractive personality traits (e.g., good social skills, confidence, and articulateness) can convince them to assess or predict our achievement to be higher. This would be consistent with popular wisdom cited in the press (CNN article), as well as with the experiments of Mobius and Rosenblat 2006. These two assumptions – that beauty causes others to perceive us as having more attractive personalities, and that more attractive personalities encourage others to rate our achievement higher – make one possible pathway from beauty to achievement[9]. Another pathway from beauty to achievement arises if one assumes that those who rate our achievement have a taste for physical beauty itself.

Can these pathways provide a simple explanation for the variation in the sign in the marginal effect of beauty described at the end of the previous section? One such explanation is that the first pathway above is more important than the second one is. If the effect of beauty on personality and grooming is what drives the beauty premium, then one would expect that the marginal effect of beauty would only be positive when personality and grooming are not controlled for, which is what I find. But can additional assumptions explain the remaining effects: that the marginal effect of beauty on achievement (and effort) in fact becomes negative when personality and grooming are controlled for, and that the beautiful spend more time in romantic activities?

In fact, simple consideration of the time use of high school students can explain these remaining results. Consider the case of a high school student who must choose between spending time studying, spending time engaging in romantic and party activities, and spending time in general relaxation. If physical beauty directly makes romance more successful, and if it indirectly makes academics more successful (as would be implied by the assumptions about beauty and personality perceptions above), then it is unclear where the beautiful students will spend their time and what their academic achievements will be relative to plain students. In fact,

[9] Likewise, there are numerous possible causal and measurement-related connections between attractiveness of grooming and physical attractiveness, and it is possible that better groomed people are considered to be more able by employers and teachers. Thus, similar arguments to those that support a pathway from beauty to personality to achievement support one from beauty to grooming to achievement as well.

depending on the specific model one writes down, it would be possible for the beautiful to have high achievement in both areas. However, if one holds personality and grooming constant, and considers the marginal effect of beauty independent of any affects that are mediated through beauty's effect on personality, then the likelihood increases that the marginal effects of beauty will be to spend less effort studying, to spend more effort on romantic activities, and to thus experience lower academic achievement, ceteris paribus. In other words, the marginal effect of beauty may be to substitute away from studying and towards socializing.

Thus, it is possible that that in a world in which (a) physical beauty causes personality and grooming to be measured as more attractive, (b) attractive personalities and grooming inspire better achievement or evaluations of achievement, and (c) physical beauty raises the marginal utility of romance and socializing, the following predictions would emerge:
(1) the marginal effect of beauty on romantic effort and romantic achievement would be positive regardless of whether or not I control for personality and other ascriptive characteristics;
(2) the marginal effect of beauty on academic effort and academic achievement might be positive or might be negative if I do not control for personality and other ascriptive characteristics; and
(3) the marginal effect of beauty on academic effort and academic achievement would be negative if I do control for personality and other ascriptive characteristics.
These predictions match well the empirical results discussed in Sections V(B) and VI, especially the results from the pooled sample and from the all women sample[10].

[10] The results in Table 5, Specification 1 show that academic effort has no significant relationship with beauty when other ascriptive characteristics are not controlled for, while the results in Table 5, Specification 2 show that academic achievement does tend to have a positive relationship with beauty when other ascriptive characteristics are not controlled for: this discrepancy is consistent with the ambiguity in the second prediction listed above.

3.8. Correspondence with Experimental Results

Another potential cause of the negative effect of beauty on achievement when other ascriptive characteristics are controlled for can be found in recent experimental results. Andreoni and Petrie find a beauty premium in a public goods experiment, but find that when people's true contributions are revealed (which shows that the beautiful contribute no more or less than average), this beauty premium becomes a beauty penalty (Andreoni and Petrie 2005). Likewise, Wilson and Eckel find that a beauty premium becomes a beauty penalty in the second stage of a game based on "trust and reciprocity" when the beautiful are revealed to not live up to the expectations of their trust that their beauty seems to have elicited.

Thus, a growing experimental literature has suggested that beauty leads to positive perceptions of personality and character, and that when these expectations are not met, the beautiful may be punished. This is another potential explanation of my finding that when personality is controlled for, many beauty premiums become beauty penalties.

3.9. Interpretation

The results described above clearly show that: (a) there is a beauty premium in some measures of academic performance when other ascriptive characteristics are not controlled for, and (b) there tends to be a beauty deficit in academic performance when other ascriptive characteristics are controlled for. Furthermore, these results demonstrate that: (c) the beautiful report putting less effort into their school work, when other ascriptive characteristics are controlled for, and (d) this less effort accounts for at least some of the beauty deficit in academic performance. Finally, the results demonstrate that, regardless of whether other ascriptive characteristics are controlled for, there is a beauty premium in romantic achievement.

Is it safe to conclude from these results that there is no "true" beauty premium in academic achievement? No: as discussed in the psychological and experimental literatures, the other ascriptive characteristics such as personality and grooming that explain the beauty premium may be themselves affected by beauty, in which case the specifications that control for the other ascriptive characteristics would suffer from over-controlling. But, in that case, it would be reasonable to conclude that *if* there is a beauty premium in academic performance, it is one that is mediated through beauty's effects on such variables as personality and grooming. This is a different conclusion from the one usually made in the empirical literature, and it is the main result of this paper.

Is it safe to conclude from these results that there is some kind of beauty deficit in academic achievement? Conditional on the strength of the identification strategy used in this area since Hamermesh & Biddle (1994), the answer is yes: if the specifications that include the other ascriptive characteristics suffer from over-controlling, then there is a beauty deficit in the sense that the effect of beauty on academic achievement that is not mediated through beauty's effect on other characteristics (such as personality and grooming) is negative. If, on the other hand, these specifications do not suffer from over-controlling, then one can conclude that the overall effect of beauty on academic achievement is likely to be negative, since these other characteristics are correlated with both beauty and outcomes, and so their omission would lead to omitted variable bias. Either way, this would be a type of beauty deficit.

Is it safe to conclude from these results that there is a beauty premium in romantic activities? Yes: it is clear that students who are more beautiful spend more time as members of a relationship, are more likely to have a relationship, and experience premiums on many related variables.

Finally, what can these results tell us about the conceptual explanation introduced in Section V? As explained at the end of Section VI, the results can largely be explained by the set of assumptions discussed in that section. But another possible explanation for some of these empirical results is the experimental findings of Andreoni and Petrie and Wilson and Eckel, who show that beautiful people are punished when it becomes clear that their personalities are not sufficiently beautiful as well.

3.10. Conclusion

In this paper, I attempt to reconcile the empirical literature that has found beauty premiums in numerous areas with an experimental literature which has found evidence that beauty premiums may be subtly related to premiums on other ascriptive characteristics, such as personality. Using a nationally representative sample, I find evidence for the first time of a beauty premium in high school performance. But because I can, for the first time, control for direct measurements of the attractiveness of personality and grooming, I am able to explain this premium on attractive physical features with premiums on attractive personality and grooming. Furthermore, I find that, as predicted by the conceptual explanation I propose in this paper, the physically beautiful experience more romantic achievement, put less effort into their school work when personality is controlled for, and experience less academic achievement in that case as well. Finally, the fact that controlling for personality is important in revealing the beauty deficit is also consistent with the experimental results of Andreoni and Petrie and Wilson and Eckel.

The literature on beauty premiums has become so ubiquitous that the recent Economist article of December 22nd, 2007 seeks to explain these premiums as a rational response of individuals to the biological *fact* that beautiful people are more intelligent than others. If the early studies of beauty had also controlled for personality and grooming measured on a one to five scale, perhaps the Economist would be opining instead on how good personalities and grooming are signs of biological superiority, or how a good personality is a positive externality in the modern workplace.

It remains to show in the other areas where beauty premiums have been found: (1) the explanatory power of premiums on the attractiveness of personality and attractiveness of grooming, (2) the relevance of the assumptions about the relationship between beauty and time use (and the predictions derived from them) that I have introduced in this paper, and (3) the persistence of the relationship between personality expectations and beauty suggested by the recent experimental literature.

Conclusion

In this dissertation, I analyze two areas related to labor demand: how labor demand responds to the supply of other labor inputs, and how academic achievement relates to ascriptive characteristics. Furthermore, I analyze the supply of workers who can choose their own hours. The results are as follows: (1) It appears that farm employers in Mexico increase their demand for adult labor when the supply of child labor decreases; (2) For some New York City taxi drivers, stopping work is more probable as their daily income increases; and (3) Any beauty premium in academic achievement in American middle and high schools is mediated through other ascriptive characteristics such as personality and grooming.

Future work must: (1) estimate the response of adult labor demand to child labor supply in other regions; (2) use structural modeling to help determine whether the daily income dependence of daily labor supply is related to reference points; and (3) replicate for labor market outcomes the study of premiums on other ascriptive characteristics.

References

Angelucci, Manuela and Giacomo De Girogi, "Indirect Effects of an Aid Program: the Case of Progresa and Consumption." Job Market Paper, University College London, 2005.

Basu, Kaushik. "Child Labor: Cause, Consequence, and Cure, with Remarks on International Labor Standards." *Journal of Economic Literature*, Vol. 37, No.3, 1999.

Basu, Kaushik. "The Intriguing Relation Between Adult Minimum Wage and Child Labour." *The Economic Journal*, 110 (March), 2000.

Basu, Kaushik and Zafiris Tzannatos. "The Global Child Labor Problem: What do we know and What Can We Do?" *The World Bank Economic Review*, Vol. 17, No. 2, 2003.

Basu, Kaushik and Pham Hoang Van. "The Economics of Child Labor." *The American Economic Review*, Vol. 88, No. 3, 1998.

Behrman, Jere R. and Petra Todd. "Randomness in the Experimental Samples of PROGRESA (Education, Health, and Nutrition Program)." Report Submitted to PROGRESA. Washington: International Food Policy Research Institute, 1999.

Biddle, Jeff E. and Hamermesh, Daniel S. "Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre," *Journal of Labor Economics*, University of Chicago Press, vol. 16(1), January 1998: pages 172-201.

Brown, Drusilla, Alan V. Deardorff and Robert M. Stern. "The Determinants of Child Labor: Theory and Evidence." *Research Seminar in International Economics Discussion Paper Series, University of Michigan*. Discussion Paper No. 486, September, 2002.

Cahuc, Pierre and Andre Zylberberg. Labor Economics, The MIT Press, Cambridge, MA (2004).

Camerer, Colin; Linda Babcock; George Lowenstein; and Richard Thaler, "Labor Supply of New York City Cabdrivers: One Day at a Time," *Quarterly Journal of Economics*, May 1997, pp. 407-441.

Chou, Yuan K. "Testing Alternative Models of Labor Supply: Evidence from Taxi Drivers in Singapore," University of Melbourne, Department of Economics Research Paper No. 768, 2000.

Cipriani, G.P., and A. Zago. "Productivity or Discrimination? Beauty and the Exams" Università Degli Studi Di Verona, Dipartimento Di Scienze Economiche, Working Paper Series, #18, May 2005.

Costa, P. T., Jr., McCrae, R. R., and Siegler, I. C. "Continuity and change over the adult life cycle: Personality and personality disorders. In C. R. Cloninger (Ed.), Personality and Psychopatholgy (pp. 129-154). Washington, DC: American Psychiatric Press (1999).

Diamond, C. and Fayed, T. "Evidence on Substitutability of Adult and Child Labour." *Journal of Development Studies*. Vol. 34, No. 3 (February), 1998.

Edmonds, Eric. "Does Child Labor Decline with Improving Economic Status?" NBER Working Paper No. w10134 2003.

Farber, Henry S. "Is Tomorrow Another Day? The Labor Supply of New York City Cab Drivers," *Journal of Political Economy* 113 (February 2005).

Farber, Henry S. "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers," *American Economic Review*, (Forthcoming, 2008).

Fehr, Ernst and Lorenz Goette. "Inter-temporal Substitution at Work? Evidence from a Field Experiment," mimeo, April 2002.

French, Michael T. "Physical Appearance and Earnings: Further Evidence" *Applied Economics*, 2002, 34, 569-572.
Galli, Rossana. "The Economic Impact of Child Labour." *Discussion Paper Series, International Institute of Labour Stuties*, DP/128, 2001.
Hamermesh, Daniel S. and Biddle, Jeff E. "Beauty and the Labor Market." *The American Economic Review* 84 (1994): 1174-1194.
Handa, Sudhanshu, Mari-Carmen Huerta, Raul Perez, and Beatriz Straffon. "Poverty, Inequality, And Spill-Over In Mexico's Education, Health, And Nutrition Program." *International Food Policy Research Institute*, Final Report. April, 2000.
Harris, Kathleen Mullan, Francesca Florey, Joyce Tabor, Peter S. Bearman, Jo Jones, and J. Richard Udry. 2003a The National Longitudinal Study of Adolescent Health: Questions about field work[WWW document]. URL: http://www.cpc.unc.edu/projects/addhealth/faqs/aboutfieldwork
Harris, Kathleen Mullan, Francesca Florey, Joyce Tabor, Peter S. Bearman, Jo Jones, and J. Richard Udry. 2003b The National Longitudinal Study of Adolescent Health: Design Facts at a Glance[WWW document]. URL: http://www.cpc.unc.edu/projects/addhealth/designfacts
Hernández, Daniel, José Gómez de León, and Gabriela Vásquez. "El Programa de Educación, Salud y Alimentación: orientaciones y components." In *Más opportunidades par alas familias pobres: evaluación de resultados del Programa de Educación, Salud y Alimentación, primeros avances*. Mexico City: Secretaria de Desarrollo Social, 1999.
International Labor Organization, *WORLD OF WORK No. 17*, September/October 1996
De Janvry, Alain and Elisabeth Sadoulet. Making Conditional Cash Transfer Programs More Efficient: Designing for Maximum Effect of the Conditionality. University of California at Berkeley. May 2005
Levison, D., R. Anker, S. Ashraf and S. Barge. "Is Child Labor Really Necessary in India's Carpet Industry?" In R. Anker, S. Barge, S. Rajagopal, and M. P. Joseph, eds., *Economics of Child Labor in Hazardous Industries of India*. New Delhi: Hindustan Publishers, 1998.
Long, J. Scott and Jeremy Freese, 2005, *Regression Models for Categorical Outcomes Using Stata*. Second Edition. College Station, TX: Stata Press.
Mankiw, N. Gregory, "Principles of Microeconomics" 4th Edition, 2006
Oettinger, Gerald S. "An Empirical Analysis of the Daily Labor Supply of Stadium Vendors," *Journal of Political Economy* (April 1999), pp. 360-392.
Pfann, Gerard A., Jeff E. Biddle, Daniel S. Hamermesh, Ciska M. Bosman. "Business Success and Businesses' Beauty Capital" *Economics Letters* 67 (2000) 201-207.
Rabin, Matthew and Botond Kőszegi. "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics* (2006), 121(4), pp. 1133-1166.
Ranjan, Priya. "Credit Constraints and the Phenomenon of Child Labor." *Journal of Development Econonomics*. 64 February, 2001.
Ray, Ranjan. "Analysis of Child Labour in Peru and Pakistan: A Comparative Study," *Journal of Population Economics*. Vol. 13, No. 1, March, 2000.
Roberts, Brent W., Wolfgang Viechtbauer, and Kate E. Walton. "Personality Traits Change in Adulthood: Reply to Costa and McCrae" *Psychological Bulletin* (2006), Vol. 132. No 1. 29-32.
Schultz, T. P. "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program." *Economic Growth Center Paper 834*. Yale University, New Haven, 2001.

Skoufias, Emmanuel, Benjamin Davis, and Jere Behrman. "An Evaluation of the Selection of Beneficiary Householdsin the Education, Health, and Nutrition Program (PROGRESA) of Mexico." *International Food Policy Research Institute*, Final Report. June, 1999.

Skoufias, Emmanuel and Susan Parker. "Conditional Cash Transfers and their Impact on Child Work and Schooling: Evidence from the PROGRESA Program in Mexico." *Economia*, Fall 2001.

Tversky, Amos and Daniel Kahneman. "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics* 106 (November 1991), 1039-1061.

Udry, Christopher. August, 2004. "Child Labor." http://www.econ.yale.edu/~cru2/pdf/kid.pdf

Varian, Hal. *Intermediate Microeconomics: A Modern Approach*. W. W. Norton & Company; 5th edition, February, 1999.

Figure 1.1: States in Mexico where the PROGRESA Experiment took place

State	Number of Observations in 1997	Percent
Guerrero	10,419	8.29
Hidalgo	21,645	17.22
Michoacán	15,133	12.04
Puebla	19,683	15.66
Queretaro	7,310	5.82
San Luis Potosí	20,125	16.01
Veracruz	31,359	24.95
Total	*125,674*	*100.00*

Figure 1.2: Age Frequency of Jornaleros earning a salary

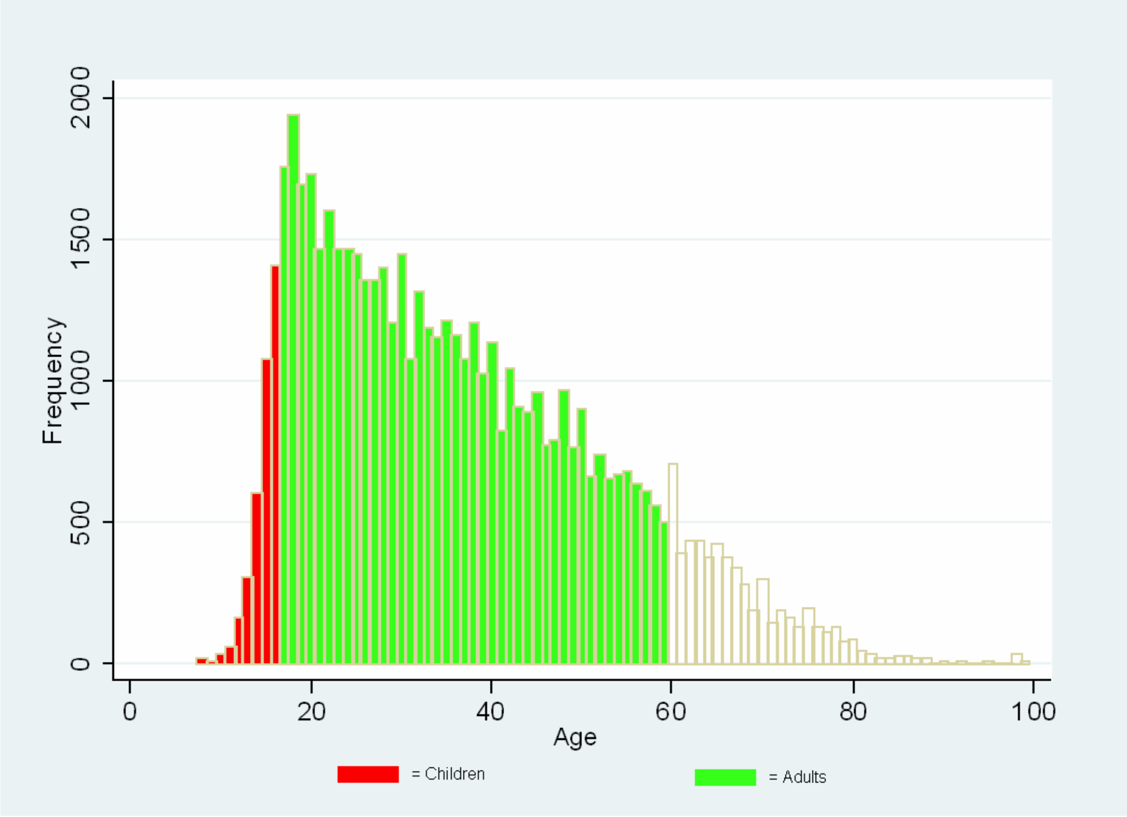


Figure 1.3: Most Paid Workers are Jornaleros

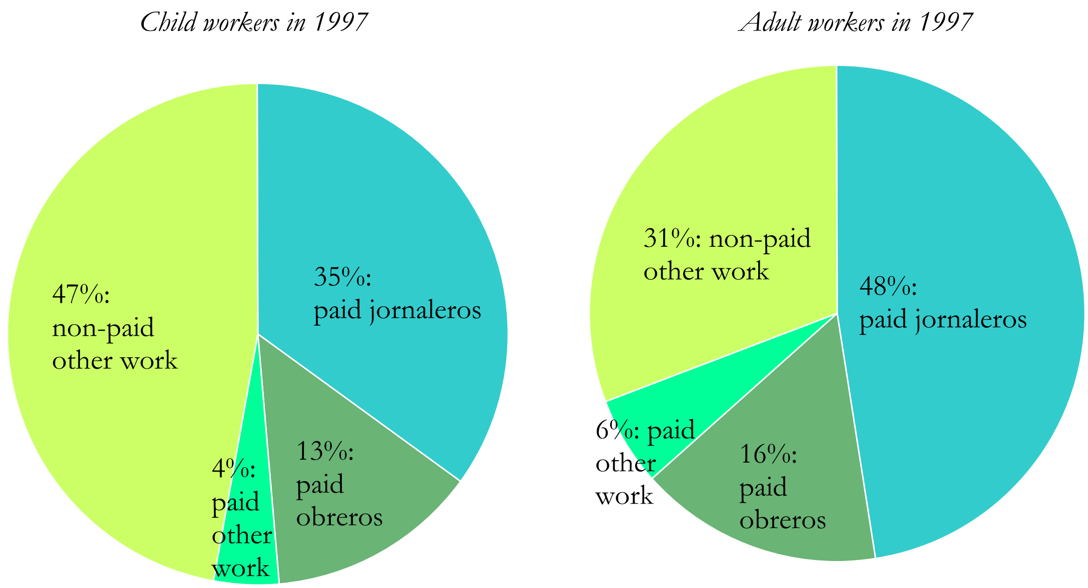
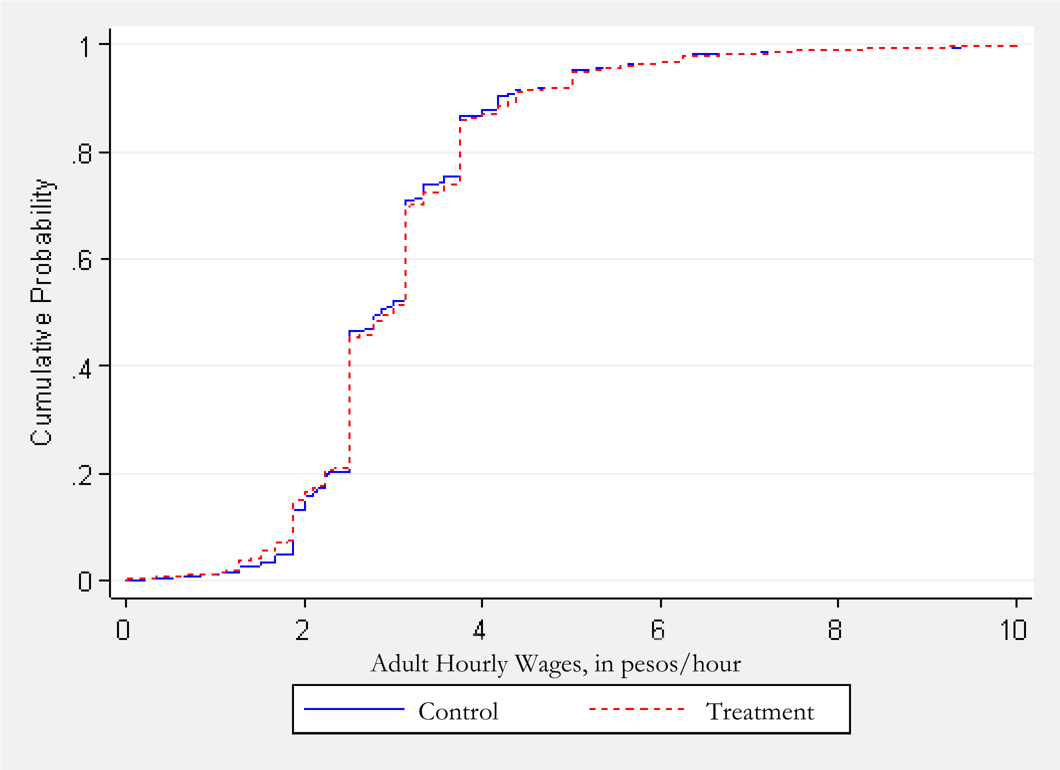


Figure 1.4: Cdfs of Hourly Wages, Control vs. Treatment, 1997 & 1999

1997:



1999:

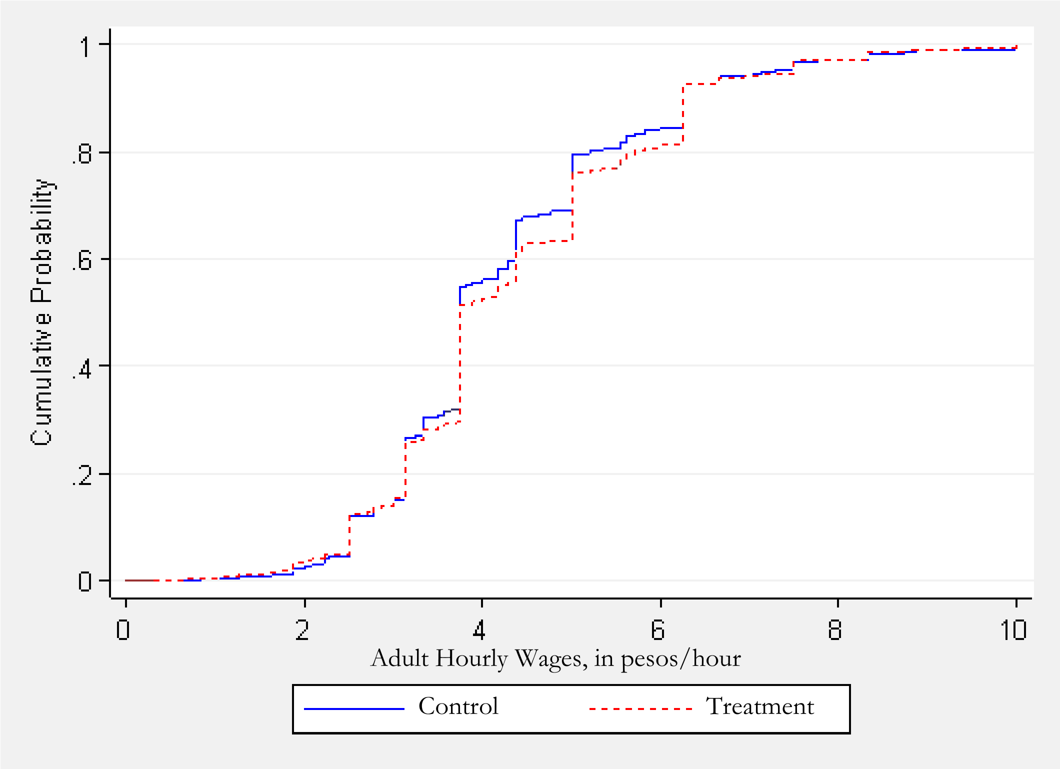
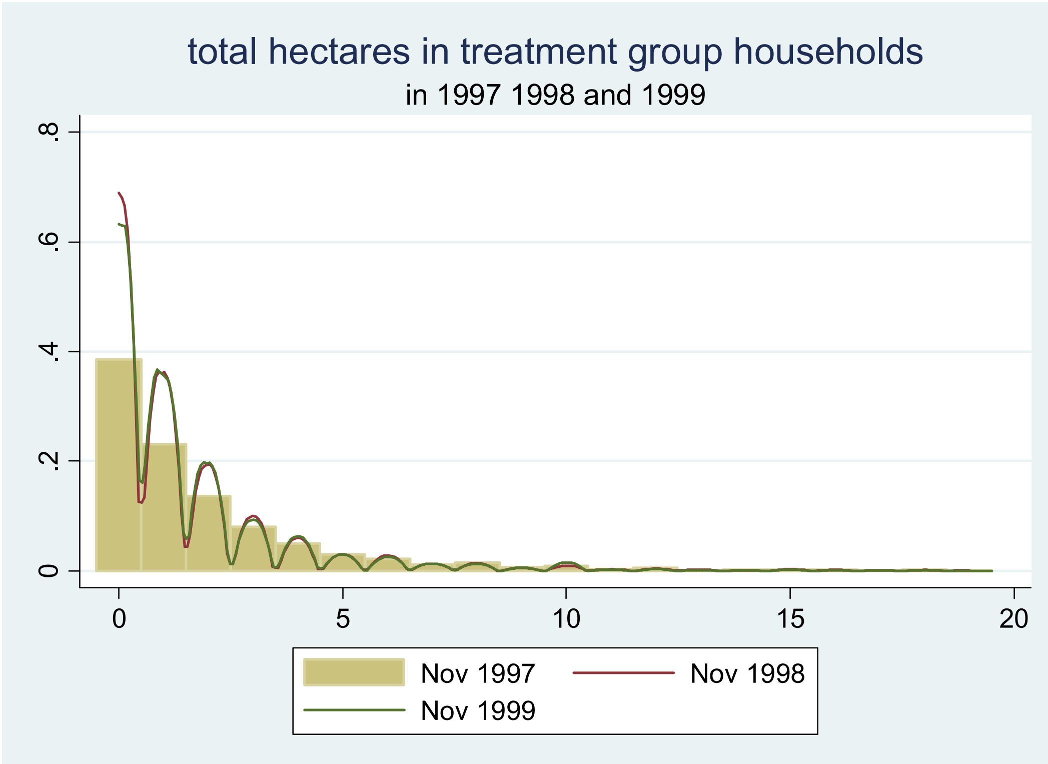


Figure 1.5: Total Hectares used or owned by treatment group vs. control group households, from 1997 through 1999



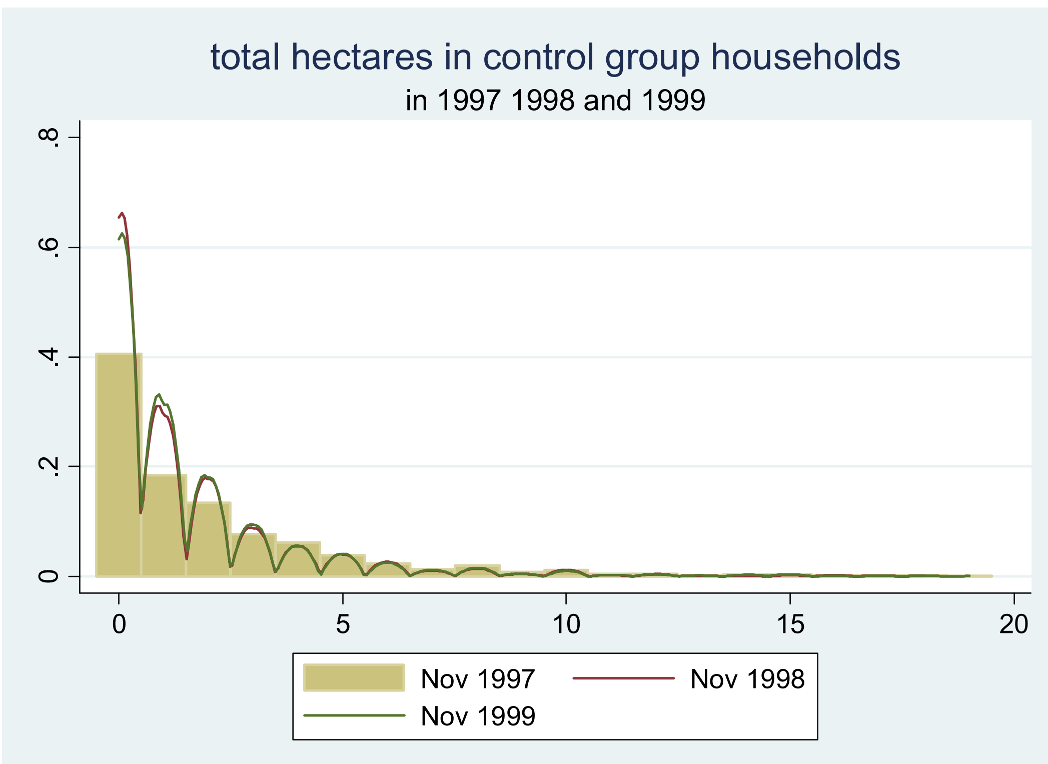


Figure 1.6. Percentage breakdown, by eligibility status, of families in treatment villages in 1997

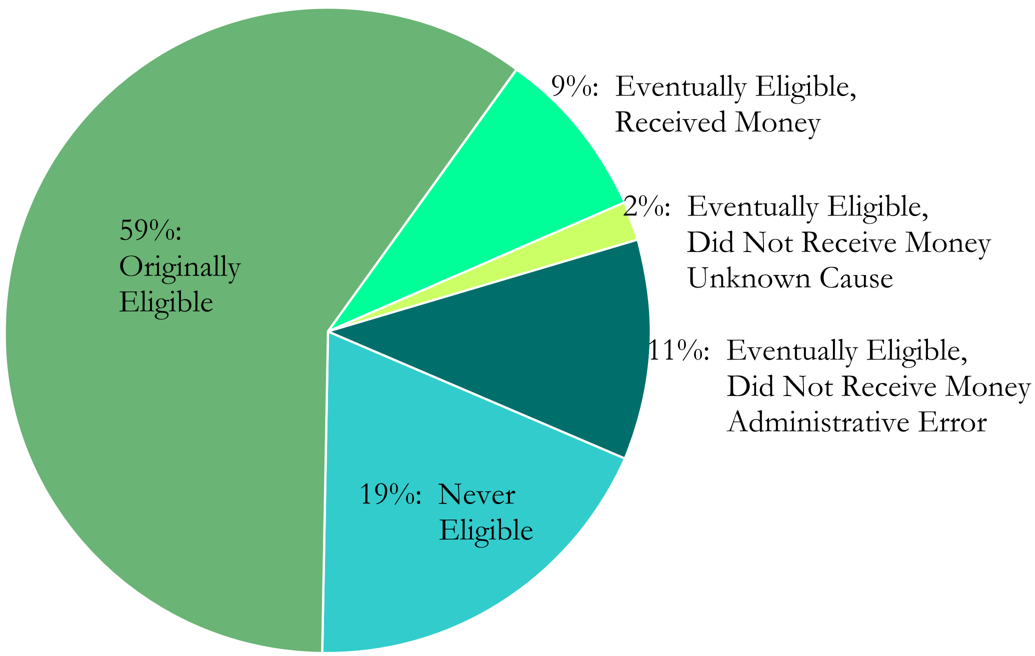


Table 1.1: PROGRESA Monthly Cash Transfer Schedule* (in Nominal Pesos)

Type of Grant	Amount of Grant Jan. to June 1998	July to Dec. 1998	Jan. to June 1999	July to Dec. 1999
Educational grant per child				
Primary (both genders)				
Third grade	65	70	75	80
Fourth grade	75	80	90	95
Fifth grade	95	100	115	125
Sixth grade	130	135	150	165
Secondary (boys)				
First grade	190	200	220	240
Second grade	200	210	235	250
Third grade	210	220	245	265
Secondary (girls)				
First grade	200	210	235	250
Second grade	220	235	260	280
Third grade	240	255	285	305
Grant for school materials per child				
Primary (September)	--	In-kind	--	110
Primary (January)	40	--	45	--
Secondary (September)	--	170	--	205
Grant for food consumption per household				
Cash Transfer	95	100	115	125
Maximum grant per household	585	635	695	750

* Source: Skoufias & Parker (2001) and Hernández, Gómez de León, & Vásquez (1999).

Table 1.2: First Response to Principal Activity & Crop Questions, Local Survey, 1997

Question	Response	Villages listing response	Percentage
Principal Activity in this village?	Agriculture	491	97.8%
	Commerical	3	0.6%
	Ganaderia	3	0.6%
	Artisan Production	1	0.2%
	Construction	1	0.2%
	Industrial Production	1	0.2%
	Services	1	0.2%
	Other	1	0.2%
Principal Crop in this village?	Corn (Maiz)	443	88.2%
	Beans	20	4.0%
	Coffee	19	3.8%
	Haba	2	0.4%
	Other	18	3.6%

Table 1.3a: Summary Statistics by Treatment Village Status and Year

Year	Variable	Control Villages	Treatment Villages
1997	Total # families	9,221 families	14,856 families
	Total # people	48,475 people	77,199 people
	% male	*50.0%*	*50.7%*
	% child (< 17 years)	46.8%	47.3%
	% adult (17 to 59 years)	45.3%	44.8%
	% worked last week	*40.0%*	*41.9%*
	% worked as jornalero	15.6%	15.2%
	Mean jornalero wage	3.36 pesos / hour	3.38 pesos / hour
1998	Total # families	9,919 families	15,927 families
	Total # people	52,299 people	85,141 people
	% male	50.0%	50.6%
	% child (< 17 years)	*47.5%*	*48.1%*
	% adult (17 to 59 years)	*44.7%*	*44.1%*
	% worked last week	35.7%	36.2%
	% worked as jornalero	21.4%	21.8%
	Mean jornalero wage	4.39 pesos / hour	4.37 pesos / hour
1999	Total # families	10,498 families	16,474 families
	Total # people	55,793 people	83,631 people
	% male	*49.6%*	*50.3%*
	% child (< 17 years)	45.9%	46.3%
	% adult (17 to 59 years)	46.0%	45.5%
	% worked last week	35.6%	36.0%
	% worked as jornalero	22.7%	22.5%
	Mean jornalero wage	5.1 pesos / hour	5.65 pesos / hour

Entries are italicized if they are significantly different between control and treatment at the 5% level.

Table 1.3b: Pre-treatment Distribution of Adults and Children across job categories

Year	% with Job Title:	Adults	Children
	Jornalero (field worker)	15,675 (50%)	1,701 (38%)
	Obrero (non-field worker)	5,320 (17%)	642 (15%)
	Self-employed	4,472 (14%)	317 (7%)
	Pattern Work	150 (0%)	9 (0%)
1997	Family Work, No Pay	3,428 (11%)	1,654 (37%)
	Other Work, No Pay	119 (0%)	50 (1%)
	Member of Cooperative	28 (0%)	3 (0%)
	Communal Farmer	2,245 (7%)	21 (0%)
	Other	229 (1%)	25 (1%)
	Total	31,666 (100%)	4,422 (100%)

Table 1.4a: OLS Treatment effects on Child Jornalero Work Participation

Dependent Variable: work participation in jornalero work force, for children aged less than 17 years old (Baseline year: 1997).

Explanatory Variables	(1) Post-Treatment: 1998	(2) Post-Treatment: 1999
Treated (post = 1 & treatment village = 1)	-0.0053 (0.0033)	-0.0071** (0.0032)
Post-treatment Dummy	-0.0101*** (0.0026)	-0.0129*** (0.0025)
Male Dummy	00727*** (0.0016)	0.0700*** (0.0016)
Age Dummies	YES	YES
Village Fixed Effects	YES	YES
Constant	0.1430*** (0.0028)	0.0991*** (0.0027)
# Observations	63488	62075
R2	0.1009	0.1025

Standard Errors are in Parenthesis
* = significant at 10%
** = significant at 5%
*** = significant at 1%

Table 1.4b Summary of OLS Treatment effects on Child Jornalero work:

Percentage change in the probability of a Child reporting Jornalero work from 1997 to . .

1998	1999
-9.81% (0.103)	**-13.05%** **(0.028)**

P-values for t-tests of significant difference from 0 are given in parenthesis.
Percentage changes are calculated by dividing the coefficient on the treated dummy from table 3a by the pre-treatment mean value of the independent variable to obtain the treatment effect.

Table 1.4c: Probit Treatment Effects on Child Jornalero Work Participation

Dependent Variable: work participation in jornalero work force,
for children aged less than 17 years old (Baseline year: 1997).

Explanatory Variables	(1) Post-Treatment: 1998	(2) Post-Treatment: 1999
Treated (post = 1 & treatment village = 1)	-0.0019** (0.0009)	-0.0021*** (0.0006)
Post-treatment Dummy	-0.0026*** (0.0007)	-0.0031*** (0.0006)
Male Dummy	0.0297*** (0.0013)	0.0243*** (0.0000)
Age Dummies	YES	YES
Village Fixed Effects	YES	YES
# Observations	61128	58852
Pseudo R2	0.33	0.36

Coefficients reported are the marginal effects
Standard Errors are in Parenthesis
* = significant at 10%
** = significant at 5%
*** = significant at 1%

Table 1.4d Summary of Probit Treatment effects on Child Jornalero work:

Percentage change in the probability of a Child reporting Jornalero work from 1997 to . .

1998	1999
-3.52%	**-3.86%**
(0.036)	**(0.002)**
[0.146]	**[0.045]**

P-values for t-tests of significant difference from 0 are given in parenthesis.
P-values for t-tests calculated with standard errors clustered at the village level are in brackets.
Percentage changes are calculated by dividing the coefficient on the treated dummy from table 3a by the pre-treatment mean value of the independent variable to obtain the treatment effect.

Table 1.5a: Treatment effects on log hourly wages and log daily income from 1997 to 1999

Dependent Variables:
Log Hourly Wages or Log Daily Income for Adult (ages 17 to 59) Jornaleros

Explanatory Variables	(1) Log Hourly Wage	(2) Log Hourly Wage	(3) Log Daily Income	(4) Log Daily Income
Treated (post = 1 & treatment village = 1)	0.065***	0.065***	0.061***	0.061***
	(0.006)	(0.02)	(0.006)	(0.016)
Treatment Village Indicator	-0.039***	-0.039**	-0.029***	-0.029
	(0.005)	(0.018)	(0.005)	(0.018)
Post-treatment Indicator	0.332***	0.332***	0.314***	0.314***
	(0.005)	(0.011)	(0.005)	(0.010)
Male Indicator	0.010	0.010	0.037***	0.037***
	(0.07)	(0.011)	(0.007)	(0.012)
Age, Schooling Level, Langauage Skills, and Marriage Status Indicators	YES	YES	YES	YES
Village Clusters		YES		YES
Constant	1.12***	1.12***	3.17***	3.17***
	(0.014)	(0.023)	(0.014)	(0.023)
# Observations	24605	24605	24605	24605
R2	0.42	0.42	0.41	0.41

Standard Errors are in Parenthesis
* = significant at 10%
** = significant at 5%
*** = significant at 1%

Table 1.5b: Treatment effects on log hours worked and days worked per week from 1997 to 1999

Dependent Variables: Log Hours Worked and Log Days Worked per Week for Adult (ages 17 to 59) Jornaleros				
Explanatory Variables	(1) Log Hours per Week	(2) Log Hours per Week	(3) Log Days per Week	(4) Log Days per Week
Treated (post = 1 & treatment village = 1)	0.035*** (0.009)	0.035* (0.020)	0.039*** (0.009)	0.039** (0.018)
Treatment Village Indicator	-0.026*** (0.007)	-0.026 (0.016)	-0.036*** (0.006)	-0.036** (0.015)
Post-treatment Indicator	-0.077*** (0.007)	-0.077*** (0.015)	-0.060*** (0.007)	-0.059*** (0.013)
Male Indicator	0.102*** (0.011)	0.102*** (0.017)	0.075*** (0.010)	0.075*** (0.015)
Age, Schooling Level, Langauage Skills, and Marriage Status Indicators	YES	YES	YES	YES
Village Clusters		YES		YES
Constant	3.60*** (0.021)	3.60*** (0.027)	1.56*** (0.019)	1.56*** (0.024)
# Observations	24575	24575	24575	24575
R2	0.02	0.02	0.01	0.01

Standard Errors are in Parenthesis
* = significant at 10%
** = significant at 5%
*** = significant at 1%

Table 1.6: Treatment effects on other quantity of labor measures from 1997 to 1999

Dependent Variables:
Quantity of Labor Measures for adult (ages 17 to 59) Jornaleros

Explanatory Variables	(1) Probit: Worked as Jornalero	(2) Probit: Worked as Jornalero	(3) OLS: Hours per Week with 0's	(4) OLS: Hours per Week with 0's
Treated (post = 1 & treatment village = 1)	3.6%* [0.08]	4.1% [0.428]	0.635*** (0.212)	0.690 (0.533)
Treatment Village Indicator		-1.5% [0.78]		-0.501 (0.561)
Post-treatment Indicator	22.7%*** [0.01]	21.9*** [0.00]	1.07*** (0.165)	-1.10*** (4.00)
Male Indicator	203%*** [0.00]	200%*** [0.00]	22.8*** (0.106)	22.9*** (0.434)
Age, Schooling Level, Langauage Skills, and Marriage Status Indicators	YES	YES	YES	YES
Village Fixed Effects	YES		YES	
Village Clusters		YES		YES
Constant			1.07*** (0.319)	38.5*** (1.2)
# Observations	103402	103402	102517	102517
R2	0.39	0.34	0.31	0.32

Standard Errors are in Parenthesis
* = significant at 10%
** = significant at 5%
*** = significant at 1%
P-values for Probits are in brackets

Table 1.7. Treatment Effects on Hectares used or owned, Total and Agricultural

	Dependent Variables		
Independent Variables:	(1) Total Hectares in 1998 minus Total Hectares in 1997	(2) Total Hectares in 1999 minus Total Hectares in 1997	(3)Agricultural Hectares in 1998 minus Agricultural Hectares in 1997
Treatment Village Dummy	0.002 (0.069)	0.018 (0.065)	0.027 (0.057)
Constant	-0.521** (0.054)	-0.608** (0.051)	-0.413** (0.045)
Village Clustering	NO	NO	NO
# Obs	22265	20648	24077
R2	0.00	0.00	0.00

Standard Errors in Parenthesis, * = significant at 10%, ** = 5%, *** = 1%. The unit of obs. is an individual household, with differences calculated from a panel data set of households. Village clustering is not performed because this shows that even under generous standard errors (i.e., those without clustering), there is no statistically significant treatment effect on hectares.

Table 1.8. Treatment Effects on the Corn Harvest in 1998

	Dependent Variables	
Independent Variables:	(1) Indicator for a non-zero Harvest in 1998 (probit)	(2) Log # Tons Corn Harvested during the 1998 Harvest (OLS)
Treatment Village Dummy	-0.023 (0.028)	-0.051 (0.096)
Age	-0.001*** (0.000)	0.006*** (0.001)
Male	0.020 (0.019)	-0.224*** (0.067)
Constant		-0.30** (.12)
Village Clustering	YES	YES
# Obs	11461	7007
R2	0.002	0.005

Standard Errors are in Parenthesis, * = significant at 10%** = significant at 5%, *** = significant at 1%

Table 1.9. Effect of increased wages for Adult Jornaleros on their Families' Consumption in 1999

	For Families *with* an Adult Jornalero			For Families *without* an Adult Jornalero		
Indep. Variables	(1) Log kg of fruits & vegetables	(2) Log kg of grains & cereals	(3) Log kg of meats and dairy	(1) Log kg of fruits & vegetables	(2) Log kg of grains & cereals	(3) Log kg of meats and dairy
Treatment Village Dummy	0.059 (0.042)	0.050* (0.027)	0.084** (0.036)	0.014 (0.040)	0.038 (0.026)	0.017 (0.038)
# of household members	0.078*** (0.006)	0.113*** (0.005)	0.071*** (0.005)	0.104*** (0.007)	0.153*** (0.006)	0.083*** (0.007)
Gender, Age, Schooling, Langauage, Marriage Status	YES	YES	YES	YES	YES	YES
Village Clusters	YES	YES	YES	YES	YES	YES
Constant	1.1*** (0.15)	2.2*** (0.10)	.08 (0.14)	.92*** (0.18)	1.9*** (0.09)	-.01 (0.13)
# Obs	3747	3766	3482	3839	3874	3574
R2	0.11	0.27	0.12	0.20	0.37	0.15

Standard Errors are in Parenthesis
* = significant at 10%
** = significant at 5%
*** = significant at 1%

Table 1.10. Probit Treatment Effects on Health Outcomes in 1999

Indep. Variables	For Families *with* an Adult Jornalero (1) Difficulty Every Day	(2) Missed Activities Every Day	(3) In Bed Every Day	For Families *without* an Adult Jornalero (1) Difficulty Every Day	(2) Missed Activities Every Day	(3) In Bed Every Day
Treatment Village Dummy	-24% (0.001)	-16% (0.073)	-9% (0.426)	-18% (0.019)	-19% (0.010)	-18% (0.027)
Gender, Age, Schooling, Langauage, Marriage Status	YES	YES	YES	YES	YES	YES
Village Clusters	YES	YES	YES	YES	YES	YES
R2	0.22	0.22	0.20	0.19	0.20	0.19
# Obs	13,468	13,449	13,447	11,115	11,114	11,114
Control Group Mean	14%	11%	0.8%	28%	25%	1.8%

P-values for t-tests of signficance of difference from zero in parenthesis

Table 2.1a: Simple Statistics across Trips, with effects of final data cleaning

	Median	Mean	Std. Dev	Min.	Max.	# Obs.
Trip's Fare (in $)	$5.90	$7.93	$8.16	$0.00	$1,750.00	165020
Trip's Fare (cleaned)	$5.90	$7.94	$6.93	$0.30	$601.00	164503
Trip's Time (in Minutes)	10	18.5	76.2	-56	1435	167084
Trip's Time (cleaned)	10	13.0	12.2	1	300	164120
Wait Time Since Previous Fare (in Minutes)	4	42.6	208	0	1439	161153
Wage Since Start of Shift (in $/minute)	$0.39	$0.42	$0.64	$0.00	$159.75	173635
Wage Since Start of Shift (cleaned)	$0.39	$0.40	$0.14	$0.00	$1.00	169978
Probability of Stopping	0	0.048	0.214	0	1	175063

Table 2.1b: Shift Level Summary Statistics over Drivers

Driver	# of Shifts	Average Trips	Working Hours	Driving Hours	Wait Hours	Break Hours	Average Total Income	Average Wage ($/hr)
Driver 1	120	23.18	6.86	4.70	2.15	1.31	$173.96	$25.37
Driver 2	114	24.58	7.11	4.33	2.78	0.39	$148.00	$20.83
Driver 3	65	15.97	3.96	2.81	1.15	0.97	$49.76	$12.57
Driver 4	24	20.42	5.08	3.55	1.53	2.21	$121.79	$23.97
Driver 5	3	13.00	2.26	1.46	0.79	0.00	$48.40	$21.46
Driver 6	37	12.27	3.87	2.38	1.49	1.23	$78.71	$20.35
Driver 7	3	11.33	2.96	2.17	0.79	3.19	$93.03	$31.48
Driver 8	166	18.77	5.89	4.01	1.88	1.20	$148.99	$25.28
Driver 9	183	14.92	5.33	4.15	1.18	2.40	$201.08	$37.74
Driver 10	178	22.01	6.27	3.89	2.38	1.69	$139.87	$22.29
Driver 11	130	21.71	6.68	4.43	2.24	0.94	$200.10	$29.98
Driver 12	23	28.87	7.80	5.02	2.78	0.89	$165.56	$21.23
Driver 13	65	19.35	5.21	3.41	1.80	0.99	$127.65	$24.52
Driver 14	82	24.40	5.77	3.59	2.17	0.37	$150.18	$26.05
Driver	70	16.20	4.31	2.98	1.33	0.36	$128.67	$29.84

15								
Driver 16	159	21.52	6.96	4.64	2.31	1.33	$178.79	$25.70
Driver 17	80	13.60	3.72	2.91	0.81	0.33	$114.81	$30.90
Driver 18	129	13.94	4.40	2.96	1.44	0.73	$124.12	$28.23
Driver 19	4	22.50	7.35	5.15	2.20	0.24	$156.68	$21.30
Driver 20	132	18.12	5.93	4.36	1.57	0.62	$134.60	$22.68
Driver 21	166	12.34	4.19	3.14	1.06	0.92	$127.02	$30.29
Driver 22	21	19.81	6.75	4.19	2.55	1.92	$149.63	$22.18
Driver 23	114	22.54	5.98	3.97	2.01	0.73	$164.73	$27.53
Driver 24	103	11.12	3.90	2.72	1.19	2.02	$98.63	$25.28
Driver 25	156	25.70	8.39	6.30	2.09	0.89	$198.77	$23.69
Driver 26	68	28.25	8.29	5.28	3.00	0.55	$175.39	$21.17
Driver 27	143	14.42	6.09	4.62	1.47	2.10	$148.60	$24.41
Driver 28	76	32.87	8.72	6.32	2.39	1.16	$227.67	$26.12
Driver 29	28	30.00	9.24	6.00	3.25	2.19	$223.16	$24.14
Driver 30	53	14.15	3.62	2.67	0.94	0.69	$85.45	$23.62
Driver 31	9	9.56	2.36	1.71	0.64	1.21	$75.40	$32.01
Driver 32	4	27.00	8.46	5.53	2.93	1.37	$192.25	$22.73
Driver 33	18	9.00	2.60	1.90	0.70	0.55	$85.87	$33.02
Driver 34	1	18.00	3.75	2.58	1.17	7.75	$138.50	$36.93
Driver 35	93	20.08	5.90	4.02	1.88	1.16	$157.36	$26.65
Driver 36	91	22.23	7.12	4.79	2.34	1.01	$203.23	$28.54
Driver 37	1	15.00	3.23	1.92	1.32	0.00	$74.80	$23.13
Driver 38	5	19.00	5.32	3.48	1.84	0.39	$139.20	$26.15
Driver 39	141	30.76	9.44	6.06	3.38	1.91	$231.31	$24.52
Driver	182	22.26	7.59	5.30	2.28	1.63	$168.81	$22.25

40								
Driver 41	17	22.94	6.96	4.73	2.23	1.17	$143.97	$20.68
Driver 42	23	13.39	4.09	2.75	1.35	0.20	$84.36	$20.61
Driver 43	137	25.84	7.18	4.76	2.42	0.85	$172.25	$24.01
Driver 44	145	15.30	5.37	3.83	1.54	0.73	$133.51	$24.85
Driver 45	125	28.71	8.57	5.58	2.99	1.04	$213.71	$24.95
Driver 46	47	24.49	7.21	4.60	2.61	1.49	$206.54	$28.66
Driver 47	1	21.00	5.75	4.08	1.67	0.00	$233.80	$40.66
Driver 48	1	11.00	5.67	3.52	2.15	0.00	$181.40	$32.01
Driver 49	12	16.25	4.59	3.03	1.56	1.24	$137.25	$29.89
Driver 50	70	19.10	6.49	4.56	1.93	1.26	$141.48	$21.81
Driver 51	66	27.41	7.55	5.10	2.44	0.51	$208.62	$27.65
Driver 52	107	22.98	6.12	4.24	1.88	0.40	$157.80	$25.79
Driver 53	119	25.08	7.06	4.68	2.38	1.11	$153.79	$21.79
Driver 54	74	22.36	5.91	4.26	1.65	0.81	$130.15	$22.02
Driver 55	44	25.68	9.19	6.69	2.50	2.45	$233.60	$25.41
Driver 56	9	23.33	8.54	6.25	2.29	3.21	$302.48	$35.44
Driver 57	31	28.03	9.16	6.58	2.58	2.35	$249.64	$27.26
Driver 58	12	13.17	4.08	2.84	1.24	0.84	$85.60	$20.96
Driver 59	46	7.72	2.71	1.98	0.73	0.49	$63.51	$23.44
Driver 60	2	5.00	1.28	1.09	0.19	0.00	$20.55	$16.01
Driver 61	4	37.50	8.79	6.85	1.93	0.75	$301.85	$34.35
Driver 62	128	20.79	5.72	3.54	2.18	7.28	$134.82	$23.58
Driver 63	6	22.67	6.73	4.24	2.48	0.47	$112.35	$16.71
Driver 64	5	7.20	2.19	1.54	0.65	1.88	$67.90	$31.00
Driver	121	25.21	7.90	4.74	3.16	0.00	$204.95	$25.93

65								
Driver 66	175	17.29	5.36	3.60	1.75	0.93	$130.23	$24.32
Driver 67	46	11.02	3.97	2.86	1.11	1.41	$126.87	$31.96
Driver 68	20	20.35	7.65	5.21	2.44	2.00	$166.08	$21.72
Driver 69	76	24.12	7.43	4.96	2.47	1.02	$248.51	$33.45
Driver 70	140	30.08	8.86	6.00	2.87	2.25	$182.73	$20.62
Driver 71	86	26.80	8.95	5.60	3.35	0.65	$213.28	$23.84
Driver 72	7	39.14	12.46	7.84	4.62	2.25	$242.74	$19.48
Driver 73	71	18.04	4.97	3.56	1.41	3.69	$142.10	$28.62
Driver 74	86	20.53	6.46	4.72	1.74	0.41	$140.17	$21.68
Driver 75	305	20.21	5.82	3.77	2.05	0.19	$119.92	$20.60
Driver 76	327	16.28	4.68	3.02	1.66	1.28	$105.46	$22.52
Driver 77	57	20.05	5.19	3.51	1.68	1.23	$125.25	$24.13
Driver 78	96	19.15	6.15	4.60	1.55	0.35	$186.68	$30.37
Driver 79	139	20.73	6.25	4.10	2.15	0.32	$170.84	$27.33
Driver 80	14	10.93	4.17	3.34	0.82	0.95	$409.16	$98.23
Driver 81	184	20.82	5.51	3.73	1.77	3.04	$147.59	$26.80
Driver 82	39	13.72	3.58	2.41	1.18	1.43	$94.54	$26.40
Driver 83	17	10.00	2.61	1.76	0.85	0.29	$78.18	$29.99
Driver 84	143	22.58	7.69	5.14	2.55	0.00	$258.40	$33.61
Driver 85	182	14.05	3.63	2.40	1.23	3.83	$85.74	$23.60
Driver 86	57	23.21	5.75	4.27	1.48	0.87	$141.50	$24.61
Driver 87	136	20.70	6.48	4.08	2.40	0.36	$122.79	$18.95
Driver 88	115	22.60	6.89	3.53	3.36	0.72	$161.47	$23.44
Driver 89	183	18.96	5.99	3.79	2.20	0.39	$142.78	$23.86
Driver	150	24.78	7.17	4.93	2.24	2.04	$158.18	$22.07

90								
Driver 91	99	23.86	7.28	4.89	2.39	0.54	$191.16	$26.26
Driver 92	142	23.33	6.94	4.85	2.09	0.77	$149.48	$21.53
Driver 93	127	8.82	2.23	1.68	0.55	0.72	$60.75	$27.27
Driver 94	51	25.33	7.28	4.41	2.88	0.09	$149.16	$20.48
Driver 95	68	30.66	8.08	5.28	2.79	1.46	$204.58	$25.33
Driver 96	3	13.00	4.18	3.02	1.16	0.80	$144.43	$34.53
Driver 97	19	19.63	6.26	3.89	2.38	1.57	$132.49	$21.15
Driver 98	4	9.50	3.31	1.96	1.35	0.82	$98.45	$29.76
Driver 99	5	18.40	5.58	3.53	2.05	1.90	$128.02	$22.94
Driver 100	2	15.50	4.95	3.34	1.61	1.25	$100.80	$20.36
Driver 101	3	8.33	2.19	1.51	0.69	0.64	$77.37	$35.26
Driver 102	5	10.80	4.25	2.66	1.59	0.73	$126.12	$29.68
Driver 103	7	3.00	0.86	0.55	0.30	1.24	$26.80	$31.27
Driver 104	132	24.63	7.62	5.74	1.87	0.17	$172.60	$22.66
Driver 105	106	24.23	6.70	4.34	2.35	1.26	$167.66	$25.04
Driver 106	97	28.90	7.93	5.35	2.58	1.48	$202.78	$25.56
Driver 107	74	25.45	8.04	5.68	2.36	0.67	$143.43	$17.83
Driver 108	3	22.00	4.93	3.23	1.71	2.02	$111.97	$22.70
Driver 109	1	5.00	1.10	0.93	0.17	0.00	$31.50	$28.64
Driver 110	22	24.41	6.61	4.35	2.26	0.00	$191.48	$28.98
Driver 111	1	34.00	7.38	5.18	2.20	0.55	$169.50	$22.96
Driver 112	10	17.60	4.53	3.38	1.14	0.00	$105.18	$23.24
Driver 113	8	15.00	4.75	3.35	1.40	0.38	$83.59	$17.60
All	8,432	19.74	5.84	3.96	1.89	1.18	$149.53	$26.15

Table 2.2: Autocorrelations of Hourly Wages across the Hours of a Shift

Hour of Work	Autocorrelation with 1st Hour	Number of Observations
1st	1.00	8398
2nd	0.19	7188
3rd	0.16	6935
4th	0.18	6495
5th	0.14	5976
6th	0.08	5384
7th	0.13	4779
8th	0.14	3929
9th	0.08	2970

Table 2.3: Analysis of Variance of Hourly Wages

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Driver ID		X			X	X	X	X
Hour of Day			X		X	X	X	X
Day of Week				X	X			
Hour of Day * Day of Week							X	X
Date						X	X	X
Date * ID								X
Degrees of Freedom Used	0	113	23	6	142	376	503	
R Squared	0.00	0.07	0.02	0.00	0.09	0.11	0.12	
Root MSE	12.72	12.29	12.57	12.71	12.12	12.17	12.13	

As in Farber (2005a), I based the statistics in this table on linear regressions where the dependent variable is the hourly wage in a given hour and the independent variables are the indicator variables listed on the left side of the table.

Table 2.4: Marginal Effects from Probit Model of Stopping Probability as a Function of Income

	(1)	(2)	(3)	(4)
Income (in $100s)	0.0026**			
	(0.0010)			
$0 < Income ≤ $10		0.0244***	-0.0028	-0.0028
		(0.0065)	(0.0029)	(0.0057)
$10 < Income ≤ $20		0.0046	-0.0041	-0.0041
		(0.0032)	(0.0023)	(0.0043)
$20 < Income ≤ $30		-0.0025	-0.0059***	-0.0059
		(0.0021)	(0.0019)	(0.0037)
$30 < Income ≤ $40		-0.0027	-0.0045**	-0.0045
		(0.0018)	(0.0018)	(0.0030)
$40 < Income ≤ $50		-0.0052***	-0.0059***	-0.0059**
		(0.0014)	(0.0015)	(0.0023)
$50 < Income ≤ $75		-0.0046***	-0.0049***	-0.0049**
		(0.0011)	(0.0013)	(0.0017)
$75 < Income ≤ $100		-0.0040***	-0.0037***	-0.0037**
		(0.0009)	(0.0011)	(0.0015)
$100 < Income ≤ $125		-0.0029***	-0.0025**	-0.0025**
		(0.0009)	(0.0010)	(0.0011)
$125 < Income ≤ $140		-0.0028***	-0.0026**	-0.0026**
		(0.0009)	(0.0010)	(0.0010)
$160 < Income ≤ $175		0.0020**	0.0012	0.0012
		(0.0012)	(0.0012)	(0.0012)
$175 < Income ≤ $200		0.0041**	0.0011	0.0011
		(0.0013)	(0.0012)	(0.0012)
$200 < Income ≤ $225		0.0076**	0.0024*	0.0024
		(0.0016)	(0.0015)	(0.0017)
$225 < Income ≤ $250		0.0125**	0.0044**	0.0044**
		(0.0022)	(0.0019)	(0.0025)
$250 < Income ≤ $275		0.0164**	0.0050**	0.0050*
		(0.0030)	(0.0023)	(0.0033)
$275 < Income		0.0194**	0.0066***	0.0066**
		(0.0031)	(0.0025)	(0.0032)
Trip Controls	Indicators	log trips	Indicators	Indicators
Hours Worked Controls	Indicators	log hours	Indicators	Indicators
Day Shift Indicator	YES	YES	YES	YES
Days of Week Controls	Indicators	Indicators	Indicators	Indicators
Months of Year Controls	Indicators	Indicators	Indicators	Indicators
Hour of Day Controls	Indicators	Indicators	Indicators	Indicators
Driver Fixed Effects	YES	YES	YES	YES
Driver Clustering	YES	NO	NO	YES
Income Joint Signif.	p = (0.01)	p = (0.00)	p = (0.00)	p = (0.01)
# of Observations	170446	169300	170446	170446
R^2	0.23	0.22	0.23	0.23

* = Significant at 10%, ** = 5%, *** = 1% Standard Errors are in Parentheses

Table 2.5: Marginal Effects of Daily Income and Daily Hours on Stopping Probability from Estimating Individual-level Probit Models

Driver	$20 Daily Income Effect	1 Daily Hour Effect	Number Obs. (# of trips)	Pseudo R^2	Income Signif.	Hours Signif.	Std. Dev. of Shift Income
40	.0008** (.0005)	.0024*** (.0009)	2744	0.37	YES	YES	0.37
41	.0005 (.0008)	.0009** (.0005)	2789	0.30	NO	YES	0.25
42	.0139 (.0111)	.0055* (.0037)	471[i]	0.15	NO	YES	0.59
47	.0032 (.0020)	.0053*** (.0013)	3095	0.16	NO	YES	0.37
48	.0012 (.0012)	.0089*** (.0015)	2701	0.32	NO	YES	0.31
49	-.0004 (.0011)	.0079*** (.0012)	3769	0.23	NO	YES	0.41
50	.0028*** (.0010)	.0065*** (.0014)	2804	0.28	YES	YES	0.32
52	.0131*** (.0034)	.0045*** (.0015)	1255	0.25	YES	YES	0.33
53	.0017 (.0020)	.0019 (.0012)	1814	0.20	NO	NO	0.47
54	.0061*** (.0024)	.0060*** (.0025)	1124	0.34	YES	YES	0.25
55	-.0020 (.0014)	.0115*** (.0021)	3179	0.24	NO	YES	0.34
56	.0166** (.0078)	.0037 (.0075)	1057	0.14	YES	NO	0.41
57	.0228*** (.0034)	-.0016 (.0025)	1796	0.20	YES	NO	0.46
60	.0026 (.0019)	.0052*** (.0016)	2390	0.25	NO	YES	0.34
61	.0113*** (.0034)	.0105*** (.0022)	2036	0.18	YES	YES	0.38
63	.0119*** (.0024)	.0051*** (.0011)	2548	0.21	YES	YES	0.36
64	.0031 (.0044)	.0198*** (.0041)	1123	0.19	NO	YES	0.49
65	.0005 (.0005)	.0047*** (.0009)	3977	0.27	NO	YES	0.45
66	.0048*** (.0018)	.0024* (.0013)	1921	0.25	YES	YES	0.27
70	.0020 (.0020)	.0077*** (.0011)	2042	0.15	NO	YES	0.60
71	.0002 (.0003)	.0003 (.0004)	2448	0.23	NO	NO	0.51
78	.0011 (.0014)	.0045*** (.0017)	1139	0.28	NO	YES	0.32

79	.0014*** (.0007)	.0004* (.0003)	1534	0.34	YES	YES	0.39
82	.0020** (.0008)	.0008 (.0005)	4173	0.21	YES	NO	0.42
83	.0032** (.0013)	.0045*** (.0007)	4012	0.21	YES	YES	0.37
86	.0037*** (.0010)	.0014*** (.0005)	3524	0.29	YES	YES	0.34
87	.0146*** (.0025)	.0068*** (.0015)	2205	0.21	YES	YES	0.48
88	.0012* (.0007)	.0033*** (.0008)	3568	0.27	YES	YES	0.27
89	.0006 (.0006)	.0015*** (.0010)	1137	0.37	NO	YES	0.28
93	.0043* (.0024)	.0057*** (.0018)	1312	0.21	YES	YES	0.34
94	.0016*** (.0008)	.0014*** (.0007)	1801	0.33	YES	YES	0.33
95	.0059*** (.0018)	.0038*** (.0011)	2433	0.28	YES	YES	0.34
96	.0002 (.0002)	.0007*** (0003)	2923	0.34	NO	YES	0.38
97	.0041* (.0025)	.0040*** (.0014)	1348	0.25	YES	YES	0.29
98	.0015 (.0021)	.0047** (.0013	1129	0.16	NO	YES	0.52
106	-.0009 (.0023)	.0031*** (.0011)	2649	0.17	NO	YES	0.66
109	.0001 (.0002)	.0008*** (.0004)	3047	0.63	NO	YES	0.32
110	.0058** (.0025)	.0041*** (.0012)	2991	0.18	YES	YES	0.43
113	.0018** (.0008)	.0034*** (.0010)	1816	0.28	YES	YES	0.29
114	-.0006 (.0007)	.0022*** (.0005)	4191	0.26	NO	YES	0.31
115	.0007 (.0010)	.0022*** (.0007)	2280	0.22	NO	YES	0.38
117	.0026 (.0044)	.0168*** (.0040)	1200	0.21	NO	YES	0.53
118	-.0009 (.0020)	.0053*** (.0013)	1759	0.26	NO	YES	0.31
121	.0014* (.0008)	.0039*** (.0005)	5733	0.23	YES	YES	0.58
122	.0002 (.0013)	.0107*** (.0012)	4380	0.21	NO	YES	0.75
123	.0193*** (.0055)	-.0039 (.0029)	1140	0.18	YES	NO	0.40
124	.0030 (.0020)	.0106*** (.0030)	1837	0.26	NO	YES	0.30

125	.0043** (.0017)	.0094*** (.0016)	2825	0.18	YES	YES	0.43
127	.0002 (.0016)	.0080*** (.0010)	3687	0.10	NO	YES	0.69
130	.0000 (.0002)	.0039*** (.0008)	3221	0.20	NO	YES	0.64
131	.0060 (.0050)	.0070*** (.0014)	2457	0.16	NO	YES	0.45
132	.0002 (.0005)	.0020*** (0013)	1316	0.37	NO	YES	0.27
133	-.0002 (.0026)	.0021 (.0015)	2812	0.25	NO	NO	0.32
134	-.0003 (.0006)	.0031*** (.0012)	2597	0.35	NO	YES	0.23
135	.0079*** (.0017)	.0052*** (.0011)	3468	0.29	YES	YES	0.36
136	.0029*** (.0008)	.0020*** (.0007)	3712	0.30	YES	YES	0.23
137	.0024*** (.0008)	.0009*** (.0003)	2358	0.32	YES	YES	0.31
138	.0021 (.0019)	.0012 (.0017)	3309	0.20	NO	NO	0.32
139	.0235 (.0145)	-.0499*** (.0165)	1113	0.12	NO	YES	0.58
140	.0017 (.0012)	.0004 (.0005)	1291	0.31	NO	NO	0.29
141	.0010*** (.0005)	.0007*** (.0004)	2069	0.37	YES	YES	0.30
150	.0011 (.0013)	.0041*** (.0010)	3236	0.20	NO	YES	0.49
151	-.0008 (.0017)	.0065*** (.0011)	2568	0.25	NO	YES	0.34
152	.0013** (.0007)	.0032*** (.0010)	2803	0.34	YES	YES	0.26
154	-.0014** (.0009)	.0015*** (.0006)	1871	0.42	YES	YES	0.31

* = Significant at 10%, ** = 5%, *** = 1%   Standard Errors are in Parentheses
All specifications are simplified to include only the following independent variables:
Daily income so far, daily hours so far, number of trips completed so far, day shift indicator, day of the week indicators, and month indicators.

[i] This driver reports over 1,000 trips, but estimation of the individual-level probit model resulted in many of these observations being dropped.

Table 2.6. Effect of 26% Fare Increase on the Hours Worked and Income Earned per Shift, for the Reference-Dependent Drivers

	(1) Log of Total Minutes in Shift	(2) Log of Total Minutes in Shift	(3) Log of Total Income in Shift
After Wage Increase	0.07*** (0.02)	0.07* (0.04)	0.27*** (0.02)
Day Shift	0.25*** (0.03)	0.25*** (0.08)	0.03 (0.03)
Thursday	0.01 (0.03)	0.01 (0.03)	0.04 (0.03)
Friday	0.06** (0.03)	0.06 (0.04)	0.08*** (0.03)
Saturday	-0.01 (0.03)	-0.01 (0.05)	-0.02 (0.03)
Sunday	-0.22*** (0.03)	-0.22*** (0.05)	-0.20*** (0.03)
Monday	-0.04 (0.03)	-0.04 (0.03)	-0.10*** (0.03)
Tuesday	-0.07** (0.03)	-0.07*** (0.03)	-0.10*** (0.03)
Driver Fixed Effects	YES	YES	YES
Driver Clustering	NO	YES	NO
R^2	0.29	0.29	0.32
# observations	3068	3028	3133

* = Significant at 10%, ** = 5%, *** = 1% Standard Errors are in Parentheses

Figure 2.1: Density of Minutes Driven Per Shift

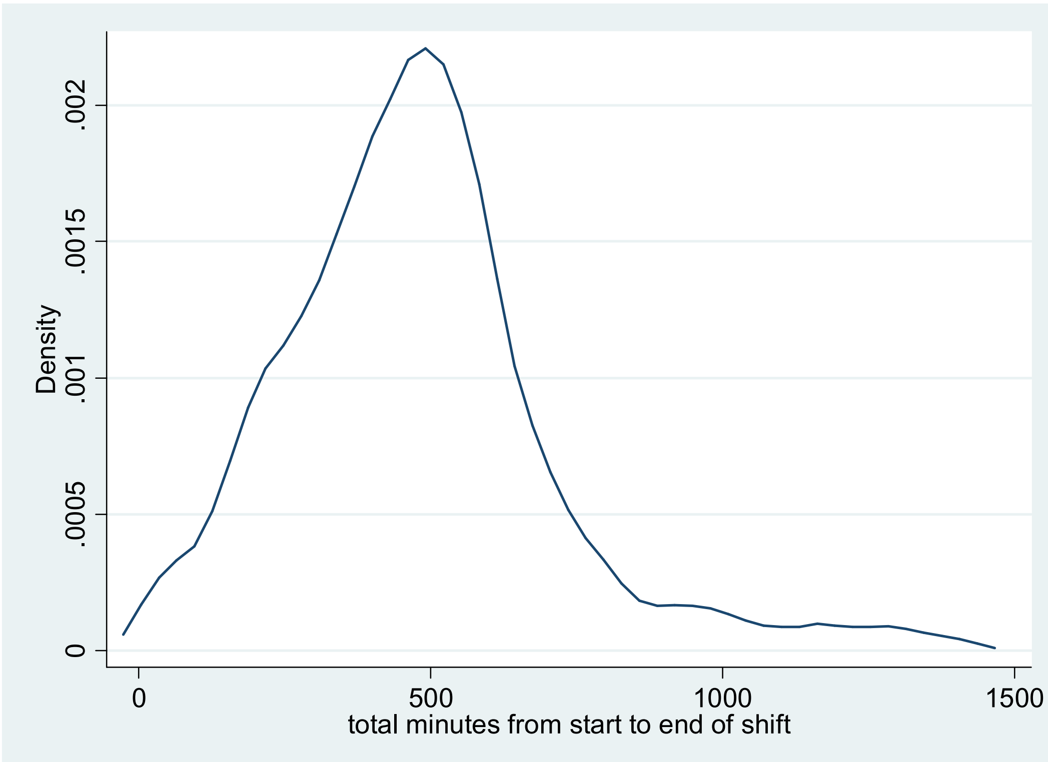
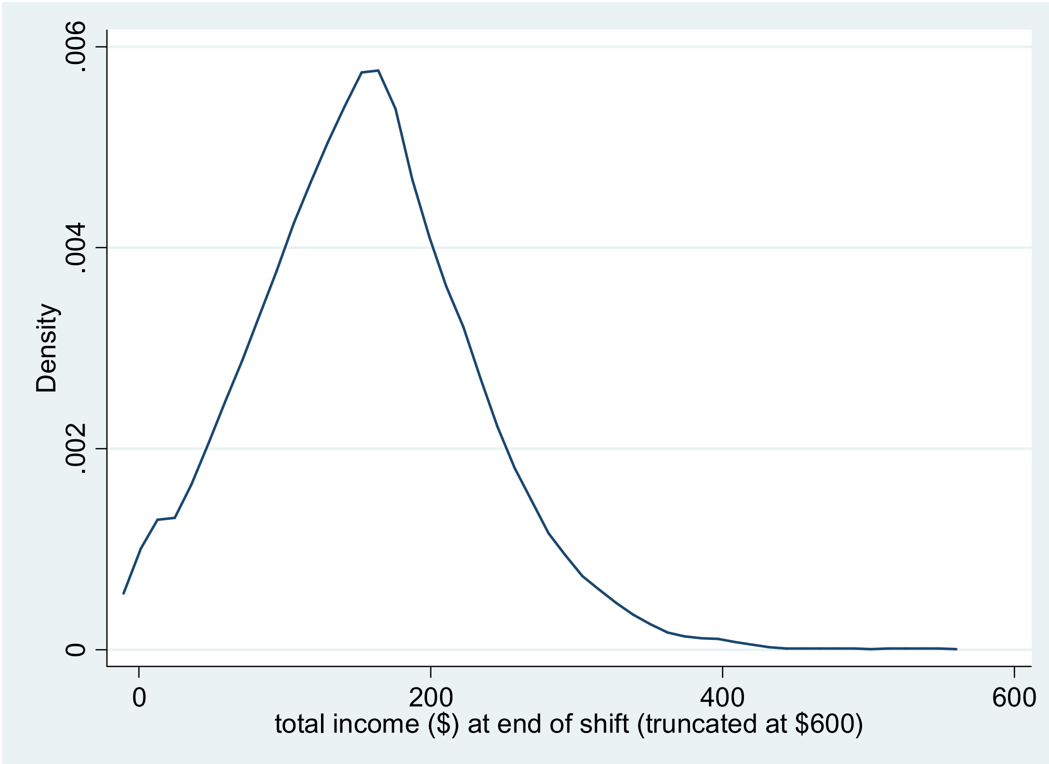
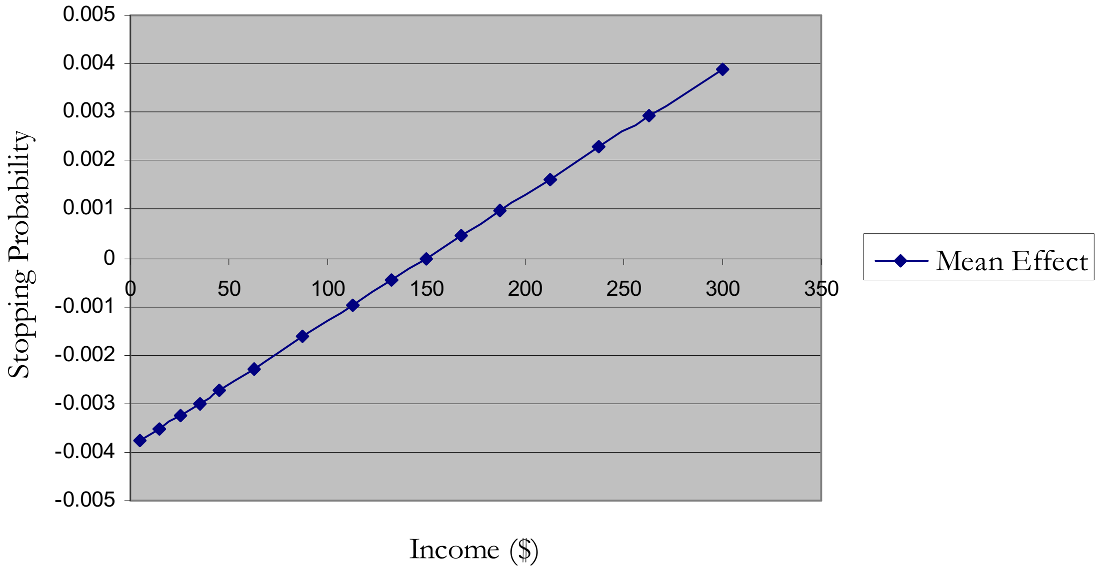


Figure 2.2: Density of Income Earned Per Shift



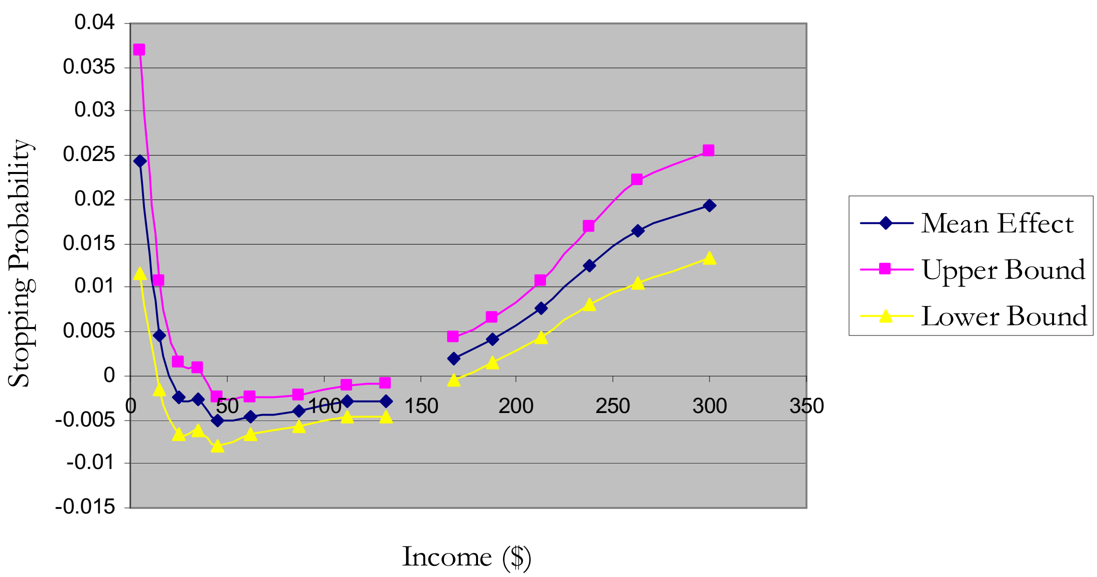
:

Figure 2.3a: Results of Probit Model of Stopping Probability as a Linear Function of Income



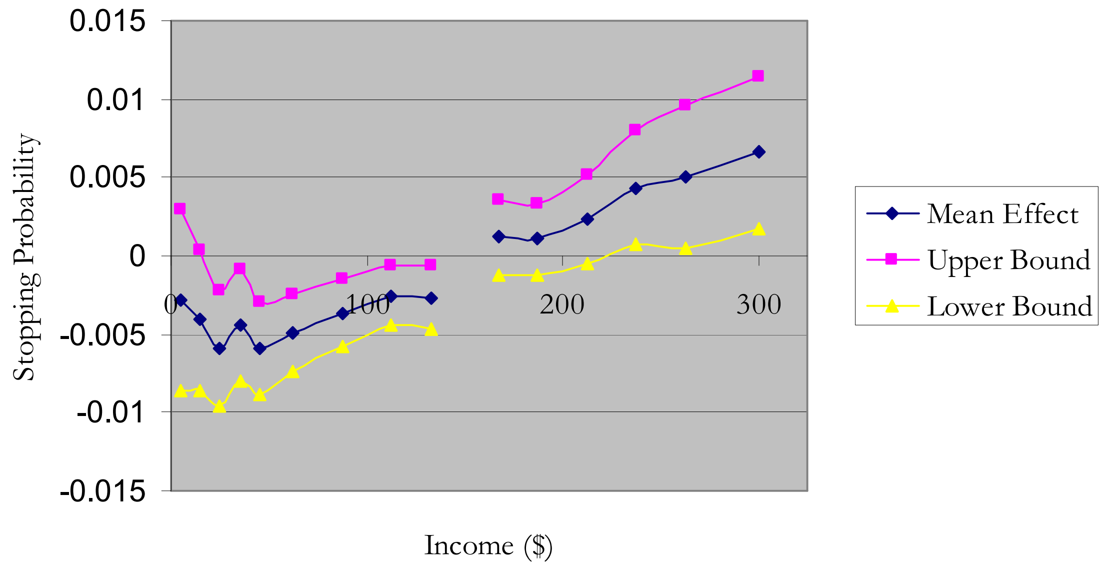
* The marginal effect is taken from Table 2, specification 1. I normalize the stopping probability to be 0 at $150 of income.

Figure 2.3b: Marginal Effects from Probit Model of Stopping Probability as a Function of Income



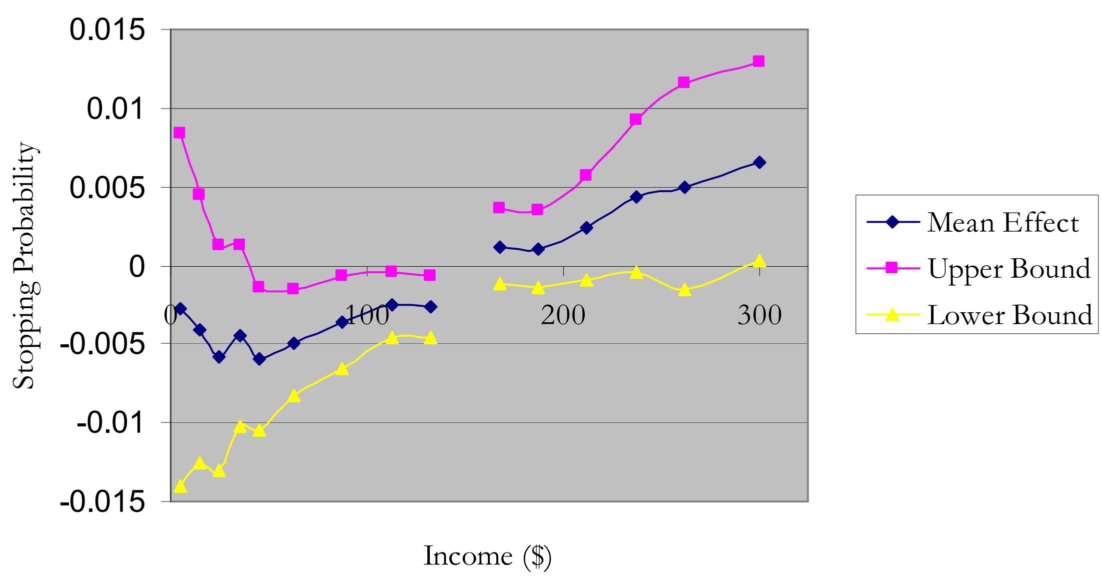
* The marginal effects are taken from Table 2, specification 2. The upper and lower bounds are calculated from unclustered standard errors.

Figure 2.3c: Marginal Effects from Probit Model of Stopping Probability as a Function of Income



* The marginal effects are taken from Table 2, specifications 3 and 4. The upper and lower bounds are calculated from the unclustered standard errors in specification 3.

Figure 2.3d: Marginal Effects from Probit Model of Stopping Probability as a Function of Income



* The marginal effects are taken from Table 2, specifications 3 and 4. The upper and lower bounds are calculated from standard errors clustered at the driver level in specification 4.

Figure 2.4: Density of Marginal Effects of Daily Income on Stopping Probability

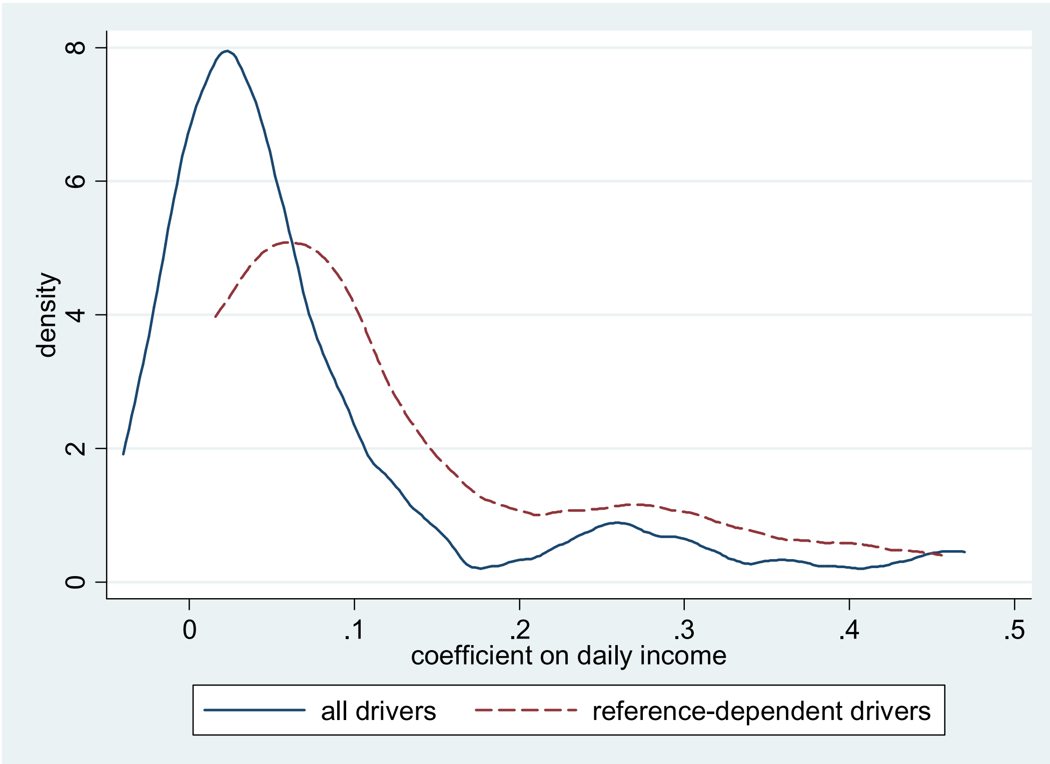


Figure 2.5: Density of Each Driver's Standard Deviation in Daily Income Across Shifts

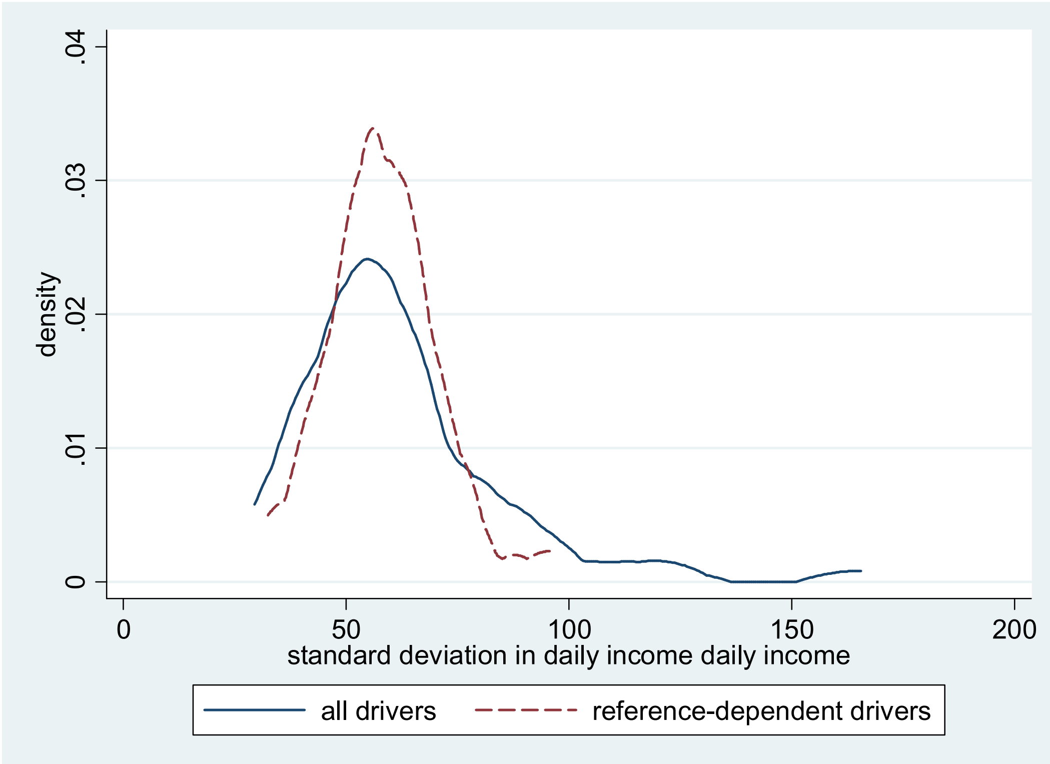


Figure 2.6 : Scatter Plot of Each Driver's Standard Deviaion in Daily Income Across Shifts vs. Each Driver's Marginal Effect of Daily Income on Stopping Probability

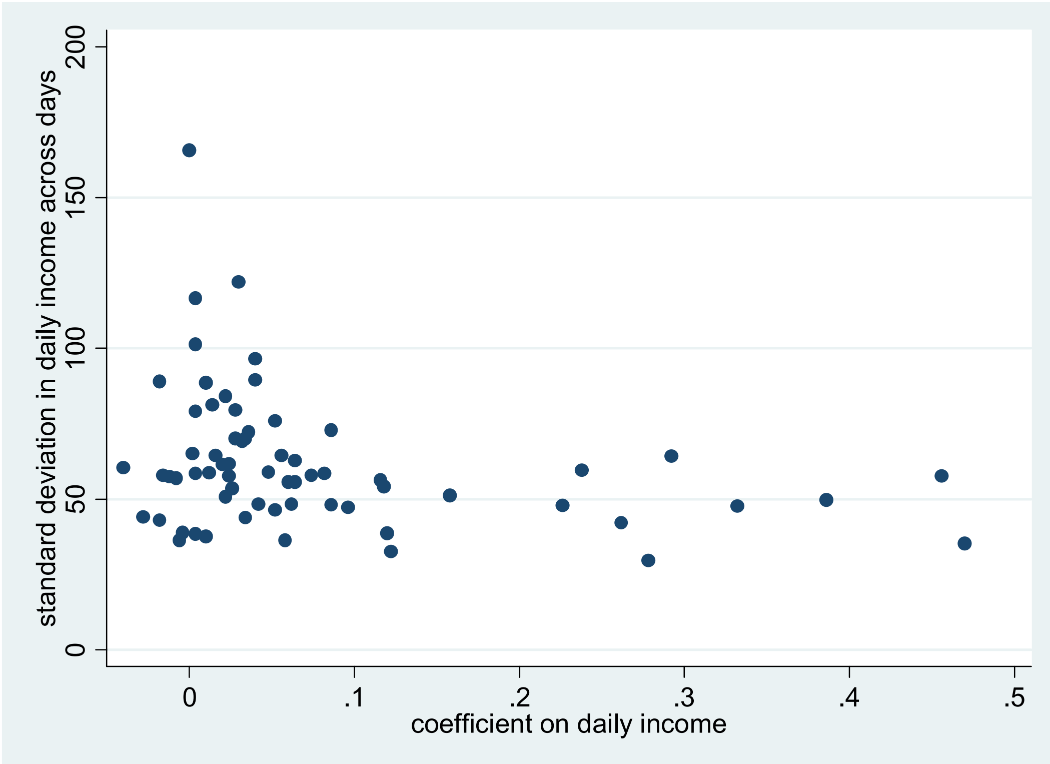


Table 3.1: Simple Statistics (std. deviation in parenthesis for continuous variables)

Variable	Mean in Pooled Sample	Mean among Men	Mean among Women
beauty = 1, very unattractive†	0.02	0.01	0.02
beauty = 2, unattractive	0.05	0.06	0.04
beauty = 3, about average	0.44	0.51	0.38
beauty = 4, attractive	0.34	0.31	0.36
beauty = 5, very attractive	0.15	0.10	0.20
Cum. Secondary School GPA	2.56 (0.84)	2.41 (0.84)	2.70 (0.81)
failure rate in Math Courses	0.12 (0.21)	0.14 (0.23)	0.10 (0.19)
failure rate in Science Courses	0.10 (0.21)	0.12 (0.23)	0.08 (0.19)
failure rate in All Courses	0.09 (0.16)	0.11 (0.17)	0.07 (0.14)
Highest math course †††	5.8 (2.2)	5.6 (2.3)	6.0 (2.1)
Total Math Credits in 4 yrs.	3.0 (1.0)	3.0 (1.1)	3.1 (1.0)
Highest science course †††	4.2 (1.5)	4.1 (1.5)	4.3 (1.4)
Total Science Credits in 4 yrs.	2.8 (1.0)	2.7 (1.0)	2.9 (1.0)
Self-Reported Academic Effort†	3.3 (0.7)	3.1 0.75)	3.4 (0.63)
Mother's* level of schooling**	4.9 (2.0)	4.9 (2.0)	4.8 (2.0)
family ann. income, in $1000s †	$45.73 ($52)	$45.69 ($51)	$45.77 ($52)
subjective health (in wave I)	4.9 (0.9)	5.0 (0.9)	4.8 (0.9)
height (inches) (in wave I)	66 in. (4 in.)	68 in. (4 in.)	64 in. (3 in.)
weight (pounds) (in wave I)	141 lb. (35 lb.)	152 lb. (37 lb.)	131 lb. (29 lb.)
Male (in wave I)	0.49	1.00	0.00
White (in wave I)	0.62	0.62	0.61
Black (in wave I)	0.23	0.22	0.24
Indian (in wave I)	0.04	0.03	0.04
Asian (in wave I)	0.08	0.08	0.07
relationship in last 18 months? †	0.56	0.54	0.57
# relationships last 18 months††	0.82 (0.93)	0.78 (0.94)	0.86 (0.91)
# months in all relationships †	7 months (28)	6 months (24)	8 months (30)
non-relationship sexual activity†	0.26	0.31	0.22
Drank more than 2/3 times? †	0.57	0.57	0.56
Usual # drinks per episode †	5.5 drinks (7.6)	6.3 drinks (8.3)	4.7 drinks (6.7)
Frequency of drunkenness †	2.4	2.5	2.2

*Sometimes, if the mother or female guardian refused to answer the survey, or did not reside with the student, a father or male guardian was interviewed instead

**1 = "8th grade or less," 2 = "more than 8th grade, but did not graduate from high school," 3 = "went to a business, trade, or vocational school instead of high school," 4 = "high school graduate," 5 = "completed a GED," 6 = "went to a business, trade or vocational school after high school," 7 = "went to college, but did not graduate," 8 = "graduated from a college or university," 9 = "professional training beyond a 4-year college or university"

† As measured in Wave I. For variable definitions, see the relevant regression tables below.

†† As measured in Wave II. For variable definitions, see the relevant regression tables below.

††† For variable definitions, see the relevant regression tables below.ACHIEVEMENT IN SCHOOL:

Table 3.2: Cumulative Secondary School GPA

Dependent Variable: Cumulative GPA across all years of secondary school, out of 4.0

OLS Regression

	(1) Pooled	(2) Pooled	(3) Men	(4) Men	(5) Women	(6) Women
Beauty = 1	0.02 (0.09)	-0.07 (0.12)	0.14 (0.14)	0.08 (0.16)	-0.09 (0.12)	-0.25 (0.17)
Beauty = 2	-0.09* (0.06)	0.02 (0.05)	-0.05 (0.07)	0.03 (0.07)	-0.15* (0.08)	-0.02 (0.08)
Beauty = 4	0.11*** (0.03)	0.00 (0.03)	0.15*** (0.04)	0.05 (0.04)	0.05 (0.04)	-0.06 (0.04)
Beauty = 5	0.09** (0.04)	-0.15*** (0.05)	0.23*** (0.06)	-0.01 (0.08)	-0.03 (0.05)	-0.27*** (0.07)
Personality = 1		-0.00 (0.11)		-0.06 (0.14)		0.05 (0.17)
Personality = 2		-0.09 (0.06)		-0.00 (0.07)		-0.24** (0.10)
Personality = 4		0.02 (0.03)		0.02 (0.05)		0.03 (0.03)
Personality = 5		0.14*** (0.05)		0.05 (0.07)		0.20*** (0.06)
Grooming = 1		0.12 (0.13)		0.13 (0.21)		0.20 (0.17)
Grooming = 2		-0.31*** (0.06)		-0.32*** (0.08)		-0.27** (0.12)
Grooming = 4		0.16*** (0.03)		0.14*** (0.04)		0.17*** (0.03)
Grooming = 5		0.23*** (0.04)		0.29*** (0.07)		0.20*** (0.05)
Ascriptives	No	Yes	No	Yes	No	Yes
R2	0.23	0.26	0.20	0.22	0.21	0.24
# obs	8350	8350	4035	4035	4315	4315

* = significant at the 10% level, ** = 5% level, *** = 1% level.

All specifications use weights to correct for the sampling structure, and include controls for: gender, race, age, body mass index, subjective health, school year, family income and parental schooling. The other ascriptives are: personality, grooming, candor and physical maturity (only the coefficients for personality and grooming are reported). All indepedent variables are observed in Wave I.

Table 3.3: Progression in Mathematics

Dependent Variable: Highest Math Course Completed with Credit by the end of Secondary School[&], Ordered Probit Regression, marginal effect and (P-value) reported

	(1) Pooled	(2) Pooled	(3) Men	(4) Men	(5) Women	(6) Women
Beauty = 1	0.013 (p=0.18)	0.009 (p=0.49)	0.038** (p=0.02)	0.047** (p=0.01)	0.009 (p=0.48)	-0.029* (p=0.09)
Beauty = 2	-0.013* (p=0.056)	0.001 (p=0.82)	-0.013 (p=0.15)	0.003 (p=0.70)	0.014 (p=0.12)	0.000 (p=0.98)
Beauty = 4	0.010*** (p=0.00)	-0.001 (p=0.79)	0.014*** (p=0.00)	0.004 (p=0.38)	0.006 (p=0.15)	-0.008* (p=0.09)
Beauty = 5	0.006 (p=0.12)	-0.015*** (p=0.01)	0.016** (p=0.01)	-0.006 (p=0.51)	0.003 (p=0.53)	-0.025*** (p=0.00)
Personality = 1		-0.002 (p=0.83)		-0.013 (p=0.37)		0.013 (p=0.42)
Personality = 2		-0.018*** (p=0.00)		-0.016** (p=0.04)		-0.024** (p=0.01)
Personality = 4		0.005 (p=0.12)		0.000 (p=0.95)		0.009** (p=0.04)
Personality = 5		0.012*** (p=0.01)		0.008 (p=0.30)		0.018*** (p=0.01)
Grooming = 1		-0.003 (p=0.85)		-0.023 (p=0.22)		0.021 (p=0.22)
Grooming = 2		-0.025*** (p=0.00)		-0.027** (p=0.01)		-0.027** (p=0.04)
Grooming = 4		0.016*** (p=0.00)		0.014*** (p=0.00)		0.019*** (p=0.00)
Grooming = 5		0.014*** (p=0.00)		0.016** (p=0.04)		0.015*** (p=0.01)
Ascriptives	No	Yes	No	Yes	No	Yes
R2	NA	NA	NA	NA	NA	NA
# obs	8340	8340	4033	4033	4307	4307

[&] 0 = no math,1 = basic/remedial math, 2 = general/applied math, 3 = pre-algebra, 4 = algebra I, 5 = geometry, 6 = algebra II, 7 = advanced math, 8 = pre-calculus, 9 = calculus

* = significant at the 10% level, ** = 5% level, *** = 1% level.

All specifications use weights to correct for the sampling structure, and include controls for: gender, race, age, body mass index, subjective health, school year, family income and parental schooling. The other ascriptives are: personality, grooming, candor and physical maturity (only the coefficients for personality and grooming are reported). All indepedent variables are observed in Wave I.

Table 3.4: Progression in Science

Dependent Variable: Highest Science Course Completed with Credit by end of Secondary School[&], Ordered Probit Regression, marginal effect and (P-value) reported

	(1) Pooled	(2) Pooled	(3) Men	(4) Men	(5) Women	(6) Women
Beauty = 1	0.012 (p=0.32)	-0.004 (p=0.83)	0.033 (p=0.12)	0.021 (p=0.34)	-0.010 (p=0.60)	-0.037 (p=0.13)
Beauty = 2	-0.004 (p=0.70)	0.013 (p=0.22)	0.002 (p=0.91)	0.018 (p=0.24)	-0.015 (p=0.029)	0.008 (p=0.54)
Beauty = 4	0.013*** (p=0.01)	-0.004 (p=0.37)	0.014** (p=0.04)	0.001 (p=0.84)	0.008 (p=0.16)	-0.013* (p=0.08)
Beauty = 5	0.004 (p=0.49)	-0.027*** (p=0.00)	0.021*** (p=0.01)	0.000 (p=0.98)	-0.013* (p=0.10)	-0.049*** (p=0.00)
Personality = 1		0.014 (p=0.50)		0.027 (p=0.38)		0.016 (p=0.52)
Personality = 2		-0.026*** (p=0.01)		-0.018 (p=0.14)		-0.044*** (p=0.00)
Personality = 4		0.006 (p=0.20)		-0.003 (p=0.59)		0.015** (p=0.04)
Personality = 5		0.019** (p=0.02)		0.007 (p=0.53)		0.029** (p=0.01)
Grooming = 1		0.025 (p=0.40)		0.012 (p=0.81)		0.052 (p=0.10)
Grooming = 2		-0.031** (p=0.03)		-0.027 (p=0.11)		-0.039* (p=0.05)
Grooming = 4		0.025*** (p=0.00)		0.025*** (p=0.00)		0.026*** (p=0.00)
Grooming = 5		0.029*** (p=0.00)		0.032*** (p=0.00)		0.027** (p=0.01)
Ascriptives	No	Yes	No	Yes	No	Yes
R2	NA	NA	NA	NA	NA	NA
# obs	8293	8293	4000	4000	4293	4293

* = significant at the 10% level, ** = 5% level, *** = 1% level.

[&] 0 = no science,1 = basic/remedial science, 2 = general/earth science, 3 = biology, 4 = chemistry, 5 = advanced science, 6 = physics

All specifications use weights to correct for the sampling structure, and include controls for: gender, race, age, body mass index, subjective health, school year, family income and parental schooling. The other ascriptives are: personality, grooming, candor and physical maturity (only the coefficients for personality and grooming are reported). All indepedent variables are observed in Wave I.

EFFECT OF BEAUTY ON EFFORT IN SCHOOL:

Table 3.5: Effort on School Work[&]
Question: "In general, how hard do you try to do your school work well?"*
OLS Regression

	(1) Pooled	(2) Pooled	(3) Men	(4) Men	(5) Women	(6) Women
Beauty = 1	-0.01 (0.08)	-0.04 (0.08)	-0.10 (0.13)	-0.14 (0.12)	0.05 (0.08)	0.03 (0.09)
Beauty = 2	0.04 (0.05)	0 .10** (0.05)	0.07 (0.06)	0 .14** (0.07)	0.00 (0.07)	0 .04 (0.07)
Beauty = 4	0.02 (0.03)	-0.03 (0 .02)	0.05 (0.03)	-0.00 (0 .04)	-0.01 (0.03)	-0.06* (0 .03)
Beauty = 5	0.01 (0.03)	-0.10*** (0 .03)	0.06 (0.04)	-0.06 (0 .05)	-0.02 (0.04)	-0.13*** (0 .04)
Personality = 1		0.07 (0.09)		0.15 (0.15)		-0.00 (0.09)
Personality = 2		-0.09* (0.05)		-0.10 (0.07)		-0.04 (0.09)
Personality = 4		0.06*** (0.02)		0.06* (0.04)		0.06** (0.03)
Personality = 5		0.09*** (0.03)		0.10* (0.05)		0.09** (0.04)
Grooming = 1		-0.20 (0.15)		-0.43* (0.24)		0.02 (0.17)
Grooming = 2		-0.17*** (0.06)		-0.18* (0.09)		-0.14 (0.11)
Grooming = 4		0.04* (0.02)		0.02 (0.03)		0.05* (0.03)
Grooming = 5		0.11*** (0.08)		0..05 (0.06)		0.15*** (0.04)
Ascriptives	No	Yes	No	Yes	No	Yes
R2	0.07	0.08	0.06	0.07	0.06	0.07
# obs	9964	9964	4891	4891	5073	5073

* = significant at the 10% level, ** = 5% level, *** = 1% level.
[&] -1 = "I try very hard to do my best", -2 = "I try hard enough, but not as hard as I could," -3 = "I don't try very hard," -4 = "I never try at all."
All specifications use weights to correct for the sampling structure, and include controls for: gender, race, age, height, weight, subjective health, school year, family income and parental schooling. The other ascriptives are: personality, grooming, candor and physical maturity (only the coefficients for personality and grooming are reported). All indepedent variables are observed in Wave I.

ROMANTIC RELATIONSHIPS and SEXUAL ACTIVITY:

Table 3.6: Likelihood of Romantic Relationships
Question: "Have you had a relationship with anyone in the last 18 months?"
OLS Regression

	Wave I Outcome	Wave I Outcome	Wave II Outcome	Wave II Outcome	Wave I Outcome	Wave II Outcome
Wave I Attributes						
Beauty	0.03*** (0.01)		0.04*** (0.01)		0.03*** (0.01)	0.04*** (0.01)
Beauty = 1		0.10** (0.05)		0.02 (0.06)		
Beauty = 2		-0.04 (0.04)		-0.05 (0.04)		
Beauty = 4		0.05*** (0.02)		0.04** (0.02)		
Beauty = 5		0.09*** (0.02)		0.10*** (0.02)		
Other ascriptive characteristics	YES	YES	YES	YES	NO	NO
R2	0.10	0.10	0.09	0.09	0.09	0.08
# obs	13188	13213	9827	9848	13199	9838

* = significant at the 10% level, ** = 5% level, *** = 1% level.
All specifications use weights to correct for the sampling structure, and include controls for: gender, race, age, height, weight, subjective health, school year, family income and parental schooling. The other ascriptives are: personality, grooming, candor, and physical maturity.

Table 3.7: Number of Romantic Relationships
Question: "How many relationships have you had in the last 18 months?"
OLS Regression

	Wave II Outcome	Wave II Outcome	Wave II Outcome
Wave I Attributes			
Beauty	0.06*** (0.02)		0.06*** (0.02)
Beauty1		0.16 (0.13)	
Beauty2		-0.06 (0.07)	
Beauty4		0.10*** (0.03)	
Beauty5		0.17*** (0.05)	
Other ascriptive characteristics	YES	YES	NO
R2	0.05	0.05	0.05
# obs	9827	9848	9838

* = significant at the 10% level, ** = 5% level, *** = 1% level.
All specifications use weights to correct for the sampling structure, and include controls for: gender, race, age, height, weight, subjective health, school year, family income and parental schooling. The other ascriptives are: personality, grooming, candor, and physical maturity.

Table 3.8: Length of time spent in Romantic Relationships

Dependent Variable: The sum of the # of months spent in each relationship, over all reported relationships

OLS Regression

	Wave I Outcome	Wave I Outcome	Wave II Outcome	Wave II Outcome	Wave I Outcome	Wave II Outcome
Wave I Attributes						
Beauty	0.60*** (0.19)		0.56** (0.20)		0.59*** (0.14)	0.03*** (0.47)
Beauty1		1.90 (1.20)		0.02 (0.91)		
Beauty2		-0.65 (0.63)		-0.82* (0.47)		
Beauty4		0.82*** (0.31)		0.75*** (0.29)		
Beauty5		1.83*** (0.42)		1.12** (0.48)		
Other ascriptive characteristics	YES	YES	YES	YES	NO	NO
R2	0.11	0.11	0.09	0.09	0.10	0.02
# obs	13026	13051	9812	9833	13037	9838

* = significant at the 10% level, ** = 5% level, *** = 1% level.

All specifications use weights to correct for the sampling structure, and include controls for: gender, race, age, height, weight, subjective health, school year, family income and parental schooling. The other ascriptives are: personality, grooming, candor, and physical maturity.

Table 3.9: Non-relationship Sexual Activity

Question: "Not counting the people you [may] have described as romantic relationships, [since Wave I] have you [ever] had a sexual relationship with anyone?"

OLS Regression

	Wave I Outcome	Wave I Outcome	Wave II Outcome	Wave II Outcome	Wave I Outcome	Wave II Outcome
Wave I Attributes						
Beauty	0.03*** (0.01)		0.01** (0.01)		0.01** (0.01)	0.00 (0.01)
Beauty = 1		0.01 (0.04)		-0.07** (0.03)		
Beauty = 2		-0.04 (0.03)		0.02 (0.03)		
Beauty = 4		0.02 (0.01)		0.01 (0.01)		
Beauty = 5		0.09*** (0.02)		(0.03)* (0.02)		
Other ascriptive characteristics	YES	YES	YES	YES	NO	NO
R2	0.12	0.12	0.06	0.06	0.11	0.05
# obs	13116	13140	9771	9792	13127	9782

* = significant at the 10% level, ** = 5% level, *** = 1% level.

All specifications use weights to correct for the sampling structure, and include controls for: gender, race, age, height, weight, subjective health, school year, family income and parental schooling. The other ascriptives are: personality, grooming, candor, and physical maturity.

DRINKING ACTIVITIES:

Table 3.10: Likelihood of Drinking Alcohol
Question: "[Since Wave I,] have you had a drink of beer, wine, or liquor—not just a sip or a taste of someone else's drink—more than two or three times [in your life]?"
OLS Regression

	Wave I Outcome	Wave I Outcome	Wave II Outcome	Wave II Outcome	Wave I Outcome	Wave II Outcome
Wave I Attributes						
Beauty	0.04*** (0.01)		0.03*** (0.01)		0.02** (0.01)	0.01 (0.01)
Beauty = 1		0.04 (0.06)		0.04 (0.06)		
Beauty = 2		-0.04 (0.03)		-0.03 (0.04)		
Beauty = 4		0.05*** (0.02)		0.05*** (0.02)		
Beauty = 5		0.11*** (0.03)		0.08*** (0.02)		
Other ascriptive characteristics	YES	YES	YES	YES	NO	NO
R2	0.10	0.10	0.06	0.06	0.09	0.05
# obs	13156	13180	9184	9120	13166	9185

* = significant at the 10% level, ** = 5% level, *** = 1% level.
All specifications use weights to correct for the sampling structure, and include controls for: gender, race, age, height, weight, subjective health, school year, family income and parental schooling. The other ascriptives are: personality, grooming, candor, and physical maturity.

Table 3.11: Number of Drinks per Drinking Episode
Question: "Think of all the times you have had a drink during the past 12 months. How many drinks did you usually have each time? A "drink" is a glass of wine, a can of beer, a wine cooler, a shot glass of liquor, or a mixed drink."
OLS Regression

	Wave I Outcome	Wave I Outcome	Wave II Outcome	Wave II Outcome	Wave I Outcome	Wave II Outcome
Wave I Attributes						
Beauty	0.5* (0.3)		0.46** (0.22)		0.20 (0.19)	0.45* (0.23)
Beauty = 1		1.88 (3.28)		-1.24 (1.06)		
Beauty = 2		-1.51** (0.64)		0.90 (1.19)		
Beauty = 4		0.69** (0.31)		0.43 (0.38)		
Beauty = 5		1.24** (0.56)		1.20** (0.51)		
Other ascriptive characteristics	YES	YES	YES	YES	NO	NO
R2	0.04	0.05	0.05	0.07	0.04	0.05
# obs	6183	6196	4279	4288	6189	4286

* = significant at the 10% level, ** = 5% level, *** = 1% level.
All specifications use weights to correct for the sampling structure, and include controls for: gender, race, age, height, weight, subjective health, school year, family income and parental schooling. The other ascriptives are: personality, grooming, candor, and physical maturity.

Table 3.12: Frequency of Drunkenness[&]

Question: "During Over the past 12 months, on how many days have you gotten drunk or "very, very high" on alcohol?"

OLS Regression

	Wave I Outcome	Wave I Outcome	Wave II Outcome	Wave II Outcome	Wave I Outcome	Wave II Outcome
Wave I Attributes						
Beauty	0.12*** (0.04)		0.20*** (0.04)		0.04 (0.03)	0.08** (0.04)
Beauty = 1		0.00 (0.26)		-0.37 (0.25)		
Beauty = 2		-0.08 (0.16)		-0.06 (0.19)		
Beauty = 4		0.20*** (0.06)		0.16 (0.08)		
Beauty = 5		0.26*** (.09)		0.45*** (0.12)		
Other ascriptive characteristics	YES	YES	YES	YES	NO	NO
R2	0.06	0.07	0.08	0.09	0.06	0.07
# obs	6285	6298	4361	4371	6291	4368

[&]-1 = "every day or almost every day," -2 = "3 to 5 days a week," -3 = "1 or 2 days a week, " -4 = "2 or 3 days a month," -5 = "once a month or less (3-12 times in the past 12 months)," -6 = "1 or 2 days in the past 12 months", -7 = "never"

* = significant at the 10% level, ** = 5% level, *** = 1% level.

All specifications use weights to correct for the sampling structure, and include controls for: gender, race, age, height, weight, subjective health, school year, family income and parental schooling. The other ascriptives are: personality, grooming, candor, and physical maturity.

Table 3.13: Tried Smoking
Question: "[Since Wave I,] have you [ever] tried cigarette smoking, even just one or two puffs?"
OLS Regression

	Wave I Outcome	Wave I Outcome	Wave II Outcome	Wave II Outcome	Wave I Outcome	Wave II Outcome
Wave I Attributes						
Beauty	0.03*** (0.01)		0.04*** (0.01)		0.01 (0.01)	0.02* (0.01)
Beauty = 1		0.03 (0.06)		0.03 (0.06)		
Beauty = 2		0.00 (0.03)		-0.01 (0.04)		
Beauty = 4		0.04** (0.02)		0.06*** (0.02)		
Beauty = 5		0.09*** (0.02)		0.11*** (0.03)		
Other ascriptive characteristics	YES	YES	YES	YES	NO	NO
R2	0.06	0.07	0.05	0.05	0.06	0.04
# obs	13160	13183	9798	9821	13170	9810

* = significant at the 10% level, ** = 5% level, *** = 1% level.
All specifications use weights to correct for the sampling structure, and include controls for: gender, race, age, height, weight, subjective health, school year, family income and parental schooling. The other ascriptives are: personality, grooming, candor, and physical maturity.

Appendix

A1. Construction of the no-treatment money sample and comparison control group

Each person in both treatment and control villages can be identified from the surveys as a member of one of three eligibility categories: (1) originally eligible; (2) eligible under the re-calculation of eligibility status in 1998; and (3) never eligible. The PROGRESA administrators assigned people who were materially well-off to category three, and people who were less-well off to category one. Everyone in both treatment and control villages is in one of these three groups, and the method of assignment should not have varied depending on whether one is in a treatment or control village. Therefore, people within a given eligibility category should be relatively similar across treatment vs. control villages. Figure 6 shows the breakdown by eligibility status of families living in treatment villages in 1997.

In order to find out which individuals in particular did not receive treatment money, I obtained administrative records identifying the recipient households and the timing for all payments made during the PROGRESA evaluation from the PROGRESA evaluation website, at http://evaloportunidades.insp.mx/en/index.php. I found that almost everyone living in treatment villages who was in eligibility category three never received money, but that in addition many of the presumably poorer people in eligibility category two also never received money (about 60 percent of them). According to Hoddinott, Skoufias and Washburn 2000, the PROGRESA administration claims that of the households that were eligible to receive benefits but never did receive any, 85.7 percent did not receive benefits because the administrators never incorporated them into the program. Thus, it seems that there is little room for selection in this sample of non-treated people living in the treatment group. In addition, because I am able to include people in eligibility category two, my sample of non-treated people in the treatment group includes households that are not restricted to be the richest in the villages[1].

I construct a similar comparison sample in the control group by including everyone in the control group who is in eligibility category 3 and a random sample of 60 percent of the people in eligibility category 2[2]. Since the households within a given eligibility group should be fairly similar by administrative design, and since the administrators should not have used different standards for eligibility status in the control and treatment villages, this technique creates a control group comparison sample that should be fairly similar to the treatment group non-treated sample. Table A1a shows baseline (1997) summary statistics for the two samples.

Table A1b shows the results of my hourly wage specification on these samples, with five percent symmetric cropping and controls and village fixed effects as before. These results demonstrate that by 1999 there was a significant wage increase even for the much smaller group of people living in treatment villages who did not receive treatment money. Table A1c shows the results of my quantity specifications on this sample – they suggest that the quantity of adult jornalero labor in this sample also increased.

Finally, as a robustness check I consider a further subsample of the above adult jornaleros who are perfectly healthy according to the following ten criteria: days of difficulty performing activities due to bad health in the past month are 0; days of missed activities due to bad health in the past month are 0; days in bed due to bad health in the past month are 0; yes, I

---

1 As a robustness check (to avoid potential problems with selection), I also consider only the richest people in each village: those in category three who were never eligible to receive treatment according to the criteria applied to both control and treatment villages. Performing kolmogorov smirnov tests on the wage distributions in 1997 and 1999 shows that before treatment I can reject their inequality, but after treatment I cannot reject that the control distribution is smaller. This holds for the overall sample, and for the healthy-only sample described at the end of this section. Thus, the results in this section seem to be robust to restricting the sample to only the never eligible, where there are fewer potential problems with selection.

2 The results are similar when my control sample includes all the people in eligibility category two (with weights of 0.6) and all in eligibility category three (with weights of 1.0).

can currently perform vigorous activities; yes, I can currently perform moderate activities; yes, I can carry an object of 10kg 500meters with ease; yes, I can easily lift a paper of the floor; yes, I can walk 2 km with ease; yes, I can dress myself with ease; I have had no physical pain in the last month. Without updating the cropping from the larger subsample above, I perform the same difference and difference regression on wages. The results show that point estimates of the treatment effect are essentially unchanged, and remain statistically significant.

Table A1a: Comparison of baseline characteristics of no-treatment sample in treatment villages with comparison sample in control villages.[3]

Year	Variable	Control Villages	Treatment Villages
1997	# families	4,276 families	5,530 families
	# people	15,874 people	24,453 people
	% male	50.9%	51.4%
	% child (< 17 years)	33.2%	34.5%
	% adult (17 to 59 years)	54.7%	54.0%
	% worked last week	48.5%	46.7%
	% worked as jornalero	16.8%	16.5%
	Mean jornalero wage	3.80 pesos / hour	3.72 pesos / hour
	Mean age	*29.3 years*	*30.0 years*
	% with high schooling	*18%*	*20%*
	% speaking a dialect	*19.3%*	*22.4%*
	% literate	*79.3%*	*78.2%*
	% married	*39%*	*41%*
	% separated	17%	19%
	% divorced	0.18%	0.16%
	% widowed	5.4%	5.6%

[3] Italicized entries are significantly different at the 5% level in t-tests without clustering. In this baseline survey, no variables are significantly different at the 5% level in tests with clustering at the village level.

Table A1b: Treatment Effect on log hourly wages and log daily income from 1997 to 1999 for no-treatment sample and comparison control sample

Dependent Variable: log hourly wages or log daily income for Adult (ages 17 to 59) Jornaleros

Explanatory Variables	(1) Log hourly wage	(2) Log hourly wage	(3) Log daily income	(4) Log daily income
Treated (post = 1 & treatment village = 1)	0.018** (0.009)	0.022** (0.009)	0.0120** (0.008)	0.020** (0.009)
Post-treatment Dummy	0.320*** (0.007)	0.318*** (0.007)	0.301*** (0.006)	0.300*** (0.009)
Male Dummy	0.022*** (0.009)	0.020** (0.010)	0.061*** (0.009)	0.060*** (0.009)
Age	-0.000 (0.000)		-0.000** (0.000)	
Age Dummies		YES		YES
Schooling Level Dummies		YES		YES
Language Skills Dummies		YES		YES
Marriage Status Dummies		YES		YES
Village Fixed Effects	YES	YES	YES	YES
Constant	1.14*** (0.01)	1.14*** (0.02)	3.20*** (0.01)	3.17*** (0.02)
# Observations	8944	8647	8977	8653
R2	0.31	0.31	0.30	0.30

Standard Errors in Parenthesis
** = significant at 5% level
*** = significant at 1% level

Table A1c: Treatment Effect on log hours per week and log days per week from 1997 to 1999 for no-treatment sample and comparison control sample

Dependent Variable: log hours per week or days per week for Adult (ages 17 to 59) Jornaleros				
Explanatory Variables	(1) Hours per week	(2) Hours per week	(3) Days per week	(4) Days per week
Treated (post = 1 & treatment village = 1)	0.032** (0.015)	0.036** (0.016)	0.028** (0.014)	0.035** (0.014)
Post-treatment Dummy	-0.085*** (0.012)	-0.087*** (0.012)	-0.065*** (0.011)	-0.068*** (0.011)
Male Dummy	0.112*** (0.016)	0.117*** (0.016)	0.079*** (0.015)	0.081*** (0.015)
Age	-0.001*** (0.000)		-0.001** (0.000)	
Age Dummies		YES		YES
Schooling Level Dummies		YES		YES
Language Skills Dummies		YES		YES
Marriage Status Dummies		YES		YES
Village Fixed Effects	YES	YES	YES	YES
Constant	3.64*** (0.019)	3.59*** (0.034)	1.58*** (0.017)	1.56*** (0.031)
# Observations	8997	8698	9019	8716
R2	0.01	0.02	0.01	0.01

Standard Errors in Parenthesis
** = significant at 5% level
*** = significant at 1% level

A2. Which deciles of the distribution of wages are moving?

Table A2 shows quantile hourly wage regressions by decile for adult jornaleros with no controls and no cropping. It is apparent that there is significant evidence of a wage increase by 1999.

Table A2. Quantile Difference-in-Difference Treatment effects on Hourly Wages, no controls or cropping

	1997 vs. 1999
10th Percentile	0.000 (0.017)
20th Percentile	**0.131** **(0.023)**
30th Percentile	**0.179** **(0.008)**
40th Percentile	0.000 (0.050)
50th Percentile	-0.083 (0.076)
60th Percentile	**0.069** **(0.032)**
70th Percentile	0.000 (0.056)
80th Percentile	**0.625** **(0.061)**
90th Percentile	-0.020 (0.270)

Standard Errors are in parenthesis. Results significant at the 5% level are bolded

A.3 Data Collection

The data collection process for the taxi driver data occurred in four stages. First, I used a list of taxi and limousine companies registered with the city of New York (for the current resources, see: http://www.nyc.gov/html/tlc/html/current/current_licensees.shtml). I called each of the companies on this list, and/or visited them in person. Second, I identified two anonymous companies that were willing to let me use their trip sheets. I determined that one of these companies had organized their trip sheets in such a way that I could easily gather the trip sheets of individual drivers over the course of a specific period of time. Since I was looking for a panel data set of drivers' shifts around the exogenous wage change of May 2004, this was the company from which I gathered most of the data (in particular, all of the data that I use in this study is from this company). The third stage of the data gathering process involved scanning the trip sheets themselves. This third stage itself included two steps.

First, I looked in the trip sheet storage room for boxes of trip sheets that were from the first half of 2004, so that the panel would overlap May 2004. Inside each box of trip sheets, the trip sheets associated with each medallion were held together by rubber bands. According to two characteristics, many of the packets of data in each box were unusable. First of all, the number of days of the year that each box covered implied that two drivers renting a medallion should have a large number of trip sheets included – many of the packets contained very few sheets. Secondly, many drivers were clearly not filling out more than a few trips per sheet. Given the expense of encoding the data, we decided to avoid encoding useless packets, and to only scan in packets that had enough trip sheets and enough trips per trip sheet to be reasonable. Thus, the second step of the third stage involved looking for these packets in each box, and scanning their contents.

The fourth and final stage of the data gathering process was to encode the data from the trip sheets that we had scanned in. We hired research assistants to type in the data from each trip. Once the data was typed in, my research assistants and I performed the data cleaning.