ABSTRACT

Title of dissertation: ESSAYS IN LABOR ECONOMICS

Ye Zhang, Doctor of Philosophy, 2007

Dissertation directed by: Professor Jeffrey Smith, Economics
Professor Judith Hellerstein, Department of Economics

My dissertation is composed of two essays in Labor Economics. The first chapter examines how employers learn about workers' unobserved productivity when learning is asymmetric between incumbent and outside firms. I develop an asymmetric employer learning model in which endogenous job mobility is both a direct result of intensified adverse selection and a signal used by outside employers to update their expectations about workers' productive ability. I derive, from the model, empirical implications regarding the relationship between wage rates, ability, schooling and overall measures of job separations that contrasts the public learning models and the two-period mover-stayer models. Testing the model with data from the National Longitudinal Survey of Youth 1979 (NLSY-79), I find strong evidence supporting the three-period asymmetric employer learning model.

The second chapter concerns economics of fertility and investigates to what extent the observed correlation between adolescent fertility and poor maternal educational attainment is causal. Semi-parametric kernel matching estimator is applied to estimate the effects of teenage childbearing on schooling outcomes. The matching method estimates the conditional moments without imposing any functional form

restrictions and attends directly to the common support condition. Using data from the NLSY-79, kernel matching estimates suggest that half of the cross-sectional educational gaps remains after controlling for individual and family covariates. The difference between matching estimates and regression-based estimates implies that part of the conditional difference in parametric models is due to the functional assumption. The robustness check following Altonji, Elder, and Taber (2005) reveals that a substantial amount of correlation is required within a parametric framework to make the negative effect of teen motherhood on educational attainment go away. Further evidence obtained by simulation-based nonparametric sensitivity analysis suggests that the matching estimates are quite robust with regard to a wide range of specifications of the simulated unobservables.

ESSAYS IN LABOR ECONOMICS

by

Ye Zhang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:
Professor Jeffrey Smith, Co-Chair
Professor Judith Hellerstein, Co-Chair
Professor Seth Sanders
Professor William Evans
Professor Anna Alberini

Preface

The development of the economics of human capital and the application of traditional price theoretic models to non-market decision making in the last fifty years has provided us a framework for the understanding of many aspects of observed behavior in the labor market regarding schooling, occupation choice, job mobility, etc.. Human capital, by definition, encompasses all the individual characteristics that make people productive. This dissertation investigates how market uncertainty about these productive skills affects wage determination and mobility patterns and how fertility timing impacts the acquisition of human capital.

Understanding the information structure of how employers learn about workers' productivity is of crucial importance in our understanding of within-firm job assignment, the wage structure and market discrimination. The simplifications of the employer learning process in the previous literature either equate the information sets of incumbent and outside firms or completely rule out learning by the recruiting employers. The model constructed in the first chapter of this thesis treats workers' endogenous job mobility a signal used by the outside market participants to assess the workers' productive ability. This more realistic learning hypothesis generates predictions that are different than both exiting public learning models and two-period mover-stayer models for the relationship between mobility and ability, between job mobility history and the effect of ability on the wage, and between the effect of schooling on wages and the extent of job mobility. I test all of these implications using micro data from the National Longitudinal Survey of Youth 1979

(NLSY-79). The empirical evidence broadly supports my three-period asymmetric employer learning model with differential learning from incumbent and outside employers.

Delayed motherhood is correlated with higher female socio-economic outcomes in both cross-sectional and time-series comparisons, and human capital accumulation is believed by many social scientists to be one of the important channels that generates this association. The second chapter of the dissertation examines that to what extent the observed correlation between teenage childbearing and poor maternal educational attainment is causal. In this co-authored chapter with Seth Sanders and Jeffrey Smith, we invoke a selection-on-observables identifying assumption and use semi-parametric kernel matching estimator to explore this question in a sample of NLSY-79. Our estimates suggest that adolescent fertility is detrimental to the schooling outcomes of teen mothers, especially for high school completion. The difference between matching estimates and regression-based estimates calls into question the findings of parametric models: they may be biased due to the functional form assumption. We also conduct sensitivity analyses to check the robustness of our matching estimates against the failure of the identifying assumption. Following Altonji, Elder, and Taber (2005), a bivariate probit model reveals that a substantial amount of correlation is required to wipe out the negative effect of teen motherhood on educational attainment. Following Ichino, Mealli, and Nannicini (2007), the results obtained by the simulation-based nonparametric method suggests that our matching estimates are quite robust to a number of configurations regarding the simulated unobservables.

Dedication

To the most beautiful place in the world, the city of Beijing, for making me the person I am.

Acknowledgments

One thing that my college teacher told me when I was getting ready for graduate school in the United States, I still remember to this day, is the importance of thesis adviser. Now that I look back, I think I am extremely lucky, because I have two of the best.

Countless hours of discussions with Jeff Smith, before and after he left Maryland, completely shaped my way of thinking economics problems. He is a mentor, a friend, and an example I will always look up to. Over the past six years, he has been a reliable source of knowledge, constructive critique and encouragement. He rescued me over and over again from unrealistic "never never land" and taught me to work from the very simplest form. His help and support is far beyond this dissertation.

Judy Hellerstein's 771 in the fall of 2001 is the reason I want to be a labor economist. She taught me how to run my first Mincer regression and helped me in limitless ways as I converged to my job market paper idea. Her ability to connect a theoretical idea with a concrete econometric model will always be an inspiration to me. I will surely miss my enjoyable coffee sessions with Judy when I move to Indianapolis. I do wish one day, I can "show her the money".

The second chapter of this dissertation is a joint work with Jeff Smith and Seth Sanders. Taking Labor Economics from Seth got me interested in the economics of fertility and family structure. I am especially grateful to him for jump starting my paper-writing process.

At various stages of writing my thesis, I have also benefited from comments by Emek Basker, Pedro Carneiro, Maureen Cropper, Flavio Cunha, Bill Evans, Jonah Gelbach, Ted Joyce, Sandy Korenman, Peter Mueser, Salvador Navarro, June O'Neill, John Rust, and participants at several workshops where I have presented versions of the work contained here.

I own very special thanks to Bill Evans for his help and advice on the job market, and I thank Anna Alberini for agreeing to serve as the Dean's representative on my committee.

I thank my family and friends for putting up with me these last seven years and for helping me to stay sane throughout.

Table of Contents

List of Tables

List of Figures

Chapter 1

Employer Learning under Asymmetric Information: Theory and Evidence

1.1 Introduction

Asymmetric and imperfect information characterize almost every aspect of the modern labor market, and economists have been interested in investigating their consequences ever since the seminal work of Akerlof (1970) and Spence (1973). This paper studies an employer's private information about a worker's productivity and argues, theoretically and empirically, that early-career job mobility plays an important role in the employer learning process about employees' productive ability.

In a world where information about workers' productivity is incomplete, it is not possible for a company that is hiring to assess the value of a job candidate's unobserved innate ability. Instead, the potential worker's employment history and other forms of information about his productivity, such as resumes and reference letters, usually serve as the basis for recruitment. This information imperfection directly motivates the statistical theory of discrimination[1] where firms distinguish between individuals with different observable characteristics based on statistical regularities. Although some information about the worker's ability is available to all

[1]See, among others, Aigner and Cain (1977), Lundberg and Startz (1983), Lang (1986), Coate and Loury (1993), and Oettinger (1996).

the firms in the market, it is reasonable to imagine that the incumbent employer accumulates further information about the worker's productive ability after production begins, and then the employer updates its beliefs accordingly. The employer's subsequent wage offers and layoff/firing choices are conditioned on the revised expectations of the worker's productivity. When the current employer and potential employers set their wage rates according to different information sets, the worker's job mobility is endogenously determined by the wage offers from the two sides, and his employment history conveys information regarding his unobserved productivity. The job change pattern of the worker, which is an inevitable consequence of the information asymmetry, provides outside employers with an additional tool to go somewhat beyond the "veil of ignorance" and learn about the worker's productive ability. As intuitively appealing as it sounds, previous research on this topic has neglected the learning process of outside employers through the worker's employment history.

The main contribution of this chapter, and the key feature of my employer learning model, is to treat endogenous job mobility[2] as an additional source of information about a worker's productivity that is available to the outside employers.[3] In the context of asymmetric information, job changing as an outcome of market adverse selection can be used by potential employers to assess the quality of the worker. By offering workers with different mobility histories different wage rates,

[2]The model endogenizes job mobility by adding non-pecuniary job characteristics to the worker's utility function. For a similar approach to modeling mobility, see Neal (1998), Acemoglu and Pischke (1998), and Schonberg (2005).

[3]Gibbons and Katz (2001) allow outside firms to learn the reasons for prior job separations and condition their wage offers on them, but in reality, discerning the cause of a prior job change is much more challenging than obtaining the employment history of a job applicant.

market selection intensifies over time. In contrast, under the hypothesis that learning is symmetric between incumbent and outside employers,[4] job separations have no implications for the worker's expected productivity and mobility plays no role in employer learning. While earlier research on asymmetric information in the labor market recognizes that one consequence of private learning is that workers who switch firms are of lower quality than workers who stay with their employers, it relies on two-period mover-stayer type models and ignores the informational content of job changes. These job changes help outside employers to dynamically acquire extra information about worker productivity.

Using the 1979 cohort of the National Longitudinal Survey of Youth (NLSY-79), and taking advantage of its unique cognitive ability measure, the Armed Forces Qualification Test ($AFQT$) score, which provides a summary of basic math and literacy skills that is not observed by the employers, I test the intensified adverse selection model by examining whether the relationship between a worker's $AFQT$ score and job mobility weakens as workers age. If employer learning is symmetric, the average quality of workers who change jobs will be equal to that of workers who do not. Additionally, if adverse selection does not worsen with the accumulation of labor market experience as implied by the two-period mover-stayer asymmetric employer learning model, then the correlation between the ability measure and the probability of changing jobs should stay constant over time. In contrast, a model in which job mobility serves as an ability signal to outside employers not only implies

[4] The symmetric learning models of Farber and Gibbons (1996) and Altonji and Pierret (2001) do not consider worker mobility at all.

that the more frequent are job turnovers the lower is the quality of the worker, but it also predicts that unobserved ability plays less and less of a vital role in the mobility decision with each year that the worker spends in the labor market. Thus, this implication empirically differentiates the three models.

I modify the empirical model of Farber and Gibbons (1996) and Altonji and Pierret (2001) by incorporating into the wage regressions they specify the frequency of prior job separations.[5] I show that there is a difference between frequent job movers and occasional movers in terms of how ability affects the way that the incumbent and outside firms set their wage rates conditional on labor market experience. This finding is at odds with the public employer learning model because, in the symmetric learning model, the assumption that the incumbent and recruiting firms have the same amount of information about the worker's ability implies that the *AFQT* score affects every firm's wage offer in the same fashion given experience. This finding is also not consistent with the two-period mover-stayer model of private learning. If outside employers do not exploit the job mobility history as an additional source of information to distinguish low quality from high quality job candidates, the difference in the impacts of ability on wage rates between the incumbent and outside firms is identical for workers with different mobility levels given the same experience level. However, if the outside firms' wage offers depend on the employment history as described by the three-period model constructed in this chapter, the outside employers will have a more accurate assessment about

[5]Mincer and Jovanovic (1981) use the frequency of prior moves as a control for individual heterogeneity when estimating the returns to job seniority. I use the coefficients on the interaction terms between prior mobility, test score, job tenure, and years of schooling to test the three-period asymmetric employer learning model.

the productivity of workers with more job changes, so the employer learning more closely resembles public learning for this group of individuals. According to my three-period model, both current employers and outside employers learn, although through different channels. The incumbent employer updates its expectations of the worker's productivity by observing the worker's output and, over time, relies less and less on easily observable characteristics. The outside firms learn over time about productivity through observing the job mobility history of the worker, and these outside employers also depend less and less on variables like years of schooling. The substitution of employment history for schooling as a productivity signal implied by my model allows me to test the model by examining the impact of education on wages for individuals with different job turnover patterns.

The first chapter of the dissertation unfolds as follows. Section 1.2 provides a review of the employer learning literature. Section 1.3 presents my employer learning model where a worker's employment history is used by outside firms to revise their expectations about the worker's productivity, and contrasts the empirical implications of my model with those of the public learning and two-period mover-stayer models. Section 1.4 describes the data and Section 1.5 presents empirical evidence. Section 1.6 concludes.

1.2 Previous Literature on Employer Learning

While it seems plausible that prospective employers may be less informed about the productivity of the worker than the current employer, it is assumptions about how outside firms learn that divide the literature on employer learning. The phrases "symmetric employer learning" or "public learning" refer to the body of research that assumes away asymmetric information and instead assumes that all market participants, incumbent or outside, have the same amount of information about the worker's productivity at each point in time and that the labor market operates competitively. Examples of early theoretical analyses under the hypothesis of public employer learning are Freeman (1977) and Harris and Holmstrom (1982). Another set of studies, including this paper, assume that there is some degree of information asymmetry and that the incumbent employer has more information than other firms about the employee's ability. Under this assumption, recruiting firms have an informational disadvantage relative to current employers. How the outside firms use the information contained in the worker's employment history to minimize this disadvantage motivates this paper. In the literature, efforts have been made to examine how "asymmetric employer learning", or "private learning", might generate inefficient job assignments within the firm; these include the models laid out by Waldman (1984), Milgrom and Oster (1987), and Bernhart (1995). Other theories, such as those of Greenwald (1986) and Lazear (1986), focus on the analogous implications for wage dynamics and job separations.

Two influential papers made empirical breakthroughs in testing the employer learning model: Farber and Gibbons (1996) and Altonji and Pierret (2001). Working under the hypothesis of pure symmetric employer learning, they deliver testable empirical implications that are consistent with the observed patterns in the data for experience gradients, education, and test scores in a wage regression that are hard to reconcile with a simple human capital model. Their models predict that, at labor market entry, firms rely on easy-to-observe variables that are correlated with productivity to determine wage rates. Thus, the coefficient on a variable correlated with productivity which is not observable to employers but is observed by economic analysts should increase with labor market experience. The same argument leads to the decreasing time path of the coefficient on the easy-to-observe variable that is correlated with ability if the hard-to-observe measure of ability is included in the wage regression.[6] Both papers use the NLSY-79 to test their theoretical predictions and obtain broadly supportive results. Their methodology also has been applied to datasets outside of the United States. For example, Bauer and Haisken-DeNew (2001) find some support for the symmetric employer learning model in German data for blue-collar workers, but not for white-collar workers; Galindo-Rueda (2002) obtains similar findings using data from the UK for approximately the same time period as that considered by Altonji and Pierret (2001). More recently, Lange (2005) develops an econometric model to estimate the speed of employer learning,[7] also under the pure symmetric learning assumption. He finds that employers are able to reduce their average expectation error about the productivity of a worker by 50% over the first three years and he concludes that this is rather fast. It is noteworthy that if the current employer and outside employers hold different perceptions about a worker's productivity, then his conclusions may change.

[6] Altonji and Pierret (2001) specify their learning model in logarithms while Farber and Gibbons (1996) specify the model in levels and derive that wages should follow a martingale.

[7] In an earlier paper, Altonji and Pierret (1998) recognize that the speed of employer learning plays a crucial role in statistical discrimination. They argue that the observed coefficient patterns in their earnings equation are consistent with a fast speed of employer learning and that this limits the contribution of signaling to the returns to education.

Empirical research on labor market asymmetric information is sparse and far from conclusive. Gibbons and Katz (2001) test the asymmetric learning hypothesis by comparing the earnings losses of workers who are laid off versus those who are displaced for exogenous reasons, like a plant closing. Under the assumption that information concerning a worker's ability is private to the current employer, outside market participants infer that laid-off workers are of low quality and label them as "lemons", but no such inference is warranted for exogenous job leavers. Since pre-displacement wages do not differ by cause of displacement for the two groups of workers, their asymmetric learning model predicts a greater wage loss for layoffs than for those displaced by plant closing. Their empirical examination using the CPS Displaced Workers Supplements (DWS) clearly supports their model predictions.[8]

Rodriguez-Planas (2004) extends the adverse selection model of Gibbons and Katz (2001) by allowing recalls of laid-off workers to their original employers and offers a new test of the importance of asymmetric information in the labor market. She argues that if employers have discretion over whom to recall, high-ability workers are more likely to be recalled and may choose to remain unemployed rather than to accept a low-wage job offered early in their unemployment spell. If so, unemployment can serve as a signal of productivity. In this case, her model suggests that unemployment duration may be positively related to post-displacement wages even among workers who are not recalled. In contrast, because workers displaced through plant closings cannot be recalled, a longer duration of unemployment should

[8] Hu and Taber (2005) recently challenged the results of Gibbons and Katz (2001) by showing the difference in wage loss between exogenous job leavers and layoffs varies dramatically by race and gender. They offer heterogeneous human capital and taste-based discrimination as possible explanations for the observed patterns for African Americans and females.

not have a positive signaling benefit for such workers. Her empirical results using the 1988-2000 DWS reveal that the earnings and unemployment duration experiences of the two groups behave in the predicted way and are consistent with asymmetric information in the labor market.

1.3 The Asymmetric Employer Learning Model

1.3.1 A Basic Two-Period Model

First, let us consider a simple two-period employer learning model set up in the spirit of Greenwald (1986) and Schonberg (2005) to highlight the way in which asymmetric information and adverse selection distort market transactions. I extend the model to a three-period setting in Subsection 1.3.2. This model assumes the productivity of individual i in firm j, $\chi_{i,j}$, is given by $\chi_{i,j} = \eta_i + \delta_{i,j}$, where η_i denotes the ith worker's time-invariant innate ability and $\delta_{i,j}$ is the quality of the worker-firm match. The population distributions of η_i and $\delta_{i,j}$ are independent and are common knowledge to all market participants. I further assume that $\eta_i \sim N(\mu_\eta, \sigma_\eta^2)$, $\forall i$ and $\delta_{i,j} \sim N(\mu_\delta, \sigma_\delta^2)$, $\forall i, j$. Jobs are treated as pure search goods in this model[9] and match productivity is known ex ante. In another words, there is no further information on match quality generated in the model as the match proceeds. Following the job matching literature, a new value of $\delta_{i,j}$ is drawn from its distribution with each job change and the successive drawings are independent. This guarantees that the worker's prior employment history is not relevant in assessing

[9] For "pure-search-good" models of job changes, see, among others, Lucas and Prescott (1974), Burdett (1978), Mortensen (1978), Jovanovic (1979b), and Wilde (1979).

his $\delta_{i,j}$ in a newly formed match.[10]

The risk-neutral workers are also heterogeneous with regard to a non-pecuniary utility component, $\theta^j_{i,t}$, associated with job j for time period t.[11] The inclusion of this taste parameter is in line with most of the existing work on asymmetric employer learning and is part of an easy way to endogenize mobility. As explained by Greenwald (1986), the "random" quit behavior generated by this type of heterogeneity is critical to the existence of equilibrium turnover. In particular, it facilitates trading even in the presence of adverse selection so that the market does not break down completely as in Akerlof (1970). In this model, the non-pecuniary utility measure is assumed to be transitory and workers draw a new value of $\theta^j_{i,t}$ in each period for each job. This taste shock may refer to preferences to specific colleagues and supervisors, the working environment, health and other benefit programs, *etc.* I specify the distribution of $\theta^j_{i,t}$ as $N(0, \sigma^2_\theta)$ for any i, j, t.

Wage rates are determined on the spot market and long-term contracts of any sort are assumed away. At the beginning of the first period, wages are offered simultaneously by all of the recruiting employers. Firms do not see $\chi_{i,j}$ although they know $\delta_{i,j}$ upon inspection. In addition, after production takes place, the ith worker's output for period 1 in firm j, $y_{i,j,1}$, becomes known to the incumbent firm.

[10] Another line of job search and matching models treats match-specific productivity as an experience good; see, e.g., Johnson (1978), Jovanovic (1979a), and Moscarini (2003), where match quality is not known ex ante but is learned over time as the job is "experienced". In order to concentrate attention on employer learning and sequential adverse selection, and to avoid the complications caused by employee's time varying perceptions of job quality, I model match quality as an inspection good in this chapter.

[11] I use employer and job interchangeably in this paper. Empirically, the term "job" refers to any position within a given employer rather than to a particular position with that employer. The work history file in NLSY-79 does not provide enough information to distinguish job position changes from employer changes.

The public learning models of Farber and Gibbons (1996) and Altonji and Pierret (2001) assume that the information held by employers is symmetric and all of the firms in the market observe the same sequence of output $(y_{i,j,1}, y_{i,j,2}, ..., y_{i,j,t})$ through period t. In contrast, in my model, the productivity signal is only observed by the worker's current employer. This noisy measurement, $y_{i,j,1} = \chi_{i,j} + \epsilon_{i,j,1}$, is then used by the current firm to update its expectation of the ith individual's productivity. With an additional assumption of an $i.i.d$ normal distribution for $\epsilon_{i,j,t}$, Bayes's rule yields the expected productivity at the end of period one from the perspective of the incumbent employer:

$$E(\chi_{i,j} \mid y_{i,j,1}) = \frac{\sigma_\epsilon^2}{\sigma_\eta^2 + \sigma_\epsilon^2}(\mu_\eta + \delta_{i,j}) + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\epsilon^2} y_{i,j,1}. \tag{1.1}$$

The posterior mean is simply a weighted average of the prior expectation of the worker's productivity and the noise-ridden signal, where the weights depend on the relative sizes of the prior variance and the variance of the noise term $\epsilon_{i,j,1}$. The posterior variance $Var(\chi_{i,j} \mid y_{i,j,1})$ is known to be $\frac{\sigma_\eta^2 \sigma_\epsilon^2}{\sigma_\eta^2 + \sigma_\epsilon^2}$, which is independent of the realization of $y_{i,j,1}$.

At the beginning of period two, potential employers first make wage offers. The current employer then observes those wage offers and makes a counter offer. This timing of events in wage determination is standard in the literature dealing with asymmetric information.[12] While the key empirical implications of the model remain valid if the second-period wage offers are made simultaneously by the incumbent

[12]See, among others, Waldman (1984), Greenwald (1986), and Gibbons and Katz (1991).

and outside firms, they are no longer attainable if the current employer makes the first move. In this case, the incumbent firm loses its informational advantage and reveals the productivity of its workers to the entire market by tying wage offers to the productivity signals that only it observes. To avoid a host of game-theoretic strategic considerations that lie beyond the scope of this paper, I maintain the conventional assumption on the timing of wage offers. Observing the wage offers and the new draws of the non-monetary utility component measures $\theta^j_{i,2}$, individual i makes his mobility decision. Assuming risk-neutrality, the utility of job j consists of the sum of the wage offer from employer j and the non-pecuniary taste measure, $w^j_{i,2} + \theta^j_{i,2}$, where $j = c, o$ with c denoting the current employer and o the potential alternative employer. Thus, worker i moves away from his current firm if and only if $w^c_{i,2} + \theta^c_{i,2} \leq w^o_{i,2} + \theta^o_{i,2}$. Making use of the distributional assumption about the unobserved non-pecuniary heterogeneity, the probability of moving is $\Phi(\frac{w^o_{i,2} - w^c_{2,i}}{\sqrt{2}\sigma_\theta})$. All workers are employed in both periods and retire at the end of the second period.

Working backwards from the second period and suppressing the individual subscript i, with the updated expectation of the worker's productivity as well as the outside wage offer w^o_2 in hand, the optimization problem for the incumbent firm is

$$\max_{w^c_2} \; (\frac{\sigma^2_\epsilon}{\sigma^2_\eta + \sigma^2_\epsilon}(\mu_\eta + \delta_c) + \frac{\sigma^2_\eta}{\sigma^2_\eta + \sigma^2_\epsilon} y_{c,1} - w^c_2)(1 - \Phi(\frac{w^o_2 - w^c_2}{\sqrt{2}\sigma_\theta})), \tag{1.2}$$

while the outside employer maximizes

$$\max_{w^o_2} \; (\mu_\eta + \delta_o - w^o_2)\Phi(\frac{w^o_2 - w^c_2}{\sqrt{2}\sigma_\theta}). \tag{1.3}$$

Manipulation of the first-order conditions yields

$$w_2^c = \frac{\sigma_\epsilon^2}{\sigma_\eta^2+\sigma_\epsilon^2}(\mu_\eta+\delta_c) + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\epsilon^2} y_{c,1} - \sqrt{2}\sigma_\theta \frac{1-\Phi(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta})}{\phi(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta})}, \tag{1.4}$$

and

$$w_2^o = \mu_\eta + \delta_o - \sqrt{2}\sigma_\theta \frac{\Phi(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta})}{\phi(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta})}. \tag{1.5}$$

The monotone hazard rate feature of normal random variables, $d(\frac{1-\Phi(\theta)}{\phi(\theta)})/d\theta < 0$, implies the quasi-concavity of the objective functions so that the first-order conditions are sufficient for the maximization problems. The monotone hazard rate also guarantees that the two reaction functions defined by the two first-order conditions both have a positive slope less than one and that there is at most one intersection. The equilibrium exists and is unique.[13]

The wage offer of the current employer depends on the productivity signal sent by the worker. His first-order condition implies

$$\frac{\partial w_2^c}{\partial \eta} = \frac{\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\epsilon^2}}{1 - d(\frac{1-\Phi(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta})}{\phi(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta})})/d(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta})} > 0, \tag{1.6}$$

and

$$\frac{\partial w_2^c}{\partial \delta_c} = \frac{1}{1 - d(\frac{1-\Phi(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta})}{\phi(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta})})/d(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta})} > 0. \tag{1.7}$$

[13] This equilibrium is different from the Nash equilibrium of Greenwald (1986) due to our differing assumptions regarding the "random" quit behavior. His analysis relies on the assumption that the probability of quitting equals one if the outside offer is greater than the wage offered by the incumbent firm and equals a fixed value μ if the current employer offers a higher wage rate. As a result of that, firms in his model simply retain high ability workers by matching their outside offers.

In the context of match quality as an inspection good, the higher is the innate ability, the higher is the wage offered by the incumbent firm. The relationship between the current employer's wage offer and the worker's ability is not as strong as the relationship between the incumbent's wage offer and match quality. This simply follows from the different learning mechanisms attached to the individual's innate ability and to the match-specific productivity. Job match quality is learned instantly, without error ex ante, while ability has to be inferred from a series of noisy signals. As pointed out by Lange (2005), the parameter $\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\epsilon^2}$ plays a central role in the updating process of expected productivity. It represents the noisiness of the initial assessment of productivity relative to the noisiness of the subsequent signals. It is clear from (1.6) that if subsequent signals are more noisy than the initial expectation, that is, the smaller is $\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\epsilon^2}$, the lower the weight the incumbent firm places on innate ability in wage setting.

At the same time, private information prevents potential employers from obtaining updated expectations of unobserved productivity, as a result, the outside wage offer does not vary with η. Nevertheless, the relationship between the outside wage offer and match-specific productivity is positive, i.e.

$$\frac{\partial w_2^o}{\partial \delta_o} = \frac{1}{1 + d\left(\frac{\Phi\left(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta}\right)}{\phi\left(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta}\right)}\right) / d\left(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta}\right)} > 0, \tag{1.8}$$

which is intuitive given the assumption about the nature of job match quality. The relationship between mobility and ability generated by this model embodies adverse selection, so that

$$\frac{\partial \Phi(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta})}{\partial \eta} = -\frac{1}{\sqrt{2}\sigma_\theta}\phi(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta})\frac{\partial w_2^c}{\partial \eta} < 0. \tag{1.9}$$

That is, the probability of moving to another employer at the beginning of the second period is higher for less able workers. Again, taking the derivative with respect to the current firm's match quality,

$$\frac{\partial \Phi(\frac{w_{2,i}^o - w_{2,i}^c}{\sqrt{2}\sigma_\theta})}{\partial \delta_c} = -\frac{1}{\sqrt{2}\sigma_\theta}\phi(\frac{w_2^o - w_2^c}{\sqrt{2}\sigma_\theta})\frac{\partial w_2^c}{\partial \delta_c} < 0. \tag{1.10}$$

Equation (1.10), along with (1.8), captures the notion of a "good match" in the sense that it pays better and survives longer. Match quality has little impact on the implications of asymmetric employer learning highlighted by (1.7) and (1.9). Topel and Ward (1992),[14] using longitudinal employee-employer data, indicate that wage gains at job changes average about 10% and account for about one third of total wage growth during the first ten years in the labor market. This evidence should not be seen as contrary to the predictions of the asymmetric information model, as the match-specific productivity $\delta_{i,j}$ in my model does allow between-job wage growth, while their study does not deal with the quality of the workers across mobility levels.

To complete the model, I assume that the wage setting game on the entry-level labor market resembles the standard inspection good job matching models and the public learning models. Before period one, none of the firms in the labor market

[14]See Bureau of Labor Statistics (2006) for similar results from the NLSY-79.

knows more about the productivity of the worker than the initial expectation, the wage offers therefore do not depend on ability. Without loss of generality, I assume only two potential employers $j = J, K$ are competing for workers on the entry-level market. This particular case can be extended readily to the N-firm case. If the firms and workers share the same discount factor β, the ith individual's expected utility when working for firm J is

$$w_1^J + \theta_1^J + \beta \left[\Phi(\frac{w_2^K - w_2^J}{\sqrt{2}\sigma_\theta})(w_2^K + \theta_2^K) + (1 - \Phi(\frac{w_2^K - w_2^J}{\sqrt{2}\sigma_\theta}))(w_2^J + \theta_2^J) \right], \qquad (1.11)$$

where switching J and K yields the utility from working for firm K.

Taking the difference between the utilities from employer J and employer K produces the probability that firm J attracts the ith worker, $\Phi(\frac{w_1^J - w_1^K}{\sqrt{2}\sigma_\theta})$. Therefore, the profit maximization problem for employer J can be written as

$$\begin{aligned} \max_{w_1^J} \Phi(\frac{w_1^J - w_1^K}{\sqrt{2}\sigma_\theta})[\mu_\eta + \delta_J - w_1^J + \beta E_{\eta,\epsilon}((1 - \Phi(\frac{w_2^K - w_2^J}{\sqrt{2}\sigma_\theta})(\frac{\sigma_\epsilon^2}{\sigma_\eta^2 + \sigma_\epsilon^2}(\mu_\eta + \delta_J) \\ + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\epsilon^2} y_{J,1} - w_2^J))], \qquad (1.12) \end{aligned}$$

where $E_{\eta,\epsilon}$ denotes the expectation with respect to random variables η and ϵ. Replacing subscript J with K defines the optimization problem facing firm K. The symmetry implies that in the entry-level market equilibrium, both firms offer the same wage conditional on match quality, just as in the case of public learning, and better match quality commands a higher wage rate. Combining match-specific productivity and adverse selection on unobserved innate ability together implies that, although a "mismatch" leads to a lower wage and an early separation, job matching alone does not predict that movers are of lower quality than stayers, which is an important prediction from the two-period model.

1.3.2 A Three-Period Extension with Empirical Implications

While the two-period mover-stayer model does capture how private information held by the current employer affects the worker's mobility decision and wage determination, it is silent about the role of job mobility in sequential market trading, and it treats potential recruiting firms as completely "passive". The extension to a three-period setting allows the employment history of the workers on the second-hand labor market to serve as another signal to outside firms and provides an additional channel for recruiting employers to learn about the unobserved productivity of the workers. The two-period model suggests that worker ability and job mobility are negatively correlated because of adverse selection. It is reasonable to think that outside employers take prior job mobility into account when they make subsequent wage offers. The three-period extension also sharply contrasts with the match quality story of job mobility, in which the prior employment history is independent of the quality of a new match. Here, prior employment history is the driving force behind dynamic adverse selection.

From the perspective of potential employers, at the end of period two workers can be distinguished by their mobility decisions in the previous period. Conditional on each of the two possible values of the number of job changes, $m = 0, 1$, the bidding procedure is completely comparable to the one at the end of the first period. The only difference is that the recruiting firms now know that the distribution of η is different for workers with different m because market selection takes place at the end of period one. For workers with $m = 1$, that is, those who change jobs at the end of period one, the expected productivity becomes

$$E(\chi_j \mid m = 1) = E(\eta \mid w_2^c \leq w_2^o + \theta_2^o - \theta_2^c) + \delta_j. \tag{1.13}$$

Given that $\frac{\partial w_2^c}{\partial \eta} > 0$ and that everything else in the conditioning set of the expectation of η is independent of η, the end-of-period-one adverse selection shifts the ability distribution of the $m = 1$ workers toward the left. Similarly, asymmetric employer learning shifts the distribution of η for the stayers toward the right.

Meanwhile, the incumbent firms of workers with $m = 0$ continue learning in the Bayesian style. Their updated expectation is

$$\frac{\sigma_\epsilon^2}{2\sigma_\eta^2 + \sigma_\epsilon^2}(\mu_\eta + \delta_c) + \frac{2\sigma_\eta^2}{2\sigma_\eta^2 + \sigma_\epsilon^2}\frac{(y_{c,1} + y_{c,2})}{2}. \tag{1.14}$$

For the current employer of workers with $m = 1$, expected productivity takes the form of (1.1).

With repeated market transactions as in the three-period model, potential employers make offers to workers with $m = 1$ according to

$$\max_{w_3^{o',1}} \; (E(\eta \mid w_2^c \leq w_2^o + \theta_2^o - \theta_2^c) + \delta_{o'} - w_3^{o',1})\Phi(\frac{w_3^{o',1} - w_3^{c',1}}{\sqrt{2}\sigma_\theta}), \tag{1.15}$$

and make offers to workers with $m = 0$ according to

$$\max_{w_3^{o',0}} \; (E(\eta \mid w_2^c > w_2^o + \theta_2^o - \theta_2^c) + \delta_{o'} - w_3^{o',0})\Phi(\frac{w_3^{o',0} - w_3^{c',0}}{\sqrt{2}\sigma_\theta}), \tag{1.16}$$

where c' and o' denote the incumbent and outside employers at the end of period two and the numerical superscript on w_3 represents the value of m. We further obtain the corresponding optimization problems for the current firms

$$\max_{w_3^{c',1}} \; (\frac{\sigma_\epsilon^2}{\sigma_\eta^2 + \sigma_\epsilon^2}(\mu_\eta + \delta_{c'}) + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\epsilon^2} y_{c',2} - w_3^{c',1})(1 - \Phi(\frac{w_3^{o',1} - w_3^{c',1}}{\sqrt{2}\sigma_\theta})), \tag{1.17}$$

and

$$\max_{w_3^{c',0}} \; (\frac{\sigma_\epsilon^2}{2\sigma_\eta^2 + \sigma_\epsilon^2}(\mu_\eta + \delta_{c'}) + \frac{\sigma_\eta^2}{2\sigma_\eta^2 + \sigma_\epsilon^2}(y_{c',1} + y_{c',2}) - w_3^{c',0})(1 - \Phi(\frac{w_3^{o',0} - w_3^{c',0}}{\sqrt{2}\sigma_\theta})). \tag{1.18}$$

Comparing (1.15) and (1.16) with (1.3), it is easy to see that the outside wage offers for $m = 0$ individuals exceed those for $m = 1$ workers because $E(\eta \mid w_2^c > w_{2,i}^o + \theta_i^o - \theta_i^c) > E(\eta \mid w_{2,i}^c \leq w_{2,i}^o + \theta_i^o - \theta_i^c)$. Movers, those with $m = 1$, are adversely selected and have a worse η distribution than workers with $m = 0$. The labor market recognizes this in the third period by offering them lower wage rates. This is in contrast with the basic two-period framework where the equilibrium wage on the second-hand market does not depend on η, as suggested by (1.3). Previous research on asymmetric employer learning stops with the two-period framework and compares quality between movers and stayers in terms of some aptitude test scores such as the $AFQT$ score. However, that approach neglects the intensified adverse selection that is induced in a third period and beyond by the information contained in the worker's employment history, and does not generate empirical implications about the time path of the effect of ability on the wage offers from the incumbent and from the outside employers. The three-period extension argues that the correlation between the outside market equilibrium wage and unobserved ability increases with labor market experience and that market selection intensifies dynamically, so that

$$\frac{\partial\Phi(\frac{w_3^{o'}-w_3^{c'}}{\sqrt{2}\sigma_\theta})}{\partial\eta} = \frac{1}{\sqrt{2}\sigma_\theta}\phi(\frac{w_3^{o'}-w_3^{c'}}{\sqrt{2}\sigma_\theta})(\frac{\partial w_3^{o'}}{\partial\eta} - \frac{\partial w_3^{c'}}{\partial\eta}) < 0. \tag{1.19}$$

Although (1.19) is negative,[15] meaning that workers with lower values of η are still more likely to change jobs, the additional positive component $\frac{\partial w_3^{o'}}{\partial\eta}$ means fewer job changes after the second period than after the first period.[16] There is an enormous amount of heterogeneity among movers and an important tool for potential recruiting firms that want to learn about this heterogeneity is job mobility history. A typical two-period analysis, such as Schonberg (2005), predicts that the ability gradient of the job separation probability remains constant over time. In contrast, in the three-period case, incumbent firms gradually lose their informational advantage due to the accumulation of knowledge about η by outside employers with the result that employer learning on the market place converges to the public learning model over time. The intensified adverse selection implies a decreasing effect of innate ability on the job change probability. It is also obvious from (1.19), but still

[15]This is because the current employer still holds more information about η than the outside market, so that, $\frac{\partial w_3^{o'}}{\partial\eta} - \frac{\partial w_3^{c'}}{\partial\eta} < 0$.

[16]See Greenwald (1986) for a similar argument.

worth mentioning, that if the output sequence $(y_{j,1}, y_{j,2}, ..., y_{j,t})$ is available to all the firms, then $\frac{\partial w_t^o}{\partial \eta} = \frac{\partial w_t^c}{\partial \eta}$ and $\frac{\partial \Phi(\frac{w_t^o - w_t^c}{\sqrt{2}\sigma_\theta})}{\partial \eta} = 0$ for any t. When the information is imperfect but symmetric, the ability distribution is identical across mobility levels and the worker's job changing decision depends on the match quality δ and the non-pecuniary job characteristics θ.[17]

The first-order condition for (1.18) combined with (1.6) allows us to obtain

$$\frac{\partial w_3^{c',0}}{\partial \eta} = \frac{\frac{2\sigma_\eta^2}{2\sigma_\eta^2+\sigma_\epsilon^2} + [d(\frac{1-\Phi(\frac{w_3^{o',1}-w_3^{c',1}}{\sqrt{2}\sigma_\theta})}{\phi(\frac{w_3^{o',1}-w_3^{c',1}}{\sqrt{2}\sigma_\theta})})/d(\frac{w_3^{o',1}-w_3^{c',1}}{\sqrt{2}\sigma_\theta})]\frac{\partial w_3^{o',0}}{\partial \eta}}{1 - d(\frac{1-\Phi(\frac{w_3^{o',0}-w_3^{c',0}}{\sqrt{2}\sigma_\theta})}{\phi(\frac{w_3^{o',0}-w_3^{c',0}}{\sqrt{2}\sigma_\theta})})/d(\frac{w_3^{o',0}-w_3^{c',0}}{\sqrt{2}\sigma_\theta})} > \frac{\partial w_2^c}{\partial \eta}. \tag{1.20}$$

This inequality explicitly spells out employer learning: for workers staying with their initial employers for the entire three periods, wage rates depend more and more on unobserved productivity. Moreover, and perhaps more importantly, this increase in the correlation between wages and ability is larger than that in the pure symmetric employer learning model. To see this, notice that the numerator of $\frac{\partial w_3^{c',0}}{\partial \eta}$ has two parts. The first term comes from the current employer learning more over time, as argued by Farber and Gibbons (1996) and Altonji and Pierret (2001), i.e. $\frac{2\sigma_\eta^2}{2\sigma_\eta^2+\sigma_\epsilon^2} > \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\epsilon^2}$, where $\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\epsilon^2}$ appears as the numerator of (1.6). The second term is a special feature of this model and follows directly from the market feedback of the job mobility decision. It represents the additional premium put on unobserved productive ability by the current employer because he knows that outside recruiting

[17] Jovanovic (1979a) (footnote 11, p. 982) writes "...in other words, the model does not imply that "movers" should do worse than "stayers" even though empirically this appears to be true..."

firms can partially learn about the ability of the workers via the employment history. Existing asymmetric employer learning models have been unable to lay this out clearly and convincingly because they do not take into account the signaling effect of job mobility on outside wage offers.

For workers who change jobs after period one, the increase in the correlation between market wage rates and innate ability η over time also holds. The wage determination process in (1.3) implies that for $m = 1$ workers, wage offers for period two are constant over η because only μ_η enters (1.3), but the story told by (1.15) and (1.17) is that whether or not these workers decide to change jobs at the end of the second period, it is always the case that $\frac{\partial w_3^{o',1}}{\partial \eta} > 0$ and $\frac{\partial w_3^{c',1}}{\partial \eta} > 0$. The three-period asymmetric employer learning model agrees with the public learning models and the earlier two-period analyses of private employer learning in that wages are increasingly correlated with unobserved productivity as labor market experience accumulates. It departs from existing studies in terms of its implications for the differential returns to ability for people with different job changing patterns, even conditional on labor market experience and job tenure.

Public information makes $\frac{\partial w_2^c}{\partial \eta} = \frac{\partial w_2^o}{\partial \eta}$ and $\frac{\partial w_3^{c',1}}{\partial \eta} = \frac{\partial w_3^{c',0}}{\partial \eta} = \frac{\partial w_3^{o',1}}{\partial \eta} = \frac{\partial w_3^{o',0}}{\partial \eta}$, at any point in time, for workers with the same amount of labor market experience. All the wage offers, no matter where they come from, depend on η in the same way, given the independence of η and match quality δ. And, individuals with different patterns of prior job separations have the same returns to unobserved productive ability. The two-period mover-stayer model in Schonberg (2005) recognizes that innate ability has a stronger impact on wage offers for incumbent firms than for outside employers, and that the difference is greater as the informational advantage of the incumbent firm increases. Based on this implication, Schonberg (2005) predicts a positive coefficient on the variable that interacts the $AFQT$ score and job tenure in a wage regression. While in general this intuition still holds for the three-period model, the absence of learning from outside employers in the two-period model implies the independence of job mobility frequency and the differential impacts of ability on wages offers from current and potential employers. In the three-period model, when recruiting firms on the outside market take job mobility history into consideration at the end of period two, the informational advantage of the current firm in period two is lower than that in period one and the reduction is higher for workers with more frequent job changes. The more information the outside firms have, the smaller is the difference between the impacts of ability on wages for the incumbent versus outside firms. This implication is not consistent with the mover-stayer model in which learning by recruiting employers is ruled out. Thus, the signaling effect of the prior job moves implies a negative coefficient for the variable which interacts the test score, job tenure, and frequency of job mobility.

One real world application of employer learning models is to study statistical discrimination, where firms distinguish among workers on the basis of easily observable variables that may be correlated with productivity like years of eduction, gender, and race. Altonji and Pierret (2001) describe the intuition of such analyses succinctly:

"As employers learn about the productivity of workers, s [which is schooling] will get less of the credit for an association with productivity that arises because s

is correlated with z [a variable like AFQT score that is initially unobserved, but is positively correlated with both s and output], provided that z is included in the wage equation with a time dependent coefficient and can claim the credit."

Note that because a worker's education level is part of the firm's initial information set and is incorporated into the determination of first-period wages, subsequent innovations in wages can not be forecast from years of schooling.[18] The empirical regularity of a declining time path of the returns to schooling arises solely out of the relationship between education and unobserved innate ability. To include easily observed time-invariant characteristics like schooling in the model, I can redefine productivity as

$$\chi_j = rs + \eta + \delta_j, \tag{1.21}$$

where s denotes the years of schooling. Keeping everything else in the model unchanged, the time path of the returns to η is shown to be increasing as firms accumulate more information, regardless of whether it is symmetrically or asymmetrically distributed between the incumbent and potential employers. This learning effect on the impact of ability spills over to the schooling variable that firms use to statistically discriminate among new employees. Thus, following the same logic as in Altonji and Pierret (2001), given $cov(\eta, s) > 0$, the model predicts that the coefficient on s in a wage regression declines with labor market experience when an ability measure unobservable to employers is included.

Unlike years of schooling, which is a time-invariant ability signal known to

[18]Farber and Gibbons (1996) make this point and predict a zero coefficient on the interaction term between education and experience when the residualized $AFQT$ score is included in a wage regression.

all the employers upon market entry, the job mobility history serves as a time-varying signal to the outside firms in the three-period model. The information contained in a worker's employment history is utilized by potential hiring firms to evaluate the productivity of the workers. The fact that learning by the outside firms increasingly makes the time-invariant signal s redundant is another special feature of the three-period model. Traditional analyses ignore the informational content of prior job moves and imply that the effect of education on wages is independent of job mobility, conditional on experience and job tenure. The signaling effect of employment history, however, predicts a negative coefficient associated with the interaction between years of schooling and the frequency of job mobility in a wage regression.

1.4 The Data: NLSY-79

The empirical work is based on White, Black, and Hispanic males from the 1979-2000 waves of the National Longitudinal Survey of Youth (NLSY-79). A key feature of the NLSY-79 is that in addition to detailed information on family background, scholastic achievement, and labor market outcomes, its work history file provides an unusually complete picture of employment for a cohort of young workers during a period when they have made transitions from school to work. This includes records of virtually every job held. As a result, it is ideal for my study. The original NLSY-79 sample consists of 12,686 men and women (age 14-22 in 1979) who were interviewed annually between 1979 and 1994 and biennially from 1996 to the present. There are three subsamples in the NLSY-79: a cross-sectional sample representative of young people; a supplemental sample designed to oversample Hispanic, Black, and economically disadvantaged White youth; and a sample designed to represent the population of those enlisted in one of the four branches of the military. I exclude the military subsample from my analysis because, following the 1984 interview, the military subsample were no longer eligible for interview and it is hard to construct a long enough employment series for respondents from this subsample.

In order to abstract the analysis from family and fertility decisions and focus on a subpopulation with strong labor force attachment, I use male sample only. There are 5579 males in the original NLSY-79 sample after eliminating the military respondents. I exclude employment and wage observations from before a person leaves school and begins to accumulate labor market experience, and only count job changes from that point. My definition of the school-to-work transition date follows that of Altonji and Pierret (2001):[19] the month and year of the respondent's most recent enrollment in school at the first interview when the respondent is not currently enrolled. I lose 49 individuals from the original sample because their school exit date is indeterminate according to this definition. I also exclude 1137 individuals whose labor market entry occurs before January 1978. Detailed information on employment activities is only reported from that date onwards in the work history file, so I can not construct accurate measures of overall mobility, work experience, and job tenure for workers who start their careers before January 1978. Additionally,

[19] Alternatively, Farber and Gibbons (1996) define a transition as occurring if the worker is classified as non-working for at least one year, followed by at least two consecutive years classified as working, where a worker is classified as working when she has worked at least 26 weeks, and during these weeks at least 30 hours, since the last interview.

I delete 47 individuals because their actual labor market experience or job seniority is indeterminate and another 12 individuals whose wage information is unreasonable, which brings the sample size down to 4334. Furthermore, 202 individuals in the sample did not take the Armed Services Vocational Aptitude Battery[20] (ASVAB) tests which are used by the NLSY-79 to construct the $AFQT$ score.[21] After dropping them, the remaining sample consists of 4132 individuals with 48,617 person-year observations.

Table 1.1 contains summary statistics for observations used in the analysis. Actual labor market experience is the number of weeks in which the worker works more than 30 hours divided by 52 after the transition from school to work. I do not count part-time employment, self-employment, time spent working without pay, time spent unemployed, and time spent out of the labor force.[22] Job tenure is calculated as the number of weeks divided by 52 spent in full-time employment with the same employer. The wage measure is the hourly wage at the beginning of each employment spell from the NLSY-79 work history file. Wages are deflated by the Consumer Price Index with 2002 as the base year; values below \$1 and above \$300 are considered unreasonable and dropped.[23]

The job mobility count is obtained from the work history file of the NLSY-79,

[20]The $AFQT$ score is the sum of the raw scores from the following four sections of the ASVAB: arithmetic reasoning, word knowledge, paragraph comprehension, and one half of the score from numerical operations section.

[21]The ASVAB was administered to the NSLY-79 respondents in 1980, thus, different respondents took it at different ages. To eliminate age effects, I standardize the $AFQT$ score within each birth cohort.

[22]For the individuals who work more than one job at a point in time, I only consider the job at which the respondent works the most hours during the week.

[23]I tried other cutoff values, such as \$0.5 and \$200. My empirical results are not sensitive to the changes in the values used to define unreasonable wage observations. See Bollinger and Chandra (2005) for more on this issue.

which reports the starting dates for the jobs held at the time of each interview, as well as for up to five jobs[24] that began and ended since the last interview. I link all the jobs across survey years[25] and construct a complete employment history for each individual in the sample. The frequency of job mobility is calculated as the individual's mean number of prior job separations as of time t. Table 1.2 shows the distribution of the number of job separations by each worker during the first 2, 5, and 10 years of his career as well as the total number of jobs held.

The average number of job separations in the first ten years is 5.6 with a standard deviation of 4.0. The mean number of jobs actually held[26] is 6.2 with a standard deviation of 4.0. Table 1.2 also illustrates that only 3% of individuals experience no job changes in the first 10 years of their career, while around 10% of workers remain with their initial employers during the first five years and 38% for the first two years. At the other extreme, 11% of individuals separate from 10 or more employers within the first ten years after the school to work transition; that is, they average over one job separation per year for the 10-year period. Table 1.2 demonstrates that the typical individual in the sample is quite mobile early in his career.

The data on job separations also suggest that job mobility slows over time. While this can not be said to be attributable solely to the intensified adverse selec-

[24]The NLSY-79 collects information on all jobs held by a respondent since the last interview, however, the percentage of respondents who report more than five jobs in each survey year is less than 1%.

[25]As the same employer can receive different job codes across survey years, it is necessary to use beginning and ending dates as well as a series of matching variables to determine the job code in the previous survey for every employer in the current survey and to decide whether it is a new job.

[26]Topel and Ward (1992) find that the average worker holds 6.1 jobs by the time he or she has eight years of potential labor market experience in their longitudinal employer-employee data.

tion, it is at least consistent with the three-period model where outside employers take the employment history into account. In contrast, neither the symmetric employer learning model nor the two-period mover-stayer model implies a decline in job turnover conditional on innate ability. About 30% of the sample undergoes no job changes during the second five years, and 46% undergoes at most one job separation during that time period.

Throughout the paper, I use the total number of job separations rather than the number of voluntary job separations. It is not clear how to distinguish between involuntary and voluntary job separations in the NLSY-79. The NLSY-79 codes a large number of reported reasons for each job separation, including "bad working condition", "own illness", "found better job", "spouse changed jobs", *etc.* If I delete all job separations corresponding to "layoff" and "discharged/fired", then 70% of all the job separations remain. However, those remaining job separations still include ones caused by family reasons as well as ones caused by "found better job" and "pay too low". Moreover, the explanation for over 25% of all job exits is coded as either "other" or missing, so I must either eliminate those jobs or arbitrarily assign them to voluntary or involuntary categories.

1.5 Econometric Specification and Empirical Results

One of the empirical implications of an employer learning model in which information about a worker's productivity is public is a correlation of zero between the worker's innate ability and his probability of changing jobs. Both the two-period mover-stayer model of asymmetric employer learning and my three-period extension challenge this by showing that the average quality of the job-changing pool is lower than that of the pool of stayers. What differentiates these two versions of the asymmetric information model is the prediction regarding how the relationship between ability and the job change probability changes over time. In the absence of learning by outside employers, the mover-stayer story implies a constant correlation between η and the probability of job change. On the other hand, information accumulation by potential employers through the observed job mobility history implies that this relationship becomes weaker and weaker over time.

I test this implication of the learning model by estimating a probit model where the dependent variable is an indicator of whether the worker experiences a job changes in a given period,

$$\begin{aligned} Pr(JobChange_{i,t} = 1) =& \Phi(\beta_0 + \beta_1 AFQT_i + \beta_2 (Exp_{i,t}/10) \\ &+ \beta_3 (AFQT_i \times Exp_{i,t}/10) + \beta_X' X_{i,t}), \end{aligned} \tag{1.22}$$

where i is an individual, t is a survey year, $Exp_{i,t}$ is actual labor market experience and $X_{i,t}$ is a vector of other control variables. Throughout the empirical analysis, I normalize all the interactions between schooling and the $AFQT$ score with experience to represent the change in the regression slope between $Exp = 0$ and $Exp = 10$. Also, all of the standard errors reported in this paper are based on White/Huber standard errors that account for arbitrary forms of heteroskedasticity and correlation among the multiple observations for each individual. All of the estimates in this paper are weighted by the sampling weights provided by the NLSY-79. Coefficients β_1 and β_3 should both be zero under the assumption of public learning. All of the asymmetric learning models imply a negative β_1, but only a model with a signaling effect of the job mobility history implies a positive coefficient β_3.

The results of the job changing regressions are presented in Table 1.3. Column (1) in the table is the mean derivative estimated from a probit model where the standardized $AFQT$ score is the only explanatory variable. A one standard deviation increase in the test score is accompanied by a 3.6 percentage point decrease in the probability of changing jobs. This preliminary evidence clearly rejects the symmetric employer learning hypothesis via a highly statistically significant probit marginal effect associated with the $AFQT$ score. To distinguish the two types of asymmetric learning hypotheses, column (2) estimates the same probit with experience and the interaction between the $AFQT$ score and experience as additional independent variables. The mean marginal effect on the $AFQT$ score remains statistically significant, and there is a positive and statistically significant estimate for the interaction term of the $AFQT$ score and labor market experience. The decreasing time path of the absolute value of the impact of the $AFQT$ score on the probability of changing jobs is a unique prediction from the three-period adverse selection model. It captures the closing of the informational gap between current and outside employers about the productivity of the workers. The estimated marginal effect of 0.026 strongly suggests that not only does the current employer learn, but potential employers also accumulate new information about a worker's innate ability, so that over time ability matters less and less in job changes.

Including additional covariates in the probit regression, column (3) controls for race, industry and occupation, and year effects. These control variables weaken the correlation between the $AFQT$ score and job mobility, but by no means eliminate it. The probit marginal effects associated with the $AFQT$ score and the interaction term are still statistically significant and qualitatively tell the same story as column (2). Hansen, Heckman, and Mullen (2004) find strong evidence in the NLSY-79 suggesting that schooling is an important determinate of measured achievement such as the ASVAB scores;[27] their estimated increase in the $AFQT$ score per year of education for the average person is 0.17 standard deviation. To deal with the effect of schooling on the test score, I construct the educational level and school enrollment status at the ASVAB test date for each individual in the sample[28] and include them in the probit regression of column (4). Putting schooling information as of the test date into the model substantially reduces the magnitude of the probit coefficients: the estimated marginal effects of the $AFQT$ score and the interaction term stand at -0.012 and 0.014, respectively. Nevertheless, both are statistically significant at the 5% level, and the overall conclusion is the same as that drawn from column (2) and column (3). To summarize, the probit estimates shown in Table 1.3 are consistent with an asymmetric employer learning model in which both

[27]See Neal and Johnson (1996) and Cascio and Lewis (2006) for a similar result.

[28]The ASVAB was administered during July–October 1980. Respondents in the NLSY-79 were interviewed during January–August 1980 and January–July 1981. The NLSY-79 also includes a measure of schooling and enrollment status as of May 1 of each survey year. Since the academic year commonly ends in June, individuals advance to a higher completed grade level in June. I use the highest grade completed and enrollment status as reported in the 1980 survey as schooling and enrollment values at the test date if the interview was conducted during July–August 1980, and I use the variables reported in 1981 if the interview was conducted during January-April 1981. For the remaining respondents, I use the variables for May 1, 1981.

the incumbent and the outside employers gather information about the worker's unobserved productivity. The negative and statistically significant mean marginal effect of the $AFQT$ score on the job change probability rejects the public learning hypothesis, and the gradually decreasing association between the test score and the probability of job separation is at odds with the two-period mover-stayer model.

To further distinguish the two versions of asymmetric employer learning models, one without outside employers learning and the other with potential firms learning through the employment history of the job candidate, I make use of the empirical framework advanced by Farber and Gibbons (1996) and Altonji and Pierret (2001). Under the assumption of pure public learning, Altonji and Pierret (2001) estimate a version of the standard earnings equation with schooling and the $AFQT$ score interacted with labor market experience

$$\begin{aligned}\ln w_{i,t} =& \alpha_0 + \alpha_1 Schooling_i + \alpha_2 AFQT_i + f(Exp_{i,t}/10) + \\ & \alpha_3 (Schooling_i \times Exp_{i,t}/10) + \alpha_4 (AFQT_i \times Exp_{i,t}/10) + \alpha_X' X_{i,t} + \xi_{i,t},\end{aligned} \tag{1.23}$$

where the log wage for the ith worker at time t depends on his schooling, his $AFQT$ score, labor market experience, and other observable characteristics $X_{i,t}$. Their model shows that when the $AFQT$ score is included in the regression as an ability measure, the time path of the coefficient on schooling declines with experience while the coefficient on the $AFQT$ score increases with labor market experience. As employers learn more about the productive ability of a worker, they rely less on the easily observable variables such as education in the wage setting process. Note that my model in Section 3 explicitly demonstrates that their implications regarding the signs of α_3 and α_4 also hold even when the information about the worker's productivity is asymmetric.

Table 1.4 shows the results generated when their wage regressions are run on my sample. In addition to the explanatory variables shown in the table, all of the regressions control for race, a cubic in experience, industry and occupation, year effects, education interacted with year effects, and Black and Hispanic interacted with year effects. The first two columns report OLS estimates of (1.23). Columns three and four report two stage least squares (2SLS) estimates using potential experience as an instrument for actual labor market experience.[29] Looking across the columns, the two sets of coefficient estimates tell the same story and confirm the empirical findings of Altonji and Pierret (2001) that the impact of the $AFQT$ score on wages increases with labor market experience and the coefficient on years of schooling decreases with experience.

While these estimates support the view that employers acquire new information about workers' productivity over time, they do not allow us to distinguish among public learning, asymmetric learning without the outside employer accumulating new information, and the three-period model developed in the Section 3. When the recruiting employers gather new information about the ability of the worker through his employment history, my model predicts a declining difference be-

[29] Altonji and Pierret (2001) argue that the implications of employer learning for the wage equation may change if the intensity of work experience conveys information to employers about worker quality.

tween the impacts of ability on wage offers from the incumbent and the outside firms with increasing job mobility, and therefore a negative coefficient for the interaction term involving the $AFQT$ score, job tenure, and the frequency of job mobility. I estimate the following wage regression,

$$\begin{aligned}\ln w_{i,t} =& \gamma_0 + \gamma_1 Schooling_i + \gamma_2 AFQT_i + f_1(Exp_{i,t}/10) + f_2(Tenure_{i,t}/10) + \\ & \gamma_3 Freq_{i,t} + \gamma_4(Schooling_i \times Exp_{i,t}/10) + \gamma_5(AFQT_i \times Exp_{i,t}/10) + \\ & \gamma_6(AFQT_i \times Tenure_{i,t}/10) + \gamma_7(AFQT_i \times Freq_{i,t}) + \\ & \gamma_8(AFQT_i \times Tenure_{i,t}/10 \times Freq_{i,t}) + \gamma_X' X_{i,t} + u_{i,t}, \end{aligned} \tag{1.24}$$

where $Tenure_{i,t}$ denotes job tenure and $Freq_{i,t}$ denotes the ith worker's frequency of job mobility as of time t. The closing informational gap between the current and outside firms through employment history implies that $\gamma_8 < 0$. On the other hand, if the outside employers ignore the information concerning the worker's innate ability contained in the job mobility history as described in the two-period model, or if their learning process occurs through other channels, then we would expect to find $\gamma_8 = 0$.

The OLS estimates of (1.24) are displayed in Table 1.5. Other covariates that I control for are a cubic in experience, a cubic in job tenure, race, industry and occupation, year effects, education interacted with experience, education interacted with year effects, and interactions between the race dummies and the year effects. Column (1) provides the regression estimates before controlling the measure of job mobility. This coincides with existing tests of the asymmetric employer learning model such as those in Schonberg (2005). If employer learning is private, the impact of ability on the wage offer of the current employer exceeds that of the outside firms, which predicts $\gamma_6 > 0$ in (1.24), as opposed to the case of pure symmetric learning which implies $\gamma_6 = 0$. In line with Schonberg (2005), my estimate for the coefficient associated with the interaction term between the $AFQT$ score and job tenure shows a positive sign that is consistent with the asymmetric information model but fails to pass the significance test at conventional levels. Schonberg (2005) only finds a marginally significant estimate for γ_6 after controlling for interactions between the $AFQT$ score and higher order tenure terms for her university graduates sample.

Column (2) of Table 1.5 estimates a complete version of (24) and paints a very different picture. Not only does the positive and significant coefficient estimate of 0.021 for the $AFQT$ score and job tenure interaction favor asymmetric employer learning, but the estimated coefficient of -0.013 associated with the interaction between the $AFQT$ score, job tenure, and the frequency of job mobility also suggests that outside employers indeed acquire knowledge about the worker's ability through his job change pattern, with the result that the informational discrepancy between the incumbent and potential employers in turn diminishes with experience. Conditional on job tenure, I still see a positive coefficient estimate of 0.049 for the variable interacting experience and the $AFQT$ score, which reinforces the conclusion that learning on the labor market is not purely asymmetric. I also find a negative coefficient estimate for the frequency of job mobility[30] which suggests that early-career mobility does little to help but can do a significant amount to hurt wages. Although

[30]See Light and McGarry (1998) for similar findings.

this may not be a defining implication from the model, it is consistent with the intensified adverse selection story.

As a time-varying signal of the worker ability, the availability to the market of job mobility history also has implications for the role played by the time-invariant observables that the employers initially use to statistically discriminate among workers. To study how the worker's career path affects the employer learning through easy-to-observe characteristics like schooling, I estimate a wage equation of the type

$$\begin{aligned}\ln w_{i,t} =& \lambda_0 + \lambda_1 Schooling_i + \lambda_2 AFQT_i + f_1(Exp_{i,t}/10) + f_2(Tenure_{i,t}/10) + \\ & \lambda_3 Freq_{i,t} + \lambda_4(Schooling_i \times Exp_{i,t}/10) + \lambda_5(AFQT_i \times Exp_{i,t}/10) + \\ & \lambda_6(Schooling_i \times Ten_{i,t}/10) + \lambda_7(AFQT_i \times Tenure_{i,t}/10) + \\ & \lambda_8(Schooling_i \times Freq_{i,t}) + \lambda_9(AFQT_i \times Freq_{i,t}) + \lambda_X' X_{i,t} + v_{i,t}. \qquad (1.25)\end{aligned}$$

Table 1.6 reports the OLS estimates of (1.25) where X contains the same additional variables as in Table 1.5. Column (1) excludes the job mobility measure and its interactions with schooling and the $AFQT$ score. Although the general pattern of the coefficients on the interactions between the $AFQT$ score and schooling with experience suggested by the learning model is still borne out by the data, the highly imprecise estimates for λ_6 and λ_7 tell us nothing about the nature of employer learning. In column (2) of Table 1.6, the estimates support the three-period model in which potential employers learn from the job mobility patterns. In particular, the negative and significant coefficient estimate for the interaction term of schooling and the frequency of job mobility implies that education plays less of a signaling role as outside firms rely more on employment history to assess the value of the worker's productivity. Information revelation as an immediate consequence of intensified adverse selection helps the recruiting firms to become better informed about the quality of the workers in the job-changing pool. Also, the coefficient on the interaction of education and job tenure is negative, though only significant at the 10 percent level. It provides suggestive evidence that potential employers depend more, relative to the incumbent employer, on schooling to determine their wage offers. Taken together, these empirical results strongly surpport the aforementioned three-period model in which not only do incumbent employers learn, but outside firms also actively extract information from workers' employment histories.

1.6 Conclusion

How do firms learn about their workers' productivity? Do they use easily observed characteristics such as education and race to statistically discriminate among their workers? Do current employers have more information about the worker's productivity than outside firms? If they do, what can outside firms do to minimize their informational disadvantage? During the past decade, labor economists have developed employer learning models to better understand the answers to these questions. Although consensus has been reached, both theoretically and empirically, on the existence of employer learning in the market place, our understanding of whether learning is asymmetric and how the information asymmetry is resolved remains unsatisfactory. This chapter builds a learning model under the hypothesis that incumbent employers have superior information about the productivity of its workers. A special feature of my model is that outside employers, by observing workers' job mobility histories, also have access to information about the workers' ability. This attribute differentiates the present model from existing models of asymmetric employer learning that are based on the two-period mover-stayer model. My model also includes a match-specific productivity component that is known ex ante and I show that because the distribution of match quality is independent of worker ability and the quality of previous matches is irrelevant to newly formed job matches, the presence of match-specific productivity does not alter the nature of employer learning about the innate ability of their workers.

It is important to underscore the limits of this study. The literature has long recognized that human capital accumulation may undermine the predictions from learning models. Although the empirical evidence of intensified adverse selection established through our probit estimates is based on a robust feature of the model, the estimates of the wage regressions, especially the coefficient associated with the interaction between the $AFQT$ score and job seniority, also fit a model in which ability aids the acquisition of specific human capital.[31] This complementarity between ability and specific capital implies that more able workers command higher returns to job tenure, which implies a positive coefficient for the interaction term between the $AFQT$ score and job seniority. It is very difficult to distinguish the present model from a specific human capital model. I can only partially address this concern, following Schonberg (2005), by looking at differential returns to job

[31] See Altonji and Spletzer (1991) for such an example.

tenure by education level. The estimate from column (2) of Table 1.6, even though only marginally significant, implies lower returns for higher educated workers. If we expect individuals with more years of schooling to benefit more from job seniority as the human capital theories imply, my negative coefficient is at odds with such a prediction. My model also rules out the possibility of an experience-good nature of job match, because analysis of an asymmetric employer learning model that also allows learning about the match quality is rather complex and beyond the scope of the current study.

To conclude, the empirical evidence from the NLSY-79 broadly supports the implications from the dynamic adverse selection model: ability is negatively correlated with the probability of changing jobs but this association weakens as young workers advance in their careers; accruing information through observing the employment history on the part of outside firms gradually eliminates the knowledge gap between them and incumbent firms; this in turn reduces over time the difference of the impacts of ability on wage rates between them and the incumbent firm, and allows them to be less dependent on the easy-to-observe characteristics of the workers.

Chapter 2

Teenage Childbearing and Maternal Schooling Outcomes: Evidence from Matching

2.1 Introduction

For more than forty years, teenage pregnancy and childbearing in the United States has been a continuing source of concern to social science researchers and policymakers. Simple tabulations of any nationally representative data set reveal that adolescent parenthood reduces the teenager's educational attainment, lowers the probability of her eventual marriage, increases her welfare recipiency rate and decreases her subsequent labor market income (Trussel 1976 and 1988; Hofferth and Moore, 1979; and Upchurch and McCarthey, 1990). Despite the widely quoted decline of the teen pregnancy and birth rates since 1990, we still witness over 750,000 women aged 15-19 become pregnant each year, among which more than 265,000 pregnancies happen to women aged 15-17. Overall, 75 pregnancies occur every year per 1000 women aged 15-19 and the rate is 45 for women between 15 and 17 years old. This rate is even higher for minority groups; black women have the highest teen pregnancy rate with 134 pregnancies per 1000 women aged 15-19 and the rate for Hispanic teens is 131 per 1000 (Guttmacher Institute, 2006). Moreover, teen pregnancy and birth rates are much higher in the United States than in many other developed countries-twice as high as in England and Wales or Canada, and eight times as high as in the Netherlands or Japan (Guttmacher Institute, 2006). Not only can we see millions of American families struggle individually with the emotional and economic challenges that adolescent pregnancy and childbearing bring, it also poses a significant financial burden to society at large, as the average teen mother receives welfare assistance valued at over $1400 annually during her first 13 years of parenthood (Maynard, 1996). The summary in *Risking the Future*, the 1987 report of National Research Council, is well known: "Women who become parents as teenagers are at greater risk of social and economic disadvantage throughout their lives than those who delay childbearing". In his 1995 State of the Union address, President Clinton echoed this perspective by declaring that teenage pregnancy was "our most serious social problem".

Human capital accumulation is thought to be an important channel through which early parenthood impacts the socio-economic outcomes of teen mothers. This chapter analyzes to what extent teenage fertility affects the educational attainment of adolescents. While the strong correlation of educational attainment and motherhood timing is obvious from a simple cross tabulation, empirical economists have known the phrase "correlation does not establish causation" long enough to recognize the difficult challenge of isolating the causal effects. Non-random selection of women into the population of teen mothers expressly stands in the way of disentangling the causal link between schooling outcomes and teen fertility from other confounding factors. Specifically, those women who bear children as teenagers may well be the same women whose schooling levels would have been lowest in any case.

In other words, if there are observed or unobserved variables that are correlated with fertility and the mother's educational outcome, then a negative correlation between teenage childbearing and schooling may arise even if teenage births have no causal influence on the mother's subsequent outcomes at all.

In order to identify and consistently estimate the mean impact of teenage childbearing on the subpopulation who bear children early in their life cycle, the ideal controlled experiment should randomly assign young women pregnancy and birth timings so as to equate the distributions of observed and unobserved confounding factors between teen mothers and their non-parenting counterparts. In the absence of such data, the distribution of counterfactual educational outcomes that the adolescent mothers would have attained had they not given birth as teens may well be quite different from the factual schooling distribution of non-teen mothers that we observe in the data and with which we approximate the missing counterfactuals.[1] Numerous econometric techniques have been advanced to deal with the selection issue and uncover the causal relationship between teen fertility and various adolescent socio-economic outcomes, with focuses on different parameters of interest and a wide range of identification strategies used in the literature, the results are so far quite mixed.

This chapter takes a nationally representative sample from the National Longitudinal Survey of Youth 1979 (NLSY-79) and applies kernel matching estimators to three comparison sub-samples to re-examine the relationship between early parent-

[1] The common mean parameters of interest in the treatment effects literature only requires certain moment independence conditions, which are less restrictive than the conditions imposed on the whole distribution.

hood and maternal educational attainment. Instead of just using the entire sample of women who delay their childbearing age until after their 18th birthdays as a comparison group, we take advantage of the complete fertility history of each woman available in the NLSY-79 file and try to improve the comparability of treatment and comparison by utilizing two other comparison groups, those women who experience their first pregnancy between their 18th and 19th birthdays and have a live birth and those who become pregnant for the first time between 19 and 20 years old and have a live birth. By doing so we can avoid, to some extent, the bias arising from the unobserved confounding factors entering the fertility timing decision and the schooling choice as well, although as we discuss later in the chapter, the nature of the economic parameters we estimate changes as these two comparison groups provide counterfactuals defined by becoming pregnant and having a live birth at specific ages.

As we worry about the possibility that some of the causal estimates obtained from semi-parametric matching may not be robust to violations of our identifying assumption, we conduct two sensitivity analyses. Within a parametric regression framework, we follow Altonji, Elder, and Taber (2005) and use a bivariate probit model to assess how strong selection on unobservables would have to be in order to imply that the entire estimated effect should be attributed to selection bias. Another robustness check builds on the work of Ichino, Mealli, and Nannicini (2007) and Rosenbaum (1987). Here, the intuition is quite simple. Suppose that our selection-on-observables identifying assumption is not satisfied given the set of covariates we condition on but would be satisfied if we could observe an additional binary covariate. This nonparametric sensitivity analysis simulates this binary variable in the data and uses it as an additional matching factor; a comparison of the estimates obtained with and without matching on this simulated confounding factor tells us to what extent the estimator is robust to this specific deviation from our identifying assumption.

The remainder of the chapter unfolds as follows. Section 2.2 surveys the current literature and Section 2.3 defines our causal parameter and fleshes out the identification strategy as well as the matching estimator. The samples and the variables of NLSY79 we use along with some summary statistics are laid out in Section 2.4. For the purpose of comparison, Section 2.5 presents probit estimates of the effects of adolescent motherhood on maternal education from the same samples. The empirical results from matching methods are reported and discussed in section 2.6. Section 2.7 assesses the role played by unobserved confounding factors by utilizing both parametric bivariate probit model and a nonparametric simulation-based sensitivity analyses. Concluding remarks are offered in Section 2.8.

2.2 The Literature: A Cautionary Tale of Various Parameters of Interests and Identification Strategies

Social scientists have failed to establish convincing causal evidence of the link between teenage childbearing and mothers' educational outcomes. Two related features of the literature stand out in this debate. In order to achieve identification of teenage childbearing effects, different exogeneity assumptions have been invoked in previous work and as a result, especially in the presence of heterogeneous effects, different treatment parameters have been estimated. One key confusion within the literature over "the effect" of teen motherhood is that in fact there are many different effects and that each different effect can be estimated for different groups of the population depending on the identification strategy.

The early multivariate analyses address the problem using traditional regression-based mothods and adjust the pre-existing differences between teen mothers and girls in the comparison group by controlling for observed characteristics.[2] Large negative effects of early motherhood are typically suggested by this type of research. The identification of this linear control function model[3] relies on the assumptions that conditional on all the observables available to economic analysts, fertility timing is exogenous with respect to future schooling outcomes and that the exact functional form of the outcome equation is correctly specified by the linear model. These early papers generally confines themselves in the world of common treatment effects, or at most, heterogeneous effects in terms of observables. Aside from the potential for unobservables to compromise the exogeneity of the timing of fertility, any misspecification of the linear functional form will fail to account for the selection on the non-linear part of the covariates, which leads to omitted variable bias.

Another line of work takes advantage of the sibling method or "within-family" estimator (Griliches, 1979). The missing counterfactual is constructed using the outcomes of the teen mother's sisters who give births at an older age in order to dif-

[2]Influential examples include Card and Wise (1978), Hofferth and Moore (1979), Moore and Waite (1977), Mott and Marsiglio (1985), and Upchurch and McCarthy (1990).

[3]See Barnow, Cain, and Goldberger (1980) for details concerning this method.

ference out common family heterogeneity. Geronimus and Korenman (1992, 1993) examined three nationally representative samples, the National Longitudinal Survey Young Women's Sample (NLSYW), the NLSY-79, and the Panel Study of Income Dynamics (PSID), compared the "with-family" estimates with the traditional cross-sectional estimates and concluded that in two out of three samples, sister comparisons suggest the bias due to family background heterogeneity is important, therefore, much of the cross-sectional correlation of teenage childbearing and poor outcomes is not causal. This identification strategy makes use of the families with siblings and those families already have relatively larger family sizes. Family size is not randomly assigned and may be endogenous to other characteristics that are also correlated with teen motherhood. Moreover, the wealth of the parents and the way they invested among children may change over time. All of these factors may challenge the external validity of the causal interpretation drawn from the estimates in this kind of study.

Looking for exogenous variation that shifts the endogenous variable to achieve "clean" identification of the causal parameters lies in the center of the toolkit of modern applied economists. The term "natural experiment" was coined because, ideally and if correctly implemented, this method works like a controlled randomized experiment. Research in this category identifies the causal effects through sources of naturally occurring variations in teen fertility and uses these sources as instrumental variables in estimation.[4] When responses to treatment vary, as pointed out

[4]Some studies explicitly model the joint determination of fertility timing and maternal outcomes in a simultaneous discrete choice framework-examples include Lundberg and Plotnick (1989) and Ribar (1992, 1994)-but nonparametric identification of such models still requires exclusion restrictions (instrumental variables).

by Heckman (1997), different choices of instrumental variables identify very different local causal parameters and direct comparison of estimates obtained through different instruments often leads to meaningless conclusions. Bronars and Grogger (1993) identified the effects of adolescent motherhood by comparing the outcomes of teenage mothers who bore twins at their first birth with those of teen mothers who had a single child. Giving birth to twins is viewed as an exogenous shock to aid the identification and this "twin-first" methodology measures the local average treatment effect of an extra child that results from having twins at the first birth, without suffering from selection bias. Nevertheless, it is worthwhile to note that this effect is somewhat different from what we want to identify in this chapter, namely, the effect on educational attainment of having at least one child as a teenager relative to having no children at all before age 18. These authors did argue that this effect should be at least as large as their "twin-first" estimate.

In another innovative paper, Hotz, McElroy, and Sanders (2005) exploited an alternative type of fertility shock: some women who become pregnant experience a miscarriage and thus do not end up with a child. Using women who miscarry before 18 years old as a comparison group for the adolescent mothers and based on several assumptions[5] that validate their IV estimator, they found that "little of their disadvantages would be changed just by getting teen mothers to delay their childbearing into adulthood". Their findings suggested that teenagers who become pregnant, but

[5]Specifically, they assume that the distributions of latent pregnancy resolution types are orthogonal to the miscarriage dummy variable and that the direct effects of miscarriage and abortion on maternal outcomes are identical. In another paper, Hotz, Mullin, and Sanders (1997) show that nonparametric bounds can be constructed for the causal effects of teenage childbearing on mother's outcomes when these assumptions fail.

whose first birth was delayed by a miscarriage, do not have systematically better outcomes than their peers who carry their babies to term. By focusing on the group of women who experience at least one pregnancy as teenagers, Hotz, McElroy, and Sanders' (2005) parameter of interest is quite different from the rest of the literature; the local average effect they identify is the causal effect of having at least one child before age 18 for teen mothers compared to the counterfactual of experiencing pregnancy but not having any child as teens, thus, the estimates they obtain are not directly comparable to the estimates in the other papers in the literature. A recent study by Ashcraft and Lang (2006) challenges the findings of Hotz, McElroy, and Sanders (2005), arguing that miscarriage is only a valid instrument for the absence of birth in the absence of abortion. When abortion is an option, teenagers who miscarry are less likely to be girls who would otherwise abort their pregnancy than are teenagers who either abort or carry the child to term and since teenagers who have abortions on average come from more favored backgrounds than those who do not, girls who miscarry are not a random sample of pregnant teenagers but are, instead, drawn from more disadvantaged backgrounds. Thus, the IV estimator underestimates the true costs of teenage childbearing.

One common feature of the previous literature entails the imposition of strong parametric restrictions. The specific functional form and distributional assumptions about the error terms permit data from all observations to be smoothed into one estimator. The validity of this estimator becomes suspect when the smoothing function operates over observations with very different characteristics, in other words, when we compare the "un-comparables". The recent development of nonparametric estimators, such as the kernel matching estimator we are apply, enables researchers to estimate conditional moments without imposing any functional form assumptions. The identification of the mean effect of adolescent parenthood critically hinges on the consistent estimation of the conterfactual mean, i.e. the mean educational outcome of teenage mothers had they delayed their childbearing age. This chapter estimates this counterfactual nonparametrically using a matching estimator and compares it with the factual mean of teenage mother's schooling outcomes to identify the causal parameter. Matching allows us to utilize a larger sample size relative to the "natural experiment" type studies which need to focus on some narrowly defined subpopulation in order to make use of the exogenous variation.[6] Levine and Painter (2003) were the first to exploit matching methods in this topic. Their nearest neighbor matching estimates based on a sample from National Education Longitudinal Survey (NELS) reinforce the early findings that a substantial portion of the cross-sectional relation between teen childbearing and high school completion is due to pre-existing disadvantages of the adolescent mothers, not the childbirth itself. According to the Monte Carlo study of Frölich (2004), nearest neighbor propensity score matching estimator performs significantly worse in finite samples relative to other versions of matching estimators. In addition, the NELS data only contain individuals who are still in school at eighth grade. As such, a more careful implementation of different types of matching algorithms on a more general sample is in order.

[6]For example, in Hotz, McElroy, and Sanders (2005) there are only 68 out of 980 women in their sample who ever experienced a miscarriage.

2.3 Parameter of Interest and Matching Estimator

To define the causal effect of interest in this chapter, let Y_1 denotes a schooling outcome, for example, high school completion, when a young woman is considered to be teen mom, i.e. gives birth before she is 18 years old. Y_0 is the outcome if she has not given birth by the age of 18. Let D be the indicator which equals 1 if this woman is actually a teenage mother and 0 otherwise. The causal effect we are interested in is the average impact of having a birth as teenager versus not being a teen mom, where not being a teen mom includes being a mother later and not being a mother at all,[7] on the schooling outcomes for the subpopulation of adolescent mothers.[8] It is worth noting that our treatment parameter combines a number of treatments that can be examined separately, such as becoming a teen mother at age 14, becoming a teen mother at age 15 and so on. In terms of notation, the parameter we are identifying and estimating is

$$\Delta^{\tau\tau} = E(Y_1 - Y_0 \mid X, D = 1), \tag{2.1}$$

where X is an individual and family characteristic vector. In the program evaluation literature, this parameter is termed "treatment on the treated". It can not be identified directly from the data because the counterfactual mean, $E(Y_0 \mid X, D = 1)$, is missing. Some form of exogeneity of the decision regarding whether to be a teen

[7]Teenage childbearing in this chapter refers to giving birth before the 18th birthday. This early fertility measure ensures that most teenage births as we define them happen before high school completion.

[8]Other interesting mean causal parameters, though not useful in this context, can be found in Heckman and Robb (1985), and Heckman, LaLonde, and Smith (1999).

mom or not has to be assumed to identify $E(Y_0 \mid X, D = 1)$. With the wealth of the background covariates and cognitive ability measures available in the NLSY-79 in hand, our key identification strategy is formalized as the *Conditional Independence Assumption* (CIA), [9]

$$(Y_0, Y_1) \perp\!\!\!\perp D \mid X. \tag{2.2}$$

This assumption also underlies the traditional regression analysis, Barnow, Cain and Goldberger (1971) described a regression-based version of this condition as "selection-on-observables". However, matching and regression methods differ in that regression-based methods also require that assumption (2.2) holds linearly conditioning on X. The selection-on-observables assumption enables us to construct the hypothetical untreated state for teen mothers from those women who have not given births by age 18 conditional on all the variables that affect both the potential schooling outcomes and the probability of being a teen mother. All we have to do is to match each adolescent mother with the counterfactual schooling outcome constructed from comparison individuals who have similar observed characteristics according to some weighting algorithm. The difference between the teen mother and matched comparison outcome after integrating out the characteristic distribution equals the causal effect of teen fertility. Because we are only interested in the average effect of "treatment on the treated", we can weaken assumption (2.2) to

$$Y_0 \perp\!\!\!\perp D \mid X. \tag{2.3}$$

[9] This assumption is also known as ignorable treatment assignment assumption or unconfoundedness, see Rosenbaum and Rubin (1983), and Imbens (2004).

Furthermore, as suggested by Heckman, Ichimura, Smith, and Todd (1997), the full independence assumption can be relaxed to *Mean Independence Assumption* to identify our parameter of interest,

$$E(Y_0 \mid X, D = 1) = E(Y_0 \mid X, D = 0). \tag{2.4}$$

Through weighting comparison group data to equate the distributions of observables of the teen mothers and the comparison observations, matching methods are able to recover the mean causal effect correcting for selection based on the observed confounding variables. Another advantage of matching methods over running regressions lie in its ability to attend to the support issue. Using regression-based models, counterfactuals for teen mothers whose characteristics lie outside the common support region are derived solely from projections based on the specific functional form. Matching methods, although they do not solve the support problem directly, give us insights into the issue that regression-based mothods are unable to provide. Matching estimators directly put restrictions on the joint distribution of treatment and observed covariates to aid the identification assumption (2.2),

$$0 < Pr(D = 1 \mid X) < 1 \quad \text{for all } X. \tag{2.5}$$

That is, the positive support region for the observable distributions should be identical for teen mothers and their counterparts. Again, if the parameter of interest is "treatment on the treated", this condition can be reduced to

$$Pr(D = 1 \mid X) < 1. \tag{2.6}$$

In practice, running nonparametric regressions on high dimensional X is quite difficult due to sizable finite sample bias. When the dimension of the covariates increases, it poses a major challenge to the convergence rate of the estimator; in nonparametric econometrics, this is termed the "curse of dimensionality". For example, the strategy of building cells and matching units with exactly the same value of multi-dimensional X may fail if X takes on too many distinct values. To avoid the need to match observations on the values of all variables exactly, we must rely on inexact matching providing that we make our matches more exact as the sample size gets larger. Propensity score matching is one type of inexact matching where the propensity score is a young woman's conditional probability of becoming a teen mother. What makes the propensity score method feasible is a variation (Rosenbaum and Rubin, 1983) of assumption (2.3)

$$Y_0 \perp\!\!\!\perp D \mid Pr(D = 1 \mid X). \tag{2.7}$$

Matching estimators based on the propensity score reduce the dimension of conditioning set from all the covariates to a scalar, and adjustment for the propensity score suffices to remove all biases originating from systematic differences of the covariate distributions between treatment and comparison observations. Strictly speaking, propensity score matching still suffers from the "curse of dimensionality" because it requires estimation of $E(D \mid X)$. Practically, the propensity score is often estimated parametrically, through logit or probit models and adjusted according to some kind of specification or balancing tests. Under the identification assumptions (2.6) and (2.7), our matching estimator takes the general form

$$\hat{\Delta^{\tau\tau}} = \frac{1}{N_1} \sum_{i \in \{D=1\}} [Y_{i1} - \hat{E}(Y_{i0} \mid \hat{P}(X_i), D_i = 1)], \tag{2.8}$$

where

$$\hat{E}(Y_{i0} \mid \hat{P}(X_i), D_i = 1) = \sum_{j \in \{D=0\}} W(i,j) Y_{j0}, \tag{2.9}$$

where N_1 denotes the number of teen mothers in the sample, $\hat{P}(X_i)$ is the estimated propensity score, i represents teen mothers and j denotes women in the comparison group. The weights $W(i,j)$ determine the subset of comparison members who are matched with teen mom i and it also differentiates various matching estimators.[10] Monte Carlo evidence provided by Frölich (2004) suggests that kernel type matching estimators outperform nearest neighbor matching estimators and inverse propensity score weighting estimators on a number of circumstances. Kernel matching estimators assign non-zero weight to several, or even all, comparison observations to construct counterfactual for each treatment individual. The weight is given by

$$W(i,j) = \frac{K[\frac{\hat{P}(X_i) - \hat{P}(X_j)}{a_n}]}{\sum_{k \in \{D=0\}} K[\frac{\hat{P}(X_i) - \hat{P}(X_k)}{a_n}]}, \tag{2.10}$$

[10]See Heckman, Ichimura, and Todd (1998) and Smith and Todd (2005a) for discussions of various versions of matching estimators and their statistical properties.

where $K(\cdot)$ is the kernel function and a_n is a bandwidth parameter. The kernel function determines which part of the comparison population will participate in the formation of the estimated counterfactual for person i. All the different matching estimators are consistent assuming selection-on-observables holds.

2.4 The Data

Our analysis is based on female samples from the 1979 cohort of the National Longitudinal Survey of Youth (NLSY-79). This is a national sample of Black, Hispanic, and White men and women who were 14 to 21 years old as of December 31, 1978. The initial surveys were conducted in 1979 and respondents have been re-interviewed every year since then.[11] We use the data through the year 2002 follow-up. The female respondents were asked a wide range of questions about family background, home environment, fertility history, marital arrangements, educational attainment and welfare status. This richness of covariates makes the sample ideal for our study and makes the assumption of selection-on-observables plausible. Four independent probability samples comprise the original NLSY-79: a random cross section of the population, a supplemental oversample of civilian Hispanics and Blacks, a supplemental oversample of economically disadvantaged Whites, and a sample drawn from members of the military. Most of the females from the military sample were dropped out of the interviews after 1984 so we do not have the schooling outcome measures for them at older ages; thus, we drop them in our empirical

[11]The data of NLSY-79 were collected yearly from 1979 to 1994, and biennially from 1996 to the present.

analysis. Furthermore, all of the economically disadvantaged Whites have not been interviewed since 1990 and we believe that the criteria used to select women into this sample is not a very reliable way of identifying a representative sample of disadvantaged White women,[12] we also exclude this sample. These exclusions leave us with 4652 females in our analysis sample.

We construct complete fertility histories for all the individuals in the sample and divide them into teen mothers and non-teen-mothers according to whether a woman had a first birth before her 18th birthday. Table 2.1 reports the weighted summary statistics[13] for basic background characteristics of these women when they were 14 years old.[14] It is obvious from Table 2.1 that there are huge discrepancies in almost all the dimensions of observable background characteristics and test scores between teenage mothers and women who did not give birth before their 18th birthday. For example, teen mothers are more likely to be Black and Hispanic. The pre-existing socio-economic disadvantages of women who experience early fertility are clear. At age 14, the probability of living in an intact family is 0.15 lower for the teen mother sample; they also tend to live in households with more siblings and to have less-educated parents. Their families are less likely to have newspaper or magazine subscriptions when they were 14. As a traditional measure of cognitive skills or, ability, the last three rows of Table 2.1 present standard scores on three of the ten-part Armed Services Vocational Aptitude Battery (ASVAB) sections.[15]

[12] See Hotz, McElroy, and Sanders (2005) for more on this point.

[13] The weights account for the original design of the sample drawn in this study and differential probabilities of completing the base year interview.

[14] The Armed Services Vocational Aptitude Battery (ASVAB) test were administered in the summer and fall of 1980.

[15] Another widely used measure of cognitive skills, the Armed Forces Qualification Test (AFQT)

They are scores for “Arithmetic Reasoning”, “Word Knowledge”, and “Mathematics Knowledge” and just like other covariates presented in the table, the more favorable means all appear in the comparison sample. On average, women who did not give birth before 18 score around 8 points, or one standard deviation higher than teenage mothers in the ASVAB.

Table 2.2 shows the same summary statistics for the other two comparison groups we construct. The Birth18 sample consists of all females in NLSY-79 who experience their first pregnancy between their 18th and 19th birthdays and the pregnancy ends with a live birth, the Birth19 sample contains those women who become pregnant for the first time between 19 and 20 years old and the pregnancy is resolved through live birth. These comparison females are more like the teen mothers with regard to their fertility timing decisions. The gap in family background measures and cognitive ability measures between the teen mothers and these two samples is much smaller than the discrepancies in Table 2.1. For example, the difference in the probability of living in an intact family at age 14 shrinks to 5-8 percentage points from the 15 percentage point gap between the teen mother sample and the whole comparison sample showed in Table 2.1. Intuitively, it seems likely that using the comparison groups in Table 2.2 could lead to less selection bias than using the whole sample, although there is evidence from the observable family background variables and ability test scores that residual selection bias is still concern in both samples. This consideration leads us to separately utilize the Birth18

score, is a weighted sum of four out of 10 scores of ASVAB sections related to literacy and basic mathematic knowledge.

and the Birth19 samples in addition to the whole sample in most of our empirical analysis. We must be cautious about the nature of the treatment parameters we identify when the composition of the comparison group women changes. Although still economically interesting and policy relevent, using the Birth18 and the Birth19 comparison samples shifts the counterfactual state from not having a child before the age of 18 to delaying childbearing until either age 18 or age 19.

2.5 Probit Estimates of the Effects of Teenage Childbearing

We first present probit estimates of the causal effects of teenage childbearing on maternal high school completion and college attendance. The schooling outcome variables we are interested in are high school completion by 21 years old and college attendance by age 25. The high school completion variable reports whether the person has finished at least 12 years of education by age 21. Previous research like that in Cameron and Heckman (1993) and Tyler, Murnane, and Willett (2000) suggested different labor market rewards associated with high school graduation and General Educational Development (GED) recipiency. Thus, a GED completion outcome variable by the age of 21 is also evaluated.

Tables 2.3 and 2.4 show the coefficients on teenage motherhood binary variable in univariate probit models for the three schooling outcomes estimated from the three comparison samples: the whole sample and the Birth18 and the Birth19 samples. Numbers inside the brackets are average derivatives for the teen mother sample. Focusing first on Table 2.3, where the full sample estimates are displayed, the raw educational attainment gaps, as shown by the average marginal effects of the probit model without covariates, are 0.47 for high school completion and 0.38 for college attendance. Teenage mothers have a higher GED completion rate by age 21 than their peers who haven't given birth at 18, by 13 percentage points. These are very large gaps considering that the high school completion and college attendance rates of the full sample are 0.81 and 0.41, respectively. We progressively add covariates into the regression as we move from column two to column four. Findings in the second column indicates that inclusion of the family background controls reduces the size of the estimated marginal effects. For example, the difference in the high school completion by age 21 is reduced to 38 percentage points and the discrepancy in the college attendance decreases to 22 percentage points. This is consistent with the results documented in all of the cross-sectional studies of teen fertility effects.

A major advantage of the NLSY-79 data is the existence of controls for the individual's congitive abilities. These ability controls are based on the ASVAB and O'Neill (1990), Blackburn and Neumark (1992), and Neal and Johnson (1996) have used these ASVAB scores to control for otherwise unobserved differences in ability. We adjust the 10 ASVAB scores for age and schooling by taking the residuals of linear regressions of raw standard scores on individuals age, school enrollment status at the test date and highest grade completed at the test date[16] and use the first two principal components of the adjusted ASVAB scores as additional control variables as suggested by Heckman (1995) and Cawley, Heckman, and Vytlacil (2001). We

[16]For more detailed information about the adjustment procedure, see Hansen, Heckman, and Mullen (2004).

describe the adjustment procedure for ASVAB test scores in Table A1. Adding these cognitive ability controls to the probit in the third column, we see that the average marginal effects for high school completion and college attendence drop by 15 percentage points and 32 percentage points. At the same time, the pseudo R^2s of the probits for these two schooling outcomes increase from 0.265 and 0.206 to 0.392 and 0.317. The only exception is that the estimated teenage childbearing effect on GED completion doesn't change much after we condition on the ability measures. These achievement test scores have strong predictive power for the schooling outcome in addition to the part accounted for by the family background covariates. From the first to the second column, the pseudo R^2s for the regressions are doubled for the high school related outcomes and increase by four times in the case of college attandence, the background characteristics are powerful predictors of educational attainment and lead to decreases in the estimated teenage childbearing effects. Nevertheless, these effects are still substantial. Moreover, the positive GED impact is in agreement with previous research.[17] These estimates reinforce the idea that teen mothers try to avoid receiving less education by substituting a GED for a high school diploma.

Table 2.4 presents the probit estimates of the effects of teenage childbearing using the Birth18 and the Birth19 samples as comparison groups. The decline of the estimated effects on high school outcomes is more impressive when we use women whose first pregnancy is between their 18th and 19th birthdays and the pregnancy ends with a live birth as comparisons. With the full set of controls, the difference between the estimate in Table 2.3 and the upper panel of Table 2.4 is 14 percentage

[17] See Hotz, McElroy, and Sanders (2005).

points, a 43% decrease. Changing the comparison women to the Birth19 sample, the probit cofficient on the teen mother variable stays almost the same as the one we get using the full sample. The estimated effects on the GED outcome variable are still quite robust to the inclusion of background and ability covariates although they decrease in magnitude when we change the comparison sample. In particular, when the Birth19 comparison sample is used to estimate the counterfactual, the estimated effect on GED outcome equals 6.9 percentage points, representing more than a 40 percent reduction from the whole sample estimate. The substantively and statistically significant estimated effects on high school completion in Table 2.3 and Table 2.4 are suggestive that at least some portion of the observed cross-sectional schooling gap is causally related to early childbearing.

The estimated effect of teen motherhood on college attendance is reduced from 0.15 to 0.011 and loses its statistical significance after we replace the full sample comparison with the Birth18 sample, and it becomes 0.046 and marginally significant if we use the Birth19 sample. Unlike the high school graduation decision, which coincides with the timing of giving birth for most teenage mothers, the decision to go to college for most females in our sample comes much later, so that teenage childbearing is less likely to interfere with this outcome. Thus, it is not surprising to see much smaller estimated effects once we switch to the more comparable comparison groups.

Comparing Table 2.3 and Table 2.4, including additional covariates in the probit regressions makes a much larger difference to the marginal effects in the whole sample results in Table 2.3 than in the comparison samples in Table 2.4. For example, the effect on high school completion when the Birth18 comparison sample is used is only reduced by 4 percentage points after we add all the controls, while in the case of the full sample, this estimate is reduced by about 15 percentage points after the same set of controls is added. The degree of selection based on observable traits has been considerably reduced by only using women who have similar fertility timing to the teenage mothers as comparison samples.

2.6 Matching Estimates of the Effects of Teenage Childbearing

Before proceeding to our matching estimates, it is useful to discuss the estimation of the propensity score and examine the support condition in our samples. The conditional probability of giving birth as a teen is a function of all the control variables X that makes the condition (2.3) satisfied. The selection-on-observables assumption conditioning on the propensity score is equivalent to assumption (2.3). In fact, as pointed out by Rosenbaum and Rubin (1993),[18] controlling for a balancing score, $b(X)$,which is defined as a function of X that is finer than the propensity score $P(X)$ in the sense that $b(X) = f(P(X))$ for some function f, is sufficient to guarantee the independence of potential outcomes and treatment status as long as assumption (2.3) holds. Although an infinite number of balancing scores exist, none of them is known in practice and a misspecified propensity score model may lead to inconsistency of the matching estimator. Hence, figuring out a parametric specification is an important practical concern.

A heuristic approach for testing the misspecification of the propensity score

[18] See Theorems 1,2, and 3 of Rosenbaum and Rubin, 1983.

model is adopted in this paper. A simple statistical property of the balancing score justifies this test which is commonly referred to as balancing test in the program evaluation literature,

$$X \perp\!\!\!\perp D \mid b(X). \tag{2.11}$$

In another words, we examine whether the womens' observable characteristics are independent of the teenage childbearing variable conditional on the estimated propensity score to detect misspecification of the propensity score equation. A number of empirical strategies have been applied in the literature to conduct the balancing test. According to the Monte Carlo analysis by Smith and Zhang (2007), the regression-based balancing test suggested in Smith and Todd (2005b) generally out-performs other versions of the tests and so is used in this chapter. The regression model directly tests the balancing property by regressing each of the conditioning variables in turn on a polynomial in the propensity score and the same polynomial interacted with the treatment indicator. Specifically, this approach estimates the regression[19]

$$\begin{aligned} x_k = &\beta_0 + \beta_1 \hat{P}(X) + \beta_2 \hat{P}(X)^2 + \\ &\beta_3 \hat{P}(X)^3 + \alpha_0 D + \alpha_1 D\hat{P}(X) + \alpha_2 D\hat{P}(X)^2 + \\ &\alpha_3 D\hat{P}(X)^3 + \epsilon_k, \end{aligned} \tag{2.12}$$

and then performs an F test of the joint null that all of the coefficients on terms involving D equal zero. This test directly gets at the condition that D provides no

[19] The requirement that the user must choose an order for the polynomial in the regression test represents the primary unattractive feature of the test; we employ a cubic in this chapter.

information about each x_k conditional on the estimated propensity score. If any of the these K F-statistics exceeds the conventional 5% critical value, we add to the propensity score specification the higher order and interaction terms of those imbalanced variables and run these regressions all over again.

In total, four higher order and interaction terms[20] are added to the basic linear probit specification in order to balance the observable controls in the full sample. For the Birth18 and Birth19 comparison samples, two and five additional interaction terms[21] are needed to make the covariate distributions balance.

The estimated propensity score can help us to examine the common support condition in the sample. Remember that the support assumption that matching estimators require is (2.6) for the X, with matching performed on the balancing score $b(X)$, this condition can be rewritten as

$$Pr(D = 1 \mid b(X)) < 1 \quad \text{for all } b(X). \tag{2.13}$$

Thus, looking at the cumulative distribution functions of the estimated propensity scores for the teenage moms and the comparison group can help us learn to what extent condition (2.13) holds in our sample.

[20] These additional terms are the squared first principal component of the ASVAB test scores, an interaction between father's education and Black, an interaction of foreign language speaking family dummy and the first principal component of the ASVAB test scores, and the interaction of the frequent religious activity dummy and the first principal component of ASVAB test scores.

[21] For the Birth18 sample, these additional terms are the squared second principal component of the ASVAB test scores and interaction between the number of siblings and the first principal component of the ASVAB test scores. For the Birth19 sample, they are the squared first principal component of the ASVAB test scores, the squared second principal component of the ASVAB test scores, an interaction between Black and the second principal component of the ASVAB test scores, the product of the frequent religious activity dummy and the second principal component of the ASVAB test scores, and an interaction between intact family structure at age 14 and the second principal component of ASVAB test scores.

Figures 2.1, 2.2, and 2.3 illustrates the overlap of positive support of the estimated propensity score distributions between women who give birth as teenagers and women in the comparison samples. These are the histograms of the estimated scores for both teen mothers and their peers who have not given birth by the age of 18. Figure 2.1 plots the histogram for the full sample and Figures 2.2 and 2.3 are for the Birth18 and the Birth19 comparison samples. In all the three graphs, the darker colored bars correspond to the $D = 0$ group. The X-axis defines the intervals of the estimated score and the Y-axis is the frequency, for each group, with an estimated propensity score in the corresponding interval. The message conveyed by these graphs is that condition (2.13) is well satisfied in our sample.

For the whole sample, 15 percent of the comparison group sits below the first percentile of the teen mothers; meanwhile, around 10 percent of the treatment group lies above the 99^{th} percentile of the comparison group. Indeed, the sample averages of the estimated propensity score between the two groups are quite close, with teen mother group being 0.23 and that for the comparison sample being 0.11. The common support is even thicker when we use our alternative comparison samples. With the Birth18 comparison sample, about 3 percent of the comparisons lie below the first percentile of the treatment group and less than 4 percent of the young mothers are above the 99^{th} percentile of the comparison group. The corresponding percentages for the Birth19 comparison group are 7 and less than 1.

Running nonparametric kernel regressions to estimate the missing counterfactual requires us to select a kernel function $K(\cdot)$ and its associated bandwidth value

a_n.[22] This task is carried out by leave-one-out cross validation following Racine and Li (2003). Denote a combination of a certain kernel function $K(\cdot)$ and a bandwidth parameter a_n as θ. The leave-one-out validation mechanism chooses θ as

$$\theta^{CV} = \underset{\theta}{argmin} \sum_{j \in \{D=0\}} [Y_j - \hat{Y_{-j}}]^2. \tag{2.14}$$

Cross validation uses comparison observations to determine which of the competing combination of kernel function and bandwidth value best fit the data. It takes one comparison member out of the sample and uses the remaining $N_0 - 1$ observations to form an estimate of Y_j. The out-of-sample prediction error is thus $Y_j - \hat{Y_{-j}}$ and by repeating the process for every remaining observation in the comparison group, the mean squared error serves as a reasonable guide for selecting the combination of a kernel function and a bandwidth value. Table A2 reports the results of this validation procedure for each outcome variable we evaluate and for each of the three comparison samples we use. For eight of the nines cases, the Epanechnikov kernel estimator outperforms the biweight and the Gaussian kernel estimators across a wide range of bandwidth values. This is consistent with the cross-validation findings of

[22]The three alternative kernel functions we consider are (1) the biweight kernel, where

$$K(\psi) = \begin{cases} \frac{15}{16}(\psi^2 - 1)^2 & \text{if } \mid \psi \mid < 1; \\ 0 & \text{otherwise.} \end{cases}$$

(2) Gaussian kernel where $K(\psi)$ is just a standard normal density function.
(3) Epanechnikov kernel, where

$$K(\psi) = \begin{cases} \frac{3}{4}(1 - \psi^2) & \text{if } \mid \psi \mid < 1; \\ 0 & \text{otherwise.} \end{cases}$$

The bandwidth search grid is 0.01, 0.02,...,1.

Black and Smith (2004).[23]

Kernel matching estimates of the impact of teenage childbearing on schooling outcomes appear in Table 2.5. These estimates are based on the leave-one-out cross validated kernel functions and bandwidth values described in the previous paragraph. Three panels from the top to the bottom present estimates using three different comparison samples: the full sample, the Birth18 sample and the Birth19 sample. Standard errors are presented in parenthesis below the estimates; they come from bootstrapping with 2000 replications.[24] Accompanying each estimate, we also indicate the number of teen mothers whose estimated propensity scores exceed the range bounded by the minimum and maximum estimated scores of the comparison members as additional examination of the common support problem. The low values of these numbers (ranging from 1 to 6) provide further evidence that condition (2.13) holds.

For the whole sample, matching estimates of the effect of a teen birth on high school outcomes are negative 24.8 percentage points for high school graduation and positive 7.3 percentage points for GED completion, both statistically significant at conventional levels. These estimated effects are all smaller than the univariate probit estimates from Table 2.3, a 22% reduction for the effect on high school completion and a 38% decrease for the impact on GED completion by age 21. Substantively, these point estimates provide strong reason for concern about teenage childbearing effect estimates based on univariate probit regressions that control only linearly for

[23] Black and Smith (2004) also relied on this cross validation mechanism to guide their choices of different versions of matching estimators. Their findings reinforced the results of Frölich (2004) that nearest neighbors estimators usually perform worse than kernel estimators.

[24] The propensity score is re-estimated for each replication.

covariates. However, they support the view that there is still a significant negative impact on schooling outcomes due to teenage motherhood even when all of the available individual and family factors have been taken into account semi-parametrically. The estimates tell a similar story for college attendance by age 25. Our matching estimate stands at -0.104 while the corresponding probit estimate is -0.148. The smaller matching estimates provide evidence for the bias in traditional regression models that control only linearly for covariates but still support the conclusion that teenage fertility does lower maternal educational attainment. Put simply, there are substantively and statistically significant schooling costs to women who become mothers before they turn 18.

A similar pattern of differences between the univariate probit estimates and the matching estimates emerges when we switch to the Birth18 and the Birth19 comparison samples. For all the schooling outcomes we evaluate, the kernel matching estimates are always smaller in magnitude than the regression estimates and accompanied by generally larger standard errors. For example, using the Birth18 comparison group, teen mothers have a higher probability of getting a GED, with the difference being 7.1 percentage points according to the matching estimate, while the probit gives an estimate of 10.6 percentage points. There are no statistically significant differences in college attendance between teen mothers and their peers in the comparison samples. Taking these findings together, after controlling for family background and cognitive characteristics, a sizable negative impact of teen motherhood on high school completion still remains even when we define the counterfactual as experiencing pregnancy between 18 and 19 years old and having live birth. The lesson regarding GED completion is the same in spirit as for the probit esitmates.

In contrast to the full sample estimates, the effect of teen fertility on college attendance is substantially reduced when we change the comparison groups to the Birth18 and the Birth19 samples. This is consistent with our findings from the probit models. The -0.104 in the whole sample decreases to -0.025 for the Birth19 sample and further to -0.007 when comparing to women in the Birth18 group. Eliminating the potential selection bias arising from unobserved confounding factors affecting a woman's fertility timing realization makes a much larger difference for the effect of college attendance. Although we still need to emphasize the change in the nature of the parameter being estimated when different comparison samples are utilized, early childbearing has a far smaller impact on the mother's decision to go to college than on high school completion.

2.7 Sensitivity Analysis: the Role of Unobservables

Our estimated effects of teenage childbearing on maternal educational attainment put us at odds with most "natural experiment" type studies. For example, exploiting miscarriage as an instrumental variable, Hotz, McElroy, and Sanders (2005) suggested much smaller schooling effects of teenage childbearing. The only statistically significant effect in their paper is the 0.13 point estimate of the GED effect while their negative 11 percentage points high school diploma impact and 0.01 high school or GED impact do not attain statistical significance at the 5 percent level. In spite of the difference between our "treatment on the treated" parameter and their local average treatment effect parameter, we remain concerned that our empirical analysis does not control for all the possible covariates. And although our results suggest only a small degree of selection on observables when we use the Birth18 and the Birth19 comparison groups, it is still likely that a small amount of selection on unobservables could bias our estimates. In order to assess the sensitivity of the estimated teenage childbearing effects to the unobservables, we implement two sensitivity analyses in this section. Note, however, that what we do in this section does not test the identifying assumption of selection-on-observables; indeed, this assumption is intrinsically not testable because the data are uninformative about the distribution of untreated potential outcomes for the teen mother sample.

2.7.1 Bivariate Probit Model with Varying Correlation Coefficient

Following Altonji, Elder, and Taber (2005), we first seek help from an explicit econometric model of schooling decisions and adolescent childbearing. In particular, consider a system of equations describing the joint determination of educational attainment Y and teenage birth D:[25]

$$Y = 1(X'\beta + D\gamma + \epsilon > 0) \tag{2.15}$$

$$D = 1(X'\delta + \eta > 0)$$

[25] This kind of model is referred as a "multivariate probit model with structural shift" by Heckman (1978).

where $1(\cdot)$ is the indicator function, X still denotes the vector of family background characteristics and ability measures and ϵ and η are the unobservables in the model. We further assume, as is common in this kind of sample selection models, that $X \perp\!\!\!\perp (\epsilon, \eta)$ and (ϵ, η) is normally distributed with

$$\begin{pmatrix} \epsilon \\ \eta \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Since the coefficients and error variances in such models are only identified up to scale, we apply the standard normalization that $\sigma_\epsilon^2 = \sigma_\eta^2 = 1$. Given this normalization, the correlation coefficient between the unobserved factors, ρ, essentially determines the degree of selection on unobservables when we estimate the causal effect of teenage childbearing. The conditional independence assumption (2.2) implies that the correlation equals zero. Although maximum likelihood estimation of this bivariate probit model is easy to implement, nonparametric identification of the effect of teen fertility, γ, requires an exclusion restriction.[26] Hence, our first sensitivity check hinges on a restricted version of this model, where ρ is treated as an unidentified parameter.

The thought experiment estimates the bivariate probit model with various assumed values of ρ, the correlation between the error components in the outcome and participation equations. Tables 2.6 and 2.7 present the estimated teen fertility effects on schooling outcomes for various ρ values (0 to -0.5 for high school comple-

[26]Strictly speaking, this model can be identified parametrically from the functional form, but applied economists are typically skeptical of the stability of the estimates from such models in the absence of an exclusion restriction. Ribar (1994) estimated a similar model with three excluded variables: age at menarche, availability of obstetricians/gynecologists, and the local abortion rate.

tion and college attendance and 0 to 0.5 for the GED outcome). The full sample estimates are reported in Table 2.6. Setting the correlation of the error terms to zero reproduces the fourth column of Table 2.3, the average derivatives for the teen mother sample from the univariate probit where teen childbearing is treated as exogenous conditional on all the covariates. The raw gap in the probability of high school completion in the whole sample is 0.47, the estimated effect is -0.319 when $\rho = 0$. Changing the correlation coefficient to -0.1, the effect declines to -0.294 and further to -0.266 when we fix ρ at -0.3. Specifying teen childbearing as an endogenous determinant of completing high school weakens the estimated effect of teenage fertility. While at $\rho = -0.5$, the effect on high school completion shrinks to -8 percentage points but remains statistically significant. Its size is relatively small given the sample mean high school completion rate for the 574 teen mothers is 0.437. To wipe out the impact of teenage childbearing on high school completion and instead attribute the estimated effect entirely to selection on unobservables, the correlation between the unbserved factors in the outcome and participation equations has to be greater in absolute value than 0.5.[27] For the GED outcome, the marginal effect becomes both economically and statistically insignificant when we increase ρ over 0.4. Both the univariate probit and matching estimates suggests smaller effects on college attendance, especially in comparison to women whose fertility timing is closer to that of the teen mothers. For example, in Table 2.7, only a little bit of selection on the unobservable determinants of college attendance and teen fertility

[27] As noted in Altonji, Elder, and Taber (2005), one problem with this type of sensitivity analysis is the difficulty of judging reasonable magnitudes of the correlation coefficient without prior knowledge.

($\rho = -0.1$) can account for the estimated effect.

Table 2.7 tells the story that the effect of high school completion is less sensitive to the correlation coefficient. For example, even with $\rho = -0.5$, the estimate obtained using the Birth19 sample is -0.114 and remains highly significant. It is much larger compared to the full sample estimate when the correlation coefficient is set at the same value although they are comparable in size when the univariate probit model is analyzed. We see this as further support for the increased "comparability" when the Birth18 and the Birth19 females are taken as comparisons. It does not take too many increases in the value of ρ (about 0.2 for GED and -0.1 for college outcome) in the Birth19 sample to drive the already-small estimated impacts of teenage childbearing on GED completion and college attendance to substantive and statistical insignificance.

The variation in the estimated marginal effect of teen motherhood with the varying values of ρ captures the possibility of selection on unobservables. Generally, for the effect on the high school completion, we find that the correlation between the unobservables in the equation system (2.15) has to be quite high (over 0.5 in absolute value) to be able to explain away the whole impact. The less sensitive impact estimates generated with the help of the Birth18 and the Birth19 samples imply a smaller role for the unobservables, relative to the whole sample estimates.

2.7.2 Simulation-based Nonparametric Sensitivity Analysis

In this subsection, we follow Rosenbaum (1987) and Ichino, Mealli, and Nannicini (2007) and assess the robustness of the estimated teenage childbearing effects with respect to selection on unobservables through a simulation-based nonparametric analysis. The unobservables are assumed to be summarized by a binary variable in order to simplify the analysis, although similar techniques could be used assuming some other distribution for the unobservables. The central assumption of this analysis is that teen motherhood is not exogenous conditional on the set of observables X, thus,

$$P(D = 1 \mid Y_0, Y_1, X) \neq P(D = 1 \mid X). \tag{2.16}$$

Instead, we assume that the conditional independence assumption holds conditional on X and an unobserved binary covariate, U,

$$P(D = 1 \mid Y_0, Y_1, X, U) = P(D = 1 \mid X, U). \tag{2.17}$$

We impose the values of the parameters that characterize the distribution of U. Given these parameters, we then predict a value of the confounding factor for each teen mother and comparison group member and re-estimate the effects with the simulated U in the set of conditioning variables. By changing the assumptions about the distribution of U, we can assess the robustness of the matching estimates with respect to different hypotheses regarding the nature of the unobservables. Formally, consider our binary potential schooling outcome variable $Y_0, Y_1 \in \{0, 1\}$. Assuming that U is independent of X, we can characterize the distribution of U by specifying the parameters

$$P_{i,j} = P(U = 1 \mid D = i, Y = j, X) = P(U = 1 \mid D = i, Y = j) \tag{2.18}$$

with $i, j \in \{0, 1\}$, which gives the probability that $U = 1$ in each of the four groups defined by teen mother status and schooling outcomes.

Given meaningful values of the parameters $P_{i,j}$, this sensitivity analysis proceeds by attributing a value of U to every female, according to her belonging to one of the four groups defined by the values of i and j. We then treat U as another observable used to estimate the propensity score and to compute the matching estimates. In practice, we choose the values of the parameters $P_{i,j}$ to make the distribution of U similar to the empirical distribution of two observable binary covariates in the sample: the Black binary variable and a dummy variable indicating whether the individual's first principal component of the ASVAB test scores lies above the sample median. We choose these variables because of the significant gap in the fraction of black between teen mothers and their comparisons and the strong predictive power of the ASVAB test scores on schooling outcomes even conditional on other observables. In this case, the simulation exercise is able to reveal the extent to which our matching estimates are robust to deviations from the identifying assumption induced by the impossibility of observing factors similar to the Black variable and the ability measures. For example, the fraction of Blacks in the sample of teen mothers who complete high school by age 21 is 0.65, as shown in Table 2.8, we therefore set the mean of the simulated U equal to 0.65 when we randomly assign values of 0 or 1 to the observations in this sample. After fixing the values of the sensitivity parameters, we repeat the matching algorithm 1000 times and obtain an estimate of the effect of teenage childbearing, which is an average of the impacts over the distribution of the simulated U.[28]

During the implementation of this nonparametric sensitivity analysis, we also measure how the different configurations of $P_{i,j}$ chosen to simulate U translate into associations of U with Y_0 and D conditioning on X. More precisely, we estimate a logit model of $P(Y = 1 \mid D = 0, U, X)$ in every iteration and compute the effect of U on the relative probability of having a positive schooling outcome in the case of no teenage childbearing as the average estimated odds ratio of U

$$\Gamma = \frac{\frac{P(Y=1|D=0,U=1,X)}{P(Y=0|D=0,U=1,X)}}{\frac{P(Y=1|D=0,U=0,X)}{P(Y=0|D=0,U=0,X)}}. \tag{2.19}$$

Similarly, by estimating the logit model of $P(D = 1|U, X)$, the average odds ratio of U measures the effect of U on the relative probability of becoming a teen mother

$$\Lambda = \frac{\frac{P(D=1|U=1,X)}{P(D=0|U=1,X)}}{\frac{P(D=1|U=0,X)}{P(D=0|U=0,X)}}. \tag{2.20}$$

[28]The standard error of the matching estimate when U is included as an additional covariate is computed following Ichino, Mealli, and Nannicini (2007) as the squared root of

$$V = \frac{1}{m}\sum_{k=1}^{m} SE_k^2 + \frac{m+1}{m^2+m}\sum_{k=1}^{m}(\hat{\Delta_k^{\tau\tau}} - \bar{\hat{\Delta^{\tau\tau}}})^2$$

where k denotes the kth replication, m is the total number of replications, and $\hat{\Delta_k^{\tau\tau}}$ and SE_k^2 represent the matching estimate and the estimated variance at the kth replication and $\bar{\hat{\Delta^{\tau\tau}}}$ is the average of $\hat{\Delta_k^{\tau\tau}}$ over m replications.

Following Ichino, Mealli, and Nannicini (2007), we refer to Γ as the "outcome effect" and Λ as the "selection effect".

Results of the nonparametric sensitivity analysis are presented in Tables 2.8-2.10 for the three comparison samples. To facilitate comparisons between actual and simulated results, the first row of Tables 2.8-2.10 shows the baseline matching estimates obtained with no unobservable from Table 2.5. The second row presents the simulated estimates with a "neutral" unobservable with $P_{i,j} = 0.5$. By definition, this simulated unobservable has zero outcome effect and zero selection effect but, as we can see from the three tables, it is enough to slightly perturb the results. For example, in Table 2.9, including such a neutral unobservable changes our matching estimate from -0.127 to -0.123 for the high school completion outcome.

The next two rows of these tables show how the matching estimate changes when the binary unobservable factor U is calibrated to mimic the Black variable and the indicator variable for an above the median ability measure. Looking across these tables, the simulated unobservable plays an important role either in the propensity score estimation or in the outcome equation, but not both. For example, an unobservable imitating the empirical distribution of the ability measure binary variable in the full sample has a negative effect on the relative probability of being a teen mother with selection effect equal to 0.27 but has a much larger positive impact on high school completion in the case of no treatment with an outcome effect of 10.1, while the matching estimate differs by only three percentage points with respect to the baseline estimate obtained in the absence of unobserved confounding effects. The same pattern shows up in all three tables and none of the baseline matching estimates qualitatively change with the inclusion of the calibrated unobservables. Taken in conjunction, these simulations convey the robustness of the baseline matching estimates of the teenage childbearing effects on maternal schooling outcomes. These simulations also show that both the outcome and the selection effects of U must be strong in order to represent a threat to the significance of the estimated impacts.

2.8 Conclusion

In this chapter, we have examined the causal relationship between adolescent fertility and maternal schooling outcomes through the method of matching in the rich NLSY-79 data. Not only does the most basic regression-based econometric approach rely heavily on assumptions about the parametric form of the regression function, its ability to highlight the support issue, which may be important due to the self-selected nature of teen childbearing, is also limited. As a result, regression estimates are often questioned. The matching estimator avoids the problematic functional form assumption by estimating the counterfactual conditional mean nonparametrically. It also provides us with a better tool to attend to the common support condition through (2.13). The empirical results from kernel matching and the related sensitivity checks support four main findings.

First, in agreement with earlier arguments in the literature, a substantial portion of the correlation between low educational attainment and early motherhood is not causal. There is substantial selection based on observed background characteristics into the population of teenage mothers, but despite this selection the support condition holds fairly well in our sample. Restricting the comparison sample to females who share similar fertility timing increases the "comparability" of the treatment and comparison groups and, although it simultaneously changes the nature of the parameters, delivers more convincing estimates.

Second, over half of the raw educational gap still remains after controlling for an extensive set of covariates, indicating that being a teen mom implies hefty schooling costs, especially in terms of high school completion. Many women who give birth before 18 opt for GED as a substitute for formal high school level education. Teen fertility makes only a small difference for college attendance by age 25.

Third, the considerable difference between the univariate probit estimates and the kernel matching estimates we find in this chapter raises serious concerns over the applicability of running regressions in this context. Nonetheless, although quite different in magnitude, the negative schooling impacts of teenage childbearing do not go away when semi-parametric matching is applied. Therefore, the matching estimates still support the overall finding that teen motherhood itself contributes causally to the poor educational outcomes of teen mothers.

Fourth, this chapter implements both parametric and nonparametric sensitivity analyses to assess the role played by unobserved confounding factors. The parametric version tells the story that a substantial amount of correlation is required to make the negative effect of teen motherhood on educational attainment go away. A simulation-based nonparametric analysis sends the same general message, as the baseline kernel matching estimates are quite robust with regard to a wide range of different specifications of the simulated unobservables.

Table 1.1: Summary Statistics

Variable	Mean	S.D.
Education	13.392	2.402
Black	0.124	0.329
Hispanic	0.061	0.239
Ln(Real hourly wage)	2.556	0.592
Actual experience	7.253	4.763
Job tenure	2.913	3.356
Standardized *AFQT*	0.000	1.000

Source: Author's calculations from NLSY-79.

Note: 1. The sample consists of 4132 individuals with 48617 observations in the years from 1979 to 2000. See the Data section of the paper for details of sample construction.

Table 1.2: Job Separations and Total Number of Jobs Held During the First 2, 5, and 10 Years of Career

	2 Years		5 Years		10 Years	
	Mean	SD	Mean	SD	Mean	SD
Job separation: 0	0.384	0.486	0.103	0.304	0.032	0.177
Job separation: 1	0.340	0.474	0.190	0.392	0.083	0.277
Job separation: between 2 and 5	0.275	0.446	0.568	0.495	0.460	0.498
Job separation: between 5 and 10	0.001	0.030	0.131	0.337	0.312	0.464
Job separation: greater than 10	0	0	0.008	0.089	0.111	0.315
Job separation: Max	6		19		29	
Job separation: total	1.024	1.068	3.016	2.323	5.568	3.950
Jobs held: total	1.757	1.095	3.733	2.309	6.214	3.927
Number of Observations	4132		4132		4132	

Source: Author's calculation from NLSY-79.

Note: 1. All the estimates are weighted by the NLSY-79 sampling weights.

2. Job separation counts and total number of jobs held are obtained from the NLSY-79 work history file which reports the starting and ending dates for jobs held at the time of each interview and the same information for up to 5 jobs which began and ended since the last interview.

Table 1.3: Probit Marginal Effects of Standardized AFQT on Job Mobility

	(1)	(2)	(3)	(4)
Standardized *AFQT*	-0.036***	-0.032***	-0.025***	-0.012**
	(0.004)	(0.005)	(0.005)	(0.005)
Standardized *AFQT**Experience/10		0.026***	0.015**	0.014**
		(0.007)	(0.007)	(0.007)
Pseudo R^2	0.004	0.167	0.187	0.189
Number of Observations (Individuals)	48617 (4132)			

Source: Author's calculation from NLSY-79.

Notes: 1. All the probit marginal effects are means of the individual marginal effects.

2. The dependent variable is a dummy variable for at least one job separation during the year. Model (2) also includes experience/10 as independent variable. Model (3) also includes experience/10, black, hispanic, industry and occupation dummies, and year dummies as independent variables. Model (4) includes school enrollment status at the ASVAB test date and highest grade completed at the ASVAB test date as additional independent variables besides the ones in model (3).

3. The standard errors are in parentheses and are White/Huber standard errors accounting for potential correlation at the individual level. The standard errors of the marginal effects are derived through the delta-method. * signifies significance at the 10% level, ** at the 5% level and *** at the 1% level.

Table 1.4: The Effects of Schooling and Standardized AFQT on Wages

	OLS		IV	
	(1)	(2)	(3)	(4)
Education	0.066***	0.070***	0.067***	0.074***
	(0.016)	(0.016)	(0.016)	(0.016)
Standardized *AFQT*	0.076***	0.038***	0.072***	0.018
	(0.010)	(0.010)	(0.010)	(0.011)
Education* Experience/10	-0.003	-0.034***	-0.050***	-0.065***
	(0.012)	(0.013)	(0.018)	(0.018)
Standardized *AFQT** Experience/10		0.052***		0.073***
		(0.011)		(0.011)
R-squared	0.307	0.308	0.302	0.303
Number of Observations (Individuals)	48617 (4132)			

Source: Author's calculation from NLSY-79.

Note: 1. All the estimates are weighted by the sampling weights provided by the NLSY-79.
2. The dependent variable is the natural log of the respondent's hourly wage. All the regressions in the table contain a cubic in experience, black, hispanic, industry and occupation affiliation, year effects, education interacted with year effects, interactions between black and year effects, and between hispanic and year effects.
3. The standard errors are in parentheses and are White/Huber standard errors accounting for potential correlation at the individual level. * signifies significance at the 10% level, ** at the 5% level and *** at the 1% level.

Table 1.5: The Relationship Among Wages, Standardized AFQT, Job Tenure, and Frequency of Job Separations

	(1)	(2)
Standardized *AFQT*	0.036***	0.051***
	(0.010)	(0.013)
Standardized *AFQT** Experience/10	0.048***	0.049***
	(0.014)	(0.014)
Standardized *AFQT** Tenure/10	0.001	0.021**
	(0.015)	(0.011)
Frequency of job separations		-0.002***
		(0.001)
Standardized *AFQT**Frequency of job separations		-0.001**
		(0.000)
Standardized *AFQT**Tenure/10*Frequency of job separations		-0.013**
		(0.005)
R-squared	0.316	0.427
Number of Observations (Individuals)	48617 (4132)	

Source: Author's calculation from NLSY-79.

Note: 1. All the estimates are weighted by the sampling weights provided by the NLSY-79.

2. The dependent variable is the natural log of the respondent's hourly wage. All regressions in the table contain a cubic in experience, a cubic in job tenure, black, hispanic, industry and occupation affiliation, year effects, education interacted with experience, education interacted with year effects, interactions between black and year effects, and between hispanic and year effects.

3. The standard errors are in parentheses and are White/Huber standard errors accounting for potential correlation at the individual level. * signifies significance at the 10% level, ** at the 5% level and *** at the 1% level.

Table 1.6: The Effects of Schooling and Standardized AFQT on Wages under Asymmetric Employer Learning

	(1)	(2)
Education	0.069***	0.094***
	(0.016)	(0.017)
Standardized *AFQT*	0.036***	0.033**
	(0.010)	(0.013)
Education* Experience/10	-0.035***	-0.034***
	(0.013)	(0.013)
Standardized *AFQT** Experience/10	0.056***	0.056***
	(0.015)	(0.015)
Education* Tenure/10	-0.009	-0.008*
	(0.006)	(0.005)
Standardized *AFQT** Tenure/10	0.015	0.016**
	(0.010)	(0.008)
Freqency of job separations		-0.016***
		(0.004)
Education*Frequency of job separations		-0.001***
		(0.000)
Standardized *AFQT**Frequency of job separations		-0.002**
		(0.001)
R-squared	0.326	0.438
Number of Observations (Individuals)	48617 (4132)	

Source: Author's calculation from NLSY-79.

Note: 1. All the estimates are weighted by the sampling weights provided by the NLSY-79.

2. The dependent variable is the natural log of the respondent's hourly wage. All regressions in the table contain a cubic in experience, a cubic in job tenure, black, hispanic, industry and occupation affiliation, year effects, education interacted with year effects, interactions between black and year effects, and between hispanic and year effects.

3. The standard errors are in parentheses and are White/Huber standard errors accounting for potential correlation at the individual level. * signifies significance at the 10% level, ** at the 5% level and *** at the 1% level.

Table 2.1: Summary Statistics of NLSY-79 Female Sample by Age at First Birth

Variable	Whole sample (N=4652)		Teen mother (N=574)		Non-teen-mother (N=4078)	
	Mean	SD	Mean	SD	Mean	SD
Black*	0.143	0.350	0.334	0.472	0.124	0.329
Hispanic*	0.061	0.240	0.089	0.286	0.059	0.235
Live in urban at age 14	0.777	0.416	0.769	0.422	0.778	0.416
Live in south at age 14*	0.324	0.468	0.430	0.495	0.314	0.464
Live in intact family at age 14*	0.822	0.382	0.683	0.466	0.836	0.370
Number of siblings*	3.400	2.307	4.344	2.723	3.303	2.238
Father's education*	10.555	4.937	7.528	5.153	10.865	4.809
Mother's education*	11.001	3.657	8.867	4.005	11.219	3.548
Family magazine subscription at age 14*	0.650	0.477	0.377	0.485	0.678	0.467
Family newspaper subscription at age 14*	0.819	0.385	0.641	0.480	0.838	0.369
Family had library card at age 14*	0.771	0.420	0.595	0.491	0.789	0.408
Foreign language speaking family	0.137	0.344	0.144	0.352	0.136	0.343
Frequent religious activity*	0.361	0.480	0.232	0.422	0.374	0.484
ASVAB score (arithmetic reasoning)*	48.038	9.442	41.693	7.424	48.686	9.386
ASVAB score (word knowledge)*	49.014	9.818	41.006	10.548	49.832	9.363
ASVAB score (mathematics knowledge)*	49.250	9.616	42.083	6.553	49.983	9.579

Source: Authors' calculation from NLSY-79.

Note: 1. Sample is weighted by the revised NLSY-79 sampling weight that takes into account that we are omitting the poor white sub-sample.

2. * denotes the differences in variable means between teen mother and non-teen-mother groups are statistically significant at 5%.

Table 2.2: Summary Statistics of NLSY-79 Female Birth18 and Birth19 Samples

Variable	Birth18 sample (N=279)		Birth19 sample (N=279)	
	Mean	SD	Mean	SD
Black*+	0.196	0.398	0.219	0.414
Hispanic	0.094	0.292	0.093	0.290
Live in urban at age 14	0.773	0.420	0.726	0.447
Live in south at age 14	0.439	0.497	0.429	0.496
Live in intact family at age14*	0.765	0.424	0.739	0.440
Number of siblings+	3.941	2.543	3.824	2.290
Father's education*	8.898	4.568	8.356	4.816
Mother's education+	9.432	3.579	10.005	3.494
Family magazine subscription at age 14+	0.453	0.499	0.499	0.501
Family newspaper subscription at age 14*	0.729	0.446	0.702	0.458
Family had library card at age 14*	0.701	0.459	0.647	0.479
Foreign language speaking family	0.116	0.321	0.160	0.367
Frequent religious activity+	0.309	0.463	0.394	0.490
ASVAB score (arithmetic reasoning)*+	44.115	8.320	44.232	7.847
ASVAB score (word knowledge)*+	44.328	9.744	44.528	10.340
ASVAB score (mathematics knowledge)*+	44.675	7.809	44.693	7.691

Source: Authors' calculation from NLSY-79.

Note: 1. The Birth18 sample consists of NLSY-79 females whose first pregnancy was between their 18th and 19th birthdays and for whom the pregnancy ended with a live birth. The Birth19 sample consists of females who experienced their first pregnancy between their 19th to 20th birthdays and had the pregnancy result in a live birth.

2. Sample is weighted by the revised NLSY-79 sampling weight that takes into account that we are omitting the poor white sub-sample.

3. * denotes that the differences in variable means between teen mother and Birth18 sample are statistically significant at 5% and $^{+}$ denotes that the differences in variable means between teen mother and Birth19 sample are statistically significant at 5%.

Table 2.3: Probit Estimates of Teenage Childbearing Effects on Maternal Schooling with NLSY-79 Female Sample

Dependent variable	No covariates	Family background variables	Col.2 plus ASVAB scores
HS completion by age 21	-1.443	-1.207	-1.152
(N=4643)	(0.071)	(0.078)	(0.081)
	[-0.467]	[-0.376]	[-0.319]
	(0.026)	(0.025)	(0.023)
Pseudo R^2	0.126	0.265	0.392
GED by age 21	0.770	0.738	0.742
(N=4628)	(0.091)	(0.103)	(0.107)
	[0.126]	[0.116]	[0.117]
	(0.021)	(0.019)	(0.020)
Pseudo R^2	0.045	0.087	0.093
Attend college by age 25	-1.257	-0.972	-0.810
(N=4629)	(0.082)	(0.096)	(0.106)
	[-0.382]	[-0.218]	[-0.148]
	(0.016)	(0.016)	(0.016)
Pseudo R^2	0.043	0.206	0.317

Source: Authors' calculation from NLSY-79.

Note: 1. Family background variables include Black, Hispanic, urban at 14, south at 14, intact family at 14, number of siblings, father's education, mother's education, foreign language speaking family, frequent religious activity, family had magazine subscription at 14, family had newspaper subscription at 14, family had library card at 14, and birth year dummies. ASVAB scores consist of the first two principal components of the ten age-adjusted test scores of ASVAB.
2. Sample is weighted by the revised NLSY-79 sampling weight that takes into account that we are omitting the poor white sub-sample.
3. Robust standard errors are in parenthesis below the estimates and mean marginal effects for teen mother sample are in brackets. Standard errors of mean marginal effects are calculated by the delta method.

Table 2.4: Probit Estimates of Effects of Teenage Childbearing on Maternal Schooling with NLSY-79 Female Birth18 and Birth19 Samples

Comparison: first pregnancy between 18th and 19th birthdays and end in live birth			
Dependent variable	No covariates	Family background variables	Col.2 plus ASVAB scores
HS completion by age 21	-0.584	-0.670	-0.675
(N=852)	(0.115)	(0.124)	(0.114)
	[-0.228]	[-0.219]	[-0.183]
	(0.043)	(0.039)	(0.036)
Pseudo R^2	0.035	0.172	0.319
GED by age 21	0.589	0.715	0.835
(N=851)	(0.171)	(0.177)	(0.190)
	[0.107]	[0.100]	[0.106]
	(0.027)	(0.020)	(0.019)
Pseudo R^2	0.034	0.133	0.195
Attend college by age 25	-0.091	-0.134	-0.072
(N=849)	(0.142)	(0.149)	(0.155)
	[-0.016]	[-0.023]	[-0.011]
	(0.026)	(0.026)	(0.023)
Pseudo R^2	0.001	0.194	0.259
Comparison: first pregnancy between 19th and 20th birthdays and end in live birth			
Dependent variable	No covariates	Family background variables	Col.2 plus ASVAB scores
HS completion by age 21	-1.014	-1.077	-1.194
(N=850)	(0.124)	(0.136)	(0.149)
	[-0.369]	[-0.331]	[-0.310]
	(0.039)	(0.037)	(0.035)
Pseudo R^2	0.096	0.234	0.369
GED by age 21	0.268	0.372	0.451
(N=849)	(0.157)	(0.166)	(0.174)
	[0.058]	[0.061]	[0.069]
	(0.032)	(0.025)	(0.023)
Pseudo R^2	0.008	0.100	0.152
Attend college by age 25	-0.376	-0.372	-0.284
(N=851)	(0.132)	(0.150)	(0.156)
	[-0.079]	[-0.070]	[-0.046]
	(0.030)	(0.029)	(0.026)

Pseudo R^2	0.017	0.197	0.288

Source: Authors' calculation from NLSY-79.

Note: 1. The Birth18 sample consists of NLSY-79 females whose first pregnancy was between their 18th and 19th birthdays and for whom the pregnancy ended with a live birth. The Birth19 sample consists of females who experienced their first pregnancy between their 19th to 20th birthdays and had the pregnancy result in a live birth.

2. Family background variables include Black, Hispanic, urban at 14, south at 14, intact family at 14, number of siblings, father's education, mother's education, foreign language speaking family, frequent religious activity, family had magazine subscription at 14, family had newspaper subscription at 14, family had library card at 14, and birth year dummies. ASVAB scores consist of the first two principal components of the ten age-adjusted test scores of the ASVAB.

3. The Sample is weighted by the revised NLSY-79 sampling weight that takes into account that we are omitting the poor white sub-sample.

4. Robust standard errors are in parenthesis below the estimates and mean marginal effects for the teen mother sample are in brackets. Standard errors of mean marginal effects are calculated by the delta method.

Table 2.5: Matching Estimates of Teenage Childbearing Effects on Maternal Schooling with NLSY-79 Female Samples

HS completion by age 21	GED by age 21	Attend college by age 25
Comparison: NLSY-79 Female Sample		
-0.248	0.073	-0.104
(0.028)	(0.019)	(0.018)
N_1=573	N_1=572	N_1=572
N_0=4073	N_0=4056	N_0=4057
N_{nc}=1	N_{nc}=1	N_{nc}=1
Comparison: first pregnancy between 18th and 19th birthdays and end in live birth		
HS completion by age 21	GED by age 21	Attend college by age 25
-0.127	0.071	-0.007
(0.055)	(0.024)	(0.023)
N_1=573	N_1=572	N_1=572
N_0=279	N_0=279	N_0=277
N_{nc}=6	N_{nc}=6	N_{nc}=6
Comparison: first pregnancy between 19th and 20th birthdays and end in live birth		
HS completion by age 21	GED by age 21	Attend college by age 25
-0.243	0.052	-0.025
(0.041)	(0.024)	(0.026)
N_1=573	N_1=572	N_1=572
N_0=279	N_0=279	N_0=279
N_{nc}=3	N_{nc}=3	N_{nc}=3

Source: Authors' calculation from NLSY-79.

Note: 1. The first panel uses the whole sample, the second panel uses women who experienced their first pregnancy after their 18th birthday and before their 19th birthday and for whom the pregnancy ended with a live birth as the comparison group while the third panel utilizes women who experienced their first pregnancy after their 19th birthday and before their 20th birthday and for whom the pregnancy ended in a live birth as the comparison group.

2. These are kernel matching estimates based on probit-estimated propensity scores where the kernel types and bandwidth values are those in Table A2. They are obtained through leave-one-out cross-validation.

3. Sample is weighted by the revised NLSY-79 sampling weight that takes into account that we are omitting the poor white sub-sample.

4. Bootstrap standard errors from 2000 replications are in parentheses. N_1 denotes the number of teen mothers and N_0 is the number of comparison observations. N_{nc} is the number of treated observations that lie outside the common support region.

Table 2.6: Sensitivity Analysis with Bivariate Probit Model
(NLSY-79 Female Sample)

Correlation	HS completion by age 21	GED by age 21	Attend college by age 25
0.0	-0.319	0.117	-0.148
	(0.023)	(0.020)	(0.016)
-0.1	-0.294	0.093	-0.123
	(0.025)	(0.017)	(0.019)
-0.2	-0.266	0.075	-0.108
	(0.023)	(0.015)	(0.024)
-0.3	-0.196	0.040	-0.097
	(0.022)	(0.013)	(0.025)
-0.4	-0.147	0.0008	-0.062
	(0.019)	(0.009)	(0.026)
-0.5	-0.080	-0.008	0.022
	(0.017)	(0.009)	(0.027)

Source: Authors' calculation from NLSY-79.
Note: 1. Estimates are from bivariate probit models with the correlation coefficient of the bivariate normal distribution set to the values in the first column. The signs of the correlation coefficients are positive for the GED outcome in the third column.
2. Sample is weighted by the revised NLSY-79 sampling weight that takes into account that we are omitting the poor white sub-sample.
3. These estimates are mean marginal effects for the teen mother sample. Robust standard errors which are calculated by the delta method are in parentheses.

Table 2.7: Sensitivity Analysis with Bivariate Probit Model
(NLSY-79 Female Birth18 and Birth19 Samples)

Comparison: first pregnancy between 18th and 19th birthdays and end in live birth			
Correlation	HS completion by age 21	GED by age 21	Attend college by age 25
0.0	-0.183	0.106	-0.011
	(0.036)	(0.019)	(0.023)
-0.1	-0.163	0.096	-0.009
	(0.035)	(0.021)	(0.023)
-0.2	-0.153	0.071	0.007
	(0.036)	(0.021)	(0.022)
-0.3	-0.142	0.045	0.012
	(0.037)	(0.023)	(0.024)
-0.4	-0.120	0.015	0.028
	(0.037)	(0.023)	(0.035)
-0.5	-0.102	-0.003	0.040
	(0.035)	(0.026)	(0.024)
Comparison: first pregnancy between 19th and 20th birthdays and end in live birth			
Correlation	HS completion by age 21	GED by age 21	Attend college by age 25
0.0	-0.310	0.069	-0.046
	(0.035)	(0.023)	(0.026)
-0.1	-0.290	0.047	-0.037
	(0.034)	(0.024)	(0.025)
-0.2	-0.242	0.039	-0.025
	(0.034)	(0.025)	(0.026)
-0.3	-0.200	0.001	-0.020
	(0.034)	(0.032)	(0.024)
-0.4	-0.170	-0.003	0.003
	(0.035)	(0.029)	(0.025)
-0.5	-0.114	-0.020	0.042
	(0.035)	(0.031)	(0.025)

Source: Authors' calculation from NLSY-79.

Note: 1. The Birth18 sample consists of NLSY-79 females whose first pregnancy was between their 18th and 19th birthdays and for whom the pregnancy ended with a live birth. The Birth19 sample consists of females who experienced their first pregnancy between their 19th to 20th birthdays and had the pregnancy result in a live birth.

2. Estimates are from bivariate probit models with the correlation coefficient of the bivariate normal distribution set to the values in the first column. The signs of the correlation coefficients are positive for the GED outcome in the third column.

2. Sample is weighted by the revised NLSY-79 sampling weight that takes into account that we are omitting the poor white sub-sample.

3. These estimates are mean marginal effects for the teen mother sample. Robust standard errors which are calculated by the delta method are in parentheses.

Table 2.8: Sensitivity Analysis with Calibrated Unobservable (NLSY-79 Female Sample)

	Simulated unobservable				Outcome effect	Selection effect	Estimated effect	SE
Dependent variable: HS completion by age 21								
	P11	P10	P01	P00				
No unobservable	0	0	0	0	-	-	-0.248	0.021
Neutral unobservable	0.5	0.5	0.5	0.5	1	1	-0.241	0.025
Unobservable mimic								
Black	0.65	0.42	0.27	0.27	1.054	3.107	-0.228	0.024
Above median ASVAB score 1	0.33	0.11	0.60	0.14	10.091	0.269	-0.217	0.027
Dependent variable: GED by age 21								
	Simulated unobservable				Outcome effect	Selection effect	Estimated effect	SE
	P11	P10	P01	P00				
No unobservable	0	0	0	0	-	-	0.073	0.015
Neutral unobservable	0.5	0.5	0.5	0.5	1	1	0.072	0.017
Unobservable mimic								
Black	0.46	0.53	0.19	0.28	0.594	2.851	0.081	0.015
Above median ASVAB score 1	0.44	0.17	0.45	0.55	0.641	0.227	0.065	0.016
Dependent variable: attend college by age 25								
	Simulated unobservable				Outcome effect	Selection effect	Estimated effect	SE
	P11	P10	P01	P00				
No unobservable	0	0	0	0	-	-	-0.104	0.016

Neutral unobservable	0.5	0.5	0.5	0.5	1	1	-0.111	0.020
Unobservable mimic								
Black	0.73	0.50	0.28	0.27	1.088	3.033	-0.097	0.019
Above median ASVAB score 1	0.46	0.18	0.73	0.39	4.189	0.277	-0.088	0.021

Source: Authors' calculation from NLSY-79.

Note: 1. Let *U* be a binary unobservable and denote the fraction of *U*=1 by teen motherhood status and outcomes as: Pij=P(*U*=1 | *D* =i, *Y*=j), with i, j = {0,1}. On the basis of these parameters, a value of *U* is imputed to each individual and the teenage childbearing effect is estimated by kernel matching as in Table 2.5 with *U* in the set of covariates. There are 1000 replications in the simulation.

Table 2.9: Sensitivity Analysis with Calibrated Unobservable (NLSY-79 Female Birth18 Sample)

Dependent variable: HS completion by age 21								
	Simulated unobservable				Outcome effect	Selection effect	Estimated effect	SE
	P11	P10	P01	P00				
No unobservable	0	0	0	0	-	-	-0.127	0.055
Neutral unobservable	0.5	0.5	0.5	0.5	1	1	-0.123	0.051
Unobservable mimic								
Black	0.65	0.42	0.40	0.25	2.196	2.287	-0.128	0.050
Above median ASVAB score 1	0.33	0.11	0.42	0.15	5.777	0.584	-0.117	0.052

Dependent variable: GED by age 21								
	Simulated unobservable				Outcome effect	Selection effect	Estimated effect	SE
	P11	P10	P01	P00				
No unobservable	0	0	0	0	-	-	0.071	0.024
Neutral unobservable	0.5	0.5	0.5	0.5	1	1	0.072	0.023
Unobservable mimic								
Black	0.46	0.53	0.20	0.36	0.244	1.9	0.079	0.025
Above median ASVAB score 1	0.44	0.17	0.33	0.33	2.106	0.541	0.082	0.026

Dependent variable: attend college by age 25								
	Simulated unobservable				Outcome effect	Selection effect	Estimated effect	SE
	P11	P10	P01	P00				
No unobservable	0	0	0	0	-	-	-0.007	0.023
Neutral unobservable	0.5	0.5	0.5	0.5	1	1	-0.007	0.021

Unobservable mimic								
Black	0.73	0.50	0.57	0.32	4.359	2.112	-0.010	0.023
Above median ASVAB score 1	0.46	0.18	0.49	0.30	3.186	0.535	-0.006	0.021

Source: Authors' calculation from NLSY-79.

Note: 1. The Birth18 sample consists of NLSY-79 females whose first pregnancy was between their 18th and 19th birthdays and for whom the pregnancy ended with a live birth.

2. Let U be a binary unobservable and denote the fraction of U=1 by teen motherhood status and outcomes as: Pij=P(U=1 | D =i, Y=j), with i, j = {0,1}. On the basis of these parameters, a value of U is imputed to each individual and the teenage childbearing effect is estimated by kernel matching as in Table 2.5 with U in the set of covariates. There are 1000 replications in the simulation.

Table 2.10. Sensitivity Analysis with Calibrated Unobservable (NLSY-79 Female Birth19 Sample)

	Simulated unobservable				Outcome effect	Selection effect	Estimated effect	SE
Dependent variable: HS completion by age 21								
	P11	P10	P01	P00				
No unobservable	0	0	0	0	-	-	-0.243	0.041
Neutral unobservable	0.5	0.5	0.5	0.5	1	1	-0.241	0.043
Unobservable mimic								
Black	0.65	0.42	0.45	0.19	4.144	1.978	-0.238	0.044
Above median ASVAB score 1	0.33	0.11	0.36	0.13	4.238	0.657	-0.245	0.047
Dependent variable: GED by age 21								
	Simulated unobservable				Outcome effect	Selection effect	Estimated effect	SE
	P11	P10	P01	P00				
No unobservable	0	0	0	0	-	-	0.052	0.024
Neutral unobservable	0.5	0.5	0.5	0.5	1	1	0.055	0.023
Unobservable mimic								
Black	0.46	0.53	0.18	0.41	0.260	1.970	0.056	0.025
Above median ASVAB score 1	0.44	0.17	0.45	0.30	2.534	0.630	0.053	0.026
Dependent variable: attend college by age 25								
	Simulated unobservable				Outcome effect	Selection effect	Estimated effect	SE
	P11	P10	P01	P00				
No unobservable	0	0	0	0	-	-	-0.025	0.026
Neutral unobservable	0.5	0.5	0.5	0.5	1	1	-0.027	0.021

Unobservable mimic								
Black	0.73	0.50	0.54	0.36	2.695	1.863	-0.023	0.019
Above median ASVAB score 1	0.46	0.18	0.48	0.27	4.299	0.583	-0.024	0.020

Source: Authors' calculation from NLSY-79.

Note: 1. The Birth19 sample consists of females who experienced their first pregnancy between their 19th to 20th birthdays and had the pregnancy result in a live birth.
2. Let *U* be a binary unobservable and denote the fraction of *U*=1 by teen motherhood status and outcomes as:
Pij=P(*U*=1 | *D* =i, *Y*=j), with i, j = {0,1}. On the basis of these parameters, a value of *U* is imputed to each individual and the teenage childbearing effect is estimated by kernel matching as in Table 2.5 with *U* in the set of covariates. There are 1000 replications in the simulation.

Table A1: Construction of First Two Principal Components of Adjusted ASVAB Test Scores

Component	Eigenvalue	Difference	Proportion Explained	Cumulative Explained
1	6.640	5.806	0.664	0.664
2	0.834	0.302	0.083	0.747
3	0.532	0.079	0.053	0.801
4	0.453	0.096	0.045	0.846
5	0.356	0.028	0.036	0.882
6	0.329	0.058	0.033	0.914
7	0.271	0.018	0.027	0.941
8	0.253	0.075	0.025	0.967
9	0.177	0.022	0.018	0.985
10	0.155	-	0.016	1.000
Eigenvectors			1st PC	2nd PC
General science			0.338	-0.180
Arithmetic reasoning			0.336	-0.057
Word knowledge			0.347	-0.026
Paragraph comprehension			0.333	0.075
Numerical operations			0.292	0.564
Coding speed			0.273	0.622
Auto and shop knowledge			0.293	-0.289
Mathematics knowledge			0.327	-0.024
Mechanical comprehension			0.303	-0.286
Electrical information			0.311	-0.295

Source: Authors' calculation from NLSY-79.

Note: 1. ASVAB scores are adjusted for age by regressing each test score on birth year dummy variables and schooling information at the test date. Principal components analysis is performed on the OLS residuals from these regressions.

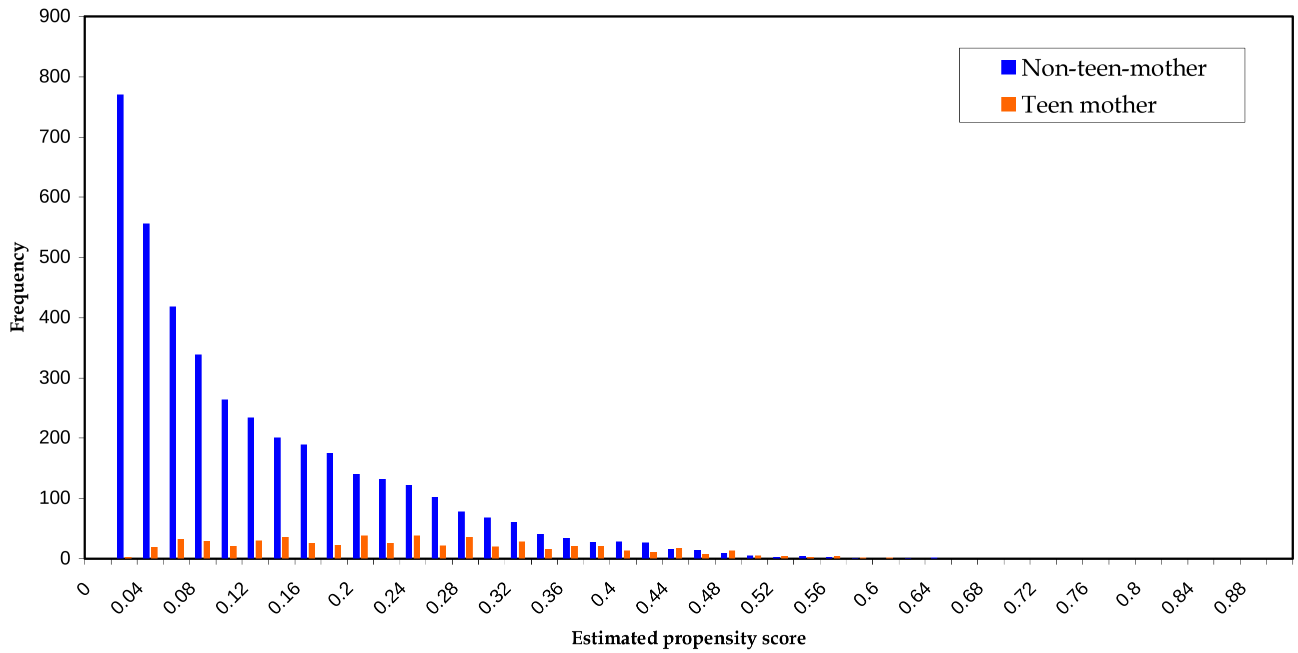
Table A2: Cross-Validation of Kernel Type and Bandwidth Value

Dependent variable	Cross-validated Kernel type	Cross-validated bandwidth value
Comparison: NLSY-79 Female Sample		
HS completion by age 21	epanechnikov	0.05
GED by age 21	gaussian	0.03
Attend college by age 25	epanechnikov	0.02
Comparison: first pregnancy between 18th and 19th birthdays and end in live birth		
Dependent variable	Cross-validated Kernel type	Cross-validated bandwidth value
HS completion by age 21	epanechnikov	0.03
GED by age 21	epanechnikov	0.02
Attend college by age 25	epanechnikov	0.02
Comparison: first pregnancy between 19th and 20th birthdays and end in live birth		
Dependent variable	Cross-validated Kernel type	Cross-validated bandwidth value
HS completion by age 21	epanechnikov	0.03
GED by age 21	epanechnikov	0.04
Attend college by age 25	epanechnikov	0.02

Note: 1. Cross-validation is done through leave-one-out mechanism. Within each comparison group, we choose the kernel type and bandwidth value as $\dot{\theta} = \arg\min_{\dot{\theta}} \sum_{i=1}^{N_0} (Y_i - \hat{m}_{-i}(\hat{P}(X))^2$

2. We cross-validate three kernel types: biweight kernel, gaussian kernel, and epanechnikov kernel.

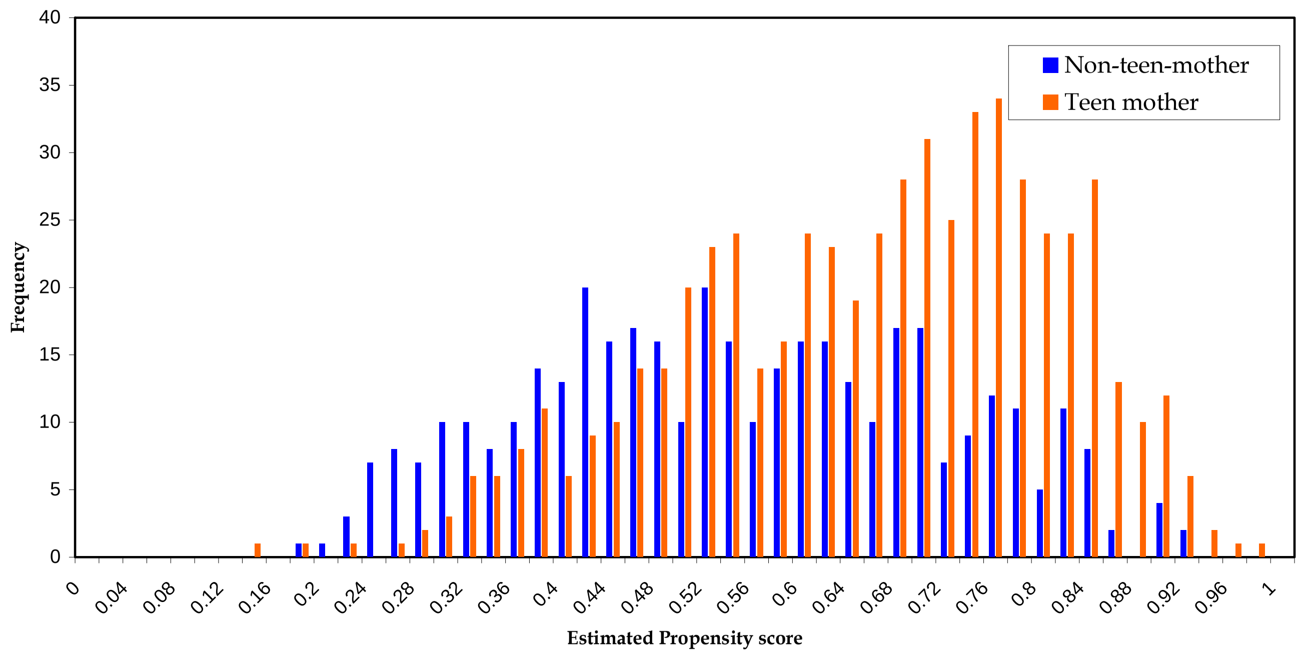
Figure 2.1: Frequency of Estimated Propensity Score by Age of First Birth (NLSY-79 Female Sample)



Source: Authors' calculation from NLSY-79.

Note: 1. The propensity score $(P(D=1 \mid X))$ is estimated through a probit model using as conditional covariates (X) : Black, Hispanic, urban at 14, south at 14, intact family at 14, number of siblings, father's education, mother's education, foreign language speaking family, frequent religious activity, family magazine subscription at 14, family newspaper subscription at 14, family has library card at 14, birth year dummies, and the first two principal components of ten adjusted ASVAB test scores. We also include four higher order and interaction terms: (ASVAB score1*ASVAB score1), (father's education*Black), (foreign language speaking family*ASVAB score1), and (frequent religious activity*ASVAB score1) to the propensity score specification following the balancing test described in the text.

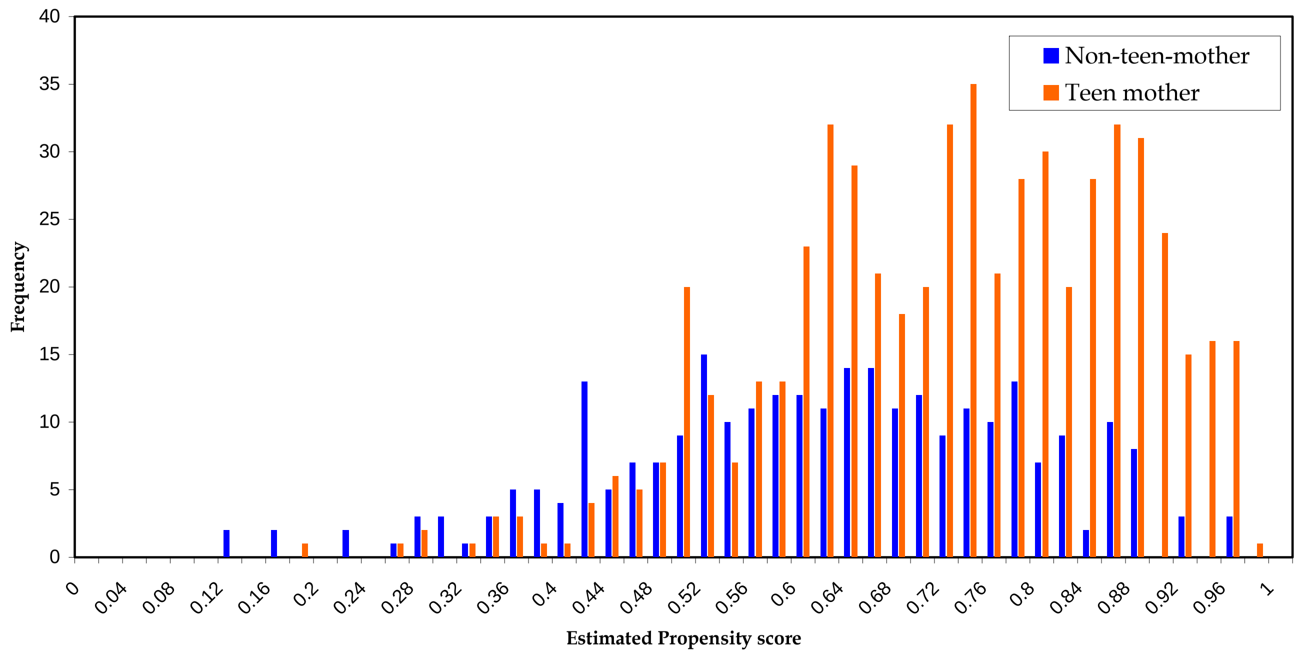
Figure 2.2: Frequency of Estimated Propensity Score by Age of First Birth
(NLSY-79 Female Birth18 Sample)



Source: Authors' calculation from NLSY-79.

Note: 1. The propensity score $(P(D=1 \mid X))$ is estimated through a probit model using as conditional covariates (X) : Black, Hispanic, urban at 14, south at 14, intact family at 14, number of siblings, father's education, mother's education, foreign language speaking family, frequent religious activity, family magazine subscription at 14, family newspaper subscription at 14, family has library card at 14, birth year dummies, and the first two principal components of ten adjusted ASVAB test scores. We also include two interaction terms: (ASVAB score2*ASVAB score2) and (sibling*ASVAB score1) to the propensity score specification following the balancing test described in the text.

Figure 2.3: Frequency of Estimated Propensity Score by Age of First Birth
(NLSY-79 Female Birth19 Sample)



Source: Authors' calculation from NLSY-79.
Note: 1. The propensity score $(P(D=1 \mid X))$ is estimated through a probit model using as conditional covariates (X) : Black, Hispanic, urban at 14, south at 14, intact family at 14, number of siblings, father's education, mother's education, foreign language speaking family, frequent religious activity, family magazine subscription at 14, family newspaper subscription at 14, family has library card at 14, birth year dummies, and the first two principal components of ten adjusted ASVAB test scores. We also include five interaction terms: (ASVAB score1*ASVAB score1), (ASVAB score2*ASVAB score2), (Black*ASVAB score2), (frequent religious activity*ASVAB score2), and (intact family at 14*ASVAB score2) to the propensity score specification following the balancing test described in the text.

Bibliography

[1] Acemoglu, D. and J. Pischke (1998), "Why Do Firms Train? Theory And Evidence." *The Quarterly Journal of Economics*, 113, 78-118.

[2] Aigner, D. and G. Cain (1977), "Statistical Theories of Discrimination in Labor Mrkets." *Industrial and Labor Relations Review*, 32, 175-187.

[3] Akerlof, G. (1970), "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics*, 84, 488-500.

[4] Altonji, J., T. Elder, and C. Taber (2005), "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113, 151-184.

[5] Altonji, J. and C. Pierret (1998), "Employer Learning and the Signaling Value of Education." in I. Ohashi and T. Tachibanaki, eds., *Internal Labour Markets, Incentives and Employment*, London, UK: Macmillan Press.

[6] Altonji, J. and C. Pierret (2001), "Employer Learning and Statistical Discrimination." *The Quarterly Journal of Economics*, 116, 313-500.

[7] Altonji, J. and R. Shakotko (1987), "Do Wages Rise with Job Seniority?." *Review of Economic Studies*, 54, 437-459.

[8] Altonji, J. and J. Spletzer (1991), "Worker Characteristics, Job Characteristics, and the Receipt of On-the-Job Training." *Industrial and Labor Relations Review*, 45, 58-79.

[9] Ashcraft, A. and K. Lang (2006), "The Consequences of Teenage Childbearing." NBER Working Paper.

[10] Bauer, T. and J. Haisken-DeNew (2001), "Employer Learning and the Returns to Schooling." *Labour Economics*, 8, 161-180.

[11] Bernhart, D. (1995), "Strategic Promotion and Compensation." *Review of Economic Studies*, 62, 315-339.

[12] Black, D. and J. Smith (2004), "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching." *Journal of Econometrics*, 121, 99-124.

[13] Blackburn, M. and D. Neumark (1992), "Unobserved Ability, Efficiency Wage, and Interindustry Wage Differentials." *The Quarterly Journal of Economics*, 107, 1421-1436.

[14] Bollinger, C. and A. Chandra (2005), "Iatrogenic Specification Error: A Cautionary Tale of Cleaning Data." *Journal of Labor Economics*, 23, 235-257.

[15] Bronars, S. and J. Grogger (1995), "The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment." *American Economic Review*, 84, 654-698.

[16] Burdett, K. (1995), "Theory of Employee Search: Quit Rates." *American Economic Review*, 68, 212-220.

[17] Bureau of Labor Statistics (2006), "Number of Jobs Held, Labor Market Activity, and Earnings Growth Among the Youngest Baby Boomers: Results from a Longitudinal Survey." *News: United States Department of Labor*, 16-1496.

[18] Cameron, S. and J. Heckman (1993), "The Nonequivalence of High School Equivalents." *Journal of Labor Economics*, 11, 1-47.

[19] Card, J. and L. Wise (1978), "Teenage Mothers and Teenage Fathers: The Impact of Early Childbearing on Parents' Personal and Professional Lives." *Family Planning Perspectives*, 10, 100-205.

[20] Cascio, E. and E. Lewis (2006), "Schooling and the Armed Forces Qualifying Test: Evidence from School-Entry Laws." *Journal of Human Resources*, 41, 294-318.

[21] Cawley, J., J. Heckman, and E. Vytlacil (2001), "Three Observations on Wages and Measured Cognitive Ability." *Labour Economics*, 8, 419-442.

[22] Coate, S. and G. Loury (1993), "Will Affirmative-Action Policies Eliminate Negative Stereotypes?." *American Economic Review*, 83, 1220-1240.

[23] Dehejia, R. and S. Wahba (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association*, 94, 1053-1062.

[24] Dehejia, R. and S. Wahba (2002), "Propensity Score Matching Methods for Nonexperimental Causal Studies." *Review of Economics and Statistics*, 84, 151-161.

[25] Farber, H. and R. Gibbons (1996), "Learning and Wage Dynamics." *The Quarterly Journal of Economics*, 111, 1007-1047.

[26] Freeman, S. (1977), "Wage Trends as Performance Displays Productive Potential: A Model and Application to Academic Early Retirement." *The Bell Journal of Economics*, 8, 419-443.

[27] Frölich, M. (2004), "Finite Sample Properties of Propensity-Score Matching and Weighting Estimators." *Review of Economics and Statistics*, 86, 77-90.

[28] Galindo-Rueda, F. (2003), "Employer Learning and Schooling-Related Statistical Discrimination in Britain." Discussion Paper, IZA.

[29] Geronimus, A. and S. Korenman (1992), "The Socioeconomic Consequences of Teen Childbearing Reconsidered." *Quarterly Journal of Economics*, 107, 1187-1214.

[30] Geronimus, A. and S. Korenman (1993), "The Socioeconomic Costs of Teenage Childbearing: Evidence and Interpretation." *Demography*, 30, 281-290.

[31] Gibbons, R. and L. Katz (1991), "Layoffs and Lemons." *Journal of Labor Economics*, 9, 315-389.

[32] Greenwald, B. (1986), "Adverse Selection in the Labour Market." *Review of Economic Studies*, 53, 325-347.

[33] Griliches, Z. (1979), "Sibling Models and Data in Economics: the Beginnings of a Survey." *Journal of Political Economy*, 87, 37-64.

[34] Guttmacher Institute (2006), *U.S. Teenage Pregnancy Statistics National and State Trends and Trends by Race and Ethnicity.* New York, NY: Guttmacher Institute.

[35] Hansen, K., J. Heckman, and K. Mullen (2004), "The Effect of Schooling and Ability on Achievement Test Scores." *Journal of Econometrics*, 121, 39-98.

[36] Harris, M. and B. Holmstrom (1982), "A Theory of Wage Dynamics." *Review of Economic Studies*, 49, 315-333.

[37] Hayes, C. (1987), *Risking the Future: Adolescent Sexuality, Pregnancy and Childbearing.* Washington, DC: National Academy Press.

[38] Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica*, 46, 931-959.

[39] Heckman, J. (1995), "Lessons from the Bell Curve." *Journal of Political Economy*, 103, 1091-1120.

[40] Heckman, J. (1997), "Instrumental Variables: A Study of Implicit Behavioral Assumptions in One Widely Used Estimator." *Journal of Human Resources*, 32, 441-462.

[41] Heckman, J., H. Ichimura, and P. Todd (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program." *Review of Economic Studies*, 64, 605-654.

[42] Heckman, J., H. Ichimura, and P. Todd (1998), "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies*, 65, 261-294.

[43] Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998), "Characterizing Selection Bias Using Experimental Data." *Econometrica*, 66, 1017-1098.

[44] Heckman, J., R. LaLonde, and J. Smith (1999), "The Economics and Econometrics of Active Labor Market Programs." in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Volume 3, Amsterdam, Netherlands: Elsevier.

[45] Heckman, J. and R. Robb (1985), "Alternative Methods for Evaluation the Impact of Interventions." in J. Heckman and B.Singer, eds., *Longitudinal Analysis of Labor Market Data*, Cambridge, UK: Cambridge University Press.

[46] Hirano, K., G. Imbens, and G. Ridder (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica*, 71, 1161-1189.

[47] Hofferth, S. and K. Moore (1979), "Early Childbearing and later Economic Well-Being." *American Sociological Review*, 44, 784-815.

[48] Hotz, V J., C. Mullin, and S. Sanders (1997), "Bounding Causal Effects Using Data from a Contaminated Natural Experiment: Analyzing the Effects of Teenage Childbearing." *Review of Economic Studies*, 64, 575-603.

[49] Hotz, V J., S. McElroy, and S. Sanders (2005), "Teenage Childbearing and its Life Cycle Consequences: Exploiting a Natural Experiment." *Journal of Human Resources*, 40, 683-715.

[50] Hu, L. and C. Taber (2005), "Layoffs, Lemons, Race, and Gender." NBER Working Paper.

[51] Ichino, A., F. Mealli, and T. Nannicini (2007), "From Temporary Help Jobs to Permanent Employment: What can We Learn from Matching Estimators and Their Sensitivity?" *Journal of Applied Econometrics*, forthcoming.

[52] Imbens, G. (2004), "Semiparametric Estimation of Average Treatment Effects Under Exogeneity: A Review." *Review of Economics and Statistics*, 86, 4-29.

[53] Johnson, W. (1978), "A Theory of Job Shopping." *The Quarterly Journal of Economics*, 92, 261-277.

[54] Jovanovic, B (1979a), "Job Matching and the Theory of Turnover." *Journal of Political Economy*, 87, 972-990.

[55] Jovanovic, B (1979b), "Firm-specific Capital and Turnover." *Journal of Political Economy*, 87, 1246-1260.

[56] Lang, K (1986), "A Language Theory of Discrimination." *The Quarterly Journal of Economics*, 101, 363-382.

[57] Lange, F. (2007), "The Speed of Employer Learning." *Journal of Labor Economics*, 25, 1-35.

[58] Lazear, E (1986), "Raids and Offer Matching." *Research in Labor Economics*, 8, 141-165.

[59] Lechner, M. (1999), "Earnings and Employment Effects of Continuous On-the-Job Training in East German after Unification." *Journal of Business and Economic Statistics*, 17, 74-90.

[60] Levine, D. and G. Painter (2003), "The Costs of Teenage Out-of-Wedlock Childbearing: Analysis with a Within-School Propensity Score Matching Estimator." *Review of Economics and Statistics*, 85, 884-900.

[61] Light, A. and K. McGarry (1998), "Job Change Patterns and the Wages of Young Men." *Review of Economics and Statistics*, 80, 276-286.

[62] Lucas, R. and E. Prescott (1982), "Equilibrium Search and Unemployment." *Journal of Economic Theory*, 2, 188-209.

[63] Lundberg, S. and R. Startz (1983), "Private Discrimination and Social Intervention in Competitive Labor Markets." *American Economic Review*, 73, 340-347.

[64] Ljungqvist, L. and T. Sargent (2000), *Recursive Macroeconomic Theory* Cambridge, MA: MIT Press.

[65] Maynard, R. (1996), *Kids Having Kids: A Robin Hood Foundation Special Report on the Costs of Adolescent Childbearing.* New York, NY: Robin Hood Foundation.

[66] Milgrom, P. and S. Oster (1987), "Job Discrimination, Market Forces, and the Invisibility Hypothesis." *The Quarterly Journal of Economics*, 102, 453-476.

[67] Mincer, J. and B. Jovanovic (1981), "Labor Mobility and Wages." in S. Rosen, eds., *Studies in Labor Markets*, Chicago, IL: University of Chicago Press.

[68] Moore, K. and L. Waite (1977), "Teenage Motherhood and Schooling." *Family Planning Perspectives*, 9, 220-225.

[69] Mortensen, D. (1998), "Specific Capital and Labor Turnover." *The Bell Journal of Economics*, 9, 572-586.

[70] Moscarini, G. (2003), "Skill and Luck in the Theory of Turnover." Working Paper, Yale University.

[71] Mott, F. and W. Marsiglio (1985), "Early Childbearing and Completion of High School." *Family Planning Perspectives*, 17, 234-237.

[72] Neal, D. (1998), "The Link between Ability and Specialization: An Explanation for Observed Correlations between Wages and Mobility Rates." *Journal of Human Resources*, 33, 173-200.

[73] Neal, D. and W. Johnson (1996), "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy*, 104, 869-895.

[74] Oettinger, G. (1996), "Statistical Discrimination and the Early Career Evolution of the Black-White Wage Gap." *Journal of Labor Economics*, 14, 52-78.

[75] O'Neill, J. (1990), "The Role of Human Capital in Earnings Differences Between Black and White Men." *Journal of Economic Perspectives*, 4, 25-45.

[76] Pagan, A. and A. Ullah (1999), *Nonparametric Econometrics.* Cambridge, UK: Cambridge University Press.

[77] Racine, J. and Q. Li (2004), "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data." *Journal of Econometrics*, 119, 99-130.

[78] Ribar, D. (1994), "Teenage Fertility and High School Completion." *Review of Economics and Statistics*, 76, 413-424.

[79] Rodriguez-Planas, N. (2004), "Signaling in the Labor Market: New Evidence on Plant Closings and Layoffs." Working Paper, IZA.

[80] Rosenbaum, P. (1987), "Sensitivity Analysis to Certain Permutation Inferences in Matched Observational Studies." *Biometrika*, 74, 13-26.

[81] Rosenbaum, P. and D. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, 70, 41-55.

[82] Rosenbaum, P. and D. Rubin (1984), "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of American Statistical Association*, 79, 516-524.

[83] Rosenbaum, P. and D. Rubin (1985), "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *The American Statistician*, 39, 33-38.

[84] Rosenzweig, M. and K. Wolpin (1995), "Sisters, Siblings, and Mothers: The Effect of Teenage Childbearing on Birth Outcomes in a Dynamic Family Context." *Econometrica*, 63, 303-326.

[85] Schonberg, U. (2005), "Testing for Asymmetric Employer Learning." Working Paper, University of Rochester.

[86] Smith, J. and P. Todd (2005a), "Does Matching Overcome LaLonde's Critique of Nonexperimental Methods?." *Journal of Econometrics*, 125, 305-353.

[87] Smith, J. and P. Todd (2005b), "Rejoinder." *Journal of Econometrics*, 125, 365-375.

[88] Smith, J. and Y. Zhang (2006), "Varieties of Balancing Tests." Working Paper, University of Maryland.

[89] Spence, A. (1973), "Job Market Signaling." *The Quarterly Journal of Economics*, 87, 355-374.

[90] Topel, R. (1991), "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority." *Journal of Political Economy*, 99, 145-176.

[91] Topel, R. and M. Ward (1992), "Job Mobility and the Careers of Young Men." *The Quarterly Journal of Economics*, 107, 439-479.

[92] Trussell, J. (1976), "The Economic Consequences of Teenage Childbearing." *Family Planning Perspectives*, 8, 184-190.

[93] Trussell, J. (1988), "Teenage Pregnancy in the United States." *Family Planning Perspectives*, 20, 262-272.

[94] Tyler, J., R. Murnane, and J. Willett (2000), "Estimating the Labor Market Signaling Value of the GED" *The Quarterly Journal of Economics*, 115, 431-468.

[95] Upchurch, D. and J. McCarthy (1990), "The Timing of a First Birth and High School Completion." *American Sociological Review*, 55, 224-234.

[96] Waldman, M. (1984), "Job Assignment, Signaling and Efficiency." *Rand Journal of Economics*, 15, 255-270.

[97] Wilde, L. (1979), "An Information-theoretic Approach to Job Quits." in S. Lippman and J. McCall, eds., *Studies in the Economics of Search*, Amsterdam: North Holland Publishing.

[98] Zhao, Z. (2004), "Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence." *Review of Economics and Statistics*, 86, 91-103.