Title
Three essays in labor economics

Author
Wang, Liang Choon

Publication Date
2010

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Three Essays in Labor Economics

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Economics

by

Liang Choon Wang

Committee in charge:

Professor Eli Berman, Chair
Professor Julian Betts, Co-Chair
Professor John Evans
Professor Nora Gordon
Professor Craig McIntosh

2010

The Dissertation of Liang Choon Wang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2010

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am extremely grateful to Eli Berman and Julian Betts, whose guidance have nurtured me to pursue a rigorous and policy relevant research agenda. I would not have undertaken many of my research topics without Eli and Julian's personal and professional encouragement and support. I hope I will continue to do good research to justify their generosities. I thank Craig McIntosh for many helpful conversations, critical suggestions, and encouragements at various stages of my research, which have shaped many of my research directions. I am thankful to Nora Gordon, who has provided me with much useful feedback and supported me in obtaining the Spencer Foundation Dissertation Fellowship. I also acknowledge John Evans's suggestions on my research on the Amish and important advice on ethics related issues at early stages of my research in Malaysia. I thank Tiffany Chou, Michael Ewens, Ben Gilbert, Ben Gillen, Jacob LaRiviere, Michael Madrid, and Alex Sawyer, for making my graduate studies a fun process. Finally, thanks to my mom and Kha Yin for being supportive of my endeavors.

Chapter 1: I thank Eli Berman, Julian Betts, Prashant Bharadwaj, Tiffany Chou, Gordon Dahl, John Evans, Michael Ewens, David Figlio, Ben Gilbert, Ben Gillen, Nora Gordon, Larry Hedges, Jacob LaRiviere, Xun Lu, Craig McIntosh, Steve Raudenbush, Alex Sawyer, and Yixiao Sun for helpful comments. I also thank seminar participants at Monash University, UC San Diego, University of Missouri at Columbia, University of Oregon, University of Queensland, University of Wisconsin Madison's WCER's Value-Added Research Center, and the World Bank. I am thankful to the Institute for Applied Economics and the Department of Economics of the University of California San Diego for research funding, and the Spencer Foundation for fellowship support. I also acknowledge the Malaysian secondary school for data provision and assistance that made this study possible. Any errors are the mistake of the author.

Chapter 2: I thank Eli Berman, Julian Betts, Tiffany Chou, John Evans, Nora Gordon, Larry Iannaccone, Jacob LaRievere, Rachel McCleary, Craig McIntosh, Gary Richardson, and Alex Sawyer for helpful comments. I also benefited from comments from participants at Meetings of the Association for the Study of Religion, Economics, and Culture, UC San Diego applied seminars, and the Graduate Workshop on Religion, Economics, and Politics. Any errors are those of the author.

Chapter 3: I thank Eli Berman, Julian Betts, Vince Crawford, John Evans, Marjorie Flavin, Nora Gordon, Craig McIntosh and participants at the Economics Graduate Student Research Seminar at UC San Diego for helpful comments.

VITA

Education:

2004 – 2010	PhD., University of California, San Diego
2002 – 2004	Master of Economics, University of Adelaide
2001 – 2001	Bachelor of Economics (First Class Honors), University of Adelaide
1998 – 2000	Bachelor of Commerce (Accounting), University of Adelaide

Work Experience:

2006 – 2010	Research assistant for Professor Eli Berman, UC San Diego
2008 – 2010	Research assistant for Professor Julian Betts, UC San Diego
2009	Lecturer, UC San Diego
2005 – 2007	Teaching assistant, UC San Diego
2001 – 2004	Teaching assistant, University of Adelaide
2002 – 2004	Research assistant for Professor Ian W. McLean, University of Adelaide
2002	Research assistant for Professor John Whitley, University of Adelaide

Honors, Scholarships, and Fellowships:

2009 – 2010	Spencer Foundation Dissertation Fellowship
2005 – 2008	University of California, San Diego Foreign Student Tuition Scholarship
2003	University of Adelaide Post Graduate Student Scholarship
2001	N.J. Thomson Memorial Prize (Best Honors Thesis in Public Economics)

Publication:

Richard Pomfret and Liang Choon Wang (2003) "Evaluating the Research Output of Australian Universities' Economics Departments," *Australian Economic Papers*, vol.42 (4), pp. 418-441

Grant:

2008 – 2009	UC San Diego Institute for Applied Economics Research Grant

2009	UC San Diego Dean of Social Science Travel Grant
2009	UC San Diego Department of Economics Graduate Student Research and Travel Fund

Professional Activities:

Journal Referee for Public Choice

Member of the American Economic Association

ABSTRACT OF THE DISSERTATION

Three Essays in Labor Economics

by

Liang Choon Wang

Doctor of Philosophy in Economics

University of California, San Diego, 2010

Professor Eli Berman, Chair

Professor Julian Betts, Co-Chair

This dissertation is comprised of three autonomous chapters on topics in labor economics. The first chapter exploits the quasi-random assignment of students into classrooms in a large secondary school in Malaysia to estimate the effects of peers on student outcomes. The estimates show that having better achieving classmates improves a student's math achievement and reduces the student's incidence of class absences and discipline violations. There is also evidence of non-linear peer effects and that average achievement may increase as a result of ability grouping.

The second chapter extends Iannaccone's (1992) religious club model to explain why the Amish would collectively object to high school education and refuse to comply with compulsory schooling laws. I utilize the surprising 1972 U.S. Supreme Court's decision in *Wisconsin vs. Yoder*, which exempts Amish children from compulsory high school education, as a policy shock to test the predictions of the model. The results show that the successful restriction on high school education helped the Amish sect exclude individuals who have high labor productivity and would lower the quality of the sect from joining. The evidence supports the idea that the Amish use the restriction on secular education as a religious sacrifice to screen out uncommitted members.

The third chapter investigates the effect of higher immigration on native fertility. Previous research shows that immigration affects wages, income, and the cost of child rearing, while standard fertility model predicts that changes in wages, income, and the cost of child rearing would affect fertility. Using the cross-state variation in the total fertility rates of native-born American women and the share of immigrants in the population between 1970 and 2005, this chapter estimates that for every one percentage point increase in the share of immigrants in the population, native total fertility rate is predicted to increase by roughly 0.01 children. The negative effect of immigration on wages is the most likely explanation, because the fertility of less educated women and women who resided in their states of birth is most affected.

CHAPTER 1

PEER EFFECTS IN THE CLASSROOM: EVIDENCE FROM A NATURAL EXPERIMENT IN MALAYSIA

Abstract

Studies of peer effects in schools often face identification problems that arise from the non-random selection of students into classrooms or schools. This paper exploits the quasi-random assignment of students into classrooms in a large secondary school in Malaysia to estimate the causal effects of peers on educational outcomes. Even with random assignment, peer effect estimates are vulnerable to a mechanical bias (Guryan et al. forthcoming). I show that existing treatments of this bias are inadequate and underestimate peer effects. I demonstrate a simple solution and find a large positive causal effect of peers on math achievement – a 0.1 standard deviation increase in the average baseline math score of classmates leads to a 0.066 standard deviation increase in a student's current math score. Effort and behavior may account for part of this large peer effect: the presence of high achieving peers lowers absence rates and the incidence of discipline violations. I also find evidence that the peer effects vary non-linearly for different types of students. The results imply that ability grouping may raise average achievement.

1.1 Introduction

Policy debates on the costs and benefits of school choice, ability grouping, and special education programs largely hinge on the existence and forms of peer effects in schools and classrooms. For example, opponents of school choice initiatives often argue that school choice may worsen the educational outcomes of students in schools from which better students have opted out. Parents opposed to ability grouping may also be concerned with the adverse effects of separating low achieving and high achieving students into different classrooms. These concerns are legitimate if peer effects are present so that reassigning high-achieving students away from low-achieving students can redistribute achievement gains. To the extent that test scores predict wages and incomes, the presence of peer effects implies that policy changes altering the ability mix of students in the same school or classroom can impact future wage and income distribution of a country (Bishop 1989; Bound and Johnson 1992; Gamoran and Mare 1989). Quantifying the effects of peers on student outcomes may also provide insights into the spillover effects of any policy intervention. If positive peer effects exist, then a policy intervention directed towards some students can also benefit other students not directly targeted by the policy through interactions between students. Hence, the central questions to policy makers are: do peer effects exist, and if so, what are their sizes and do they vary across different types of students?

Studies of peer effects in schools face numerous identification and estimation problems. The first type of identification problem is what Manski (1993) referred to as the reflection problem and it occurs when researchers cannot properly isolate the effect of a person on her peers from the effect of her peers on her. This issue is essentially an endogeneity problem as outcomes of members in the group are jointly determined (Moffitt 2001). Second, individuals in the same group may have similar behavior and outcomes because they self-select or are selected into the same group on the basis of factors such as motivation and ability, which are unobserved by researchers. In the presence of unobserved selection, group members may behave similarly because they share similar characteristics and not necessarily because they influence one another. Third, members of the same group may also have similar outcomes because they experience common shocks unrelated to their interactions. For instance, students in the same classroom may have similar achievement because they share the same teachers. Manski (1993) labeled effects due to unobserved selection and common group shocks as correlated effects. Finally, a mechanical negative correlation exists between an individual's own outcome and the average of her peers' outcomes in a typical linear regression framework because they deviate in opposite directions from the group mean. With the exception of Guryan et al. (forthcoming), this problem has largely been neglected in the literature. Failing to address these identification problems may lead to inconsistent estimates of the effect of peers.

Several studies attempt to address these identification problems. The first commonly employed strategy, especially for studies of peer effects in school, is to address unobserved selection and the reflection problem by using a fixed effects estimator and replacing current peer achievement with lagged peer achievement (Betts and Zau 2004; Hanushek et al. 2003; Vigdor and Nechyba 2007). Since variation across classrooms or cohorts within a school and lagged peer achievement may not be credibly exogenous, Hoxby and Weingarth (2007) used school-reassignment-induced variation in peers to implement instrumental variable estimation. In contrast, Burke and Sass (2008) adapted a fixed effects estimator proposed by Arcidiacono et al. (2005) and modeled individual outcomes as a function of peers' and teacher fixed effects to control for unobserved selection. Generally, the empirical findings on the size of peer effects on student outcomes are mixed. For example, Vigdor and Nechyba (2007) and Burke and Sass (2008) report either small or insignificant effects of peers on student achievements in schools, while Hanushek et al. (2003), Betts and Zau (2004), Hoxby and Weingarth (2007) report sizable positive effects of peers on student achievement. Moreover, estimated peer effects at the classroom level are typically larger than estimates at the grade level or school level (Betts and Zau 2004).

Natural or controlled experiments are another means of identifying causal effects of peers. The random assignment of students into groups removes correlated effects, since group membership is, by construction, orthogonal to group characteristics. The majority of these studies look at student outcomes as a function of college roommates or classmates (Carrell et al. 2009; De Giorgi et al. forthcoming; Foster 2006; Lyle 2007; Sacerdote 2001; Zimmerman 2003). The effects of classmates are reported to be positive and significant, while the effects of roommates on achievement are relatively weak and insignificant. It is less common to see the random assignment of students into classrooms in K-12 school, with the exception of Kang (2007) and Duflo et al. (forthcoming). Using contemporaneous achievement of a random sample of classmates, Kang (2007) finds strong positive effects of peers on math achievement in South Korean middle schools. Similarly, Duflo et al. (forthcoming) study the effects of tracking and peers for Kenyan first graders in a large scale controlled experiment and report strong positive effects of peers on math and literature achievement. Moreover, they find that students at all levels of the initial test score distribution benefited from tracking.

Using data generated from the quasi-random assignment of students into classrooms in a large secondary school in Malaysia, I investigate peer effects of a student's classmates on individual math achievement, class absences, and discipline violations in the first year of secondary school (grade 7). The sample school assigns students into classrooms using a computer program that attempts to equalize the average baseline test scores across classrooms. This quasi-random assignment method ensures that the estimated peer effects are unlikely confounded by systematic assignment of teachers of different quality into classrooms of different achievement mixes. Similarly, selection bias is effectively controlled because the school only has limited information about each student's past and the classroom assignment method is strictly enforced. The data also provide a clear definition of the relevant peer group, since students in the sample attend all classes with the same set of classmates during an academic year. As most forms of academic interaction occur at the classroom level, the effects of classmates are likely strongest and the most relevant to policy makers. In contrast to previous studies, this paper's comprehensive coverage of classmates and baseline test scores provides a more complete picture of peer effects. Observing standardized baseline test scores is important, as I show that controlling for each student's own baseline test score is essential for removing the mechanical correlation between an individual's outcome and the average outcome of peers. Moreover, because the student body consists of racially homogenous urban students of the same age, potential differential effects of race and age of peers will not confound the estimated effects of peers. This feature is particularly important given recent findings by Angrist and Lang (2004) and Cooley (2008), which show that the effects of peers may differ depending on the racial composition of students in a classroom.

I also examine the mechanical negative relationship between characteristics of individuals and peers and the test of random assignment commonly used in other studies of peer effects. Guryan et al. (forthcoming) show by simulation that there exists a mechanical negative correlation between characteristics of individuals and their peers in professional golf tournaments even when group assignment is random and they propose a modified test of random assignment that is well-behaved. First, I show both analytically and in simulations that the mechanical negative correlation increases as 1) group size (i.e., class size) grows, holding the number of groups constant, and as 2) when the number of groups formed decreases, holding group size fixed. This finding implies that a typical test of random assignment is more likely to reject the null of zero correlation when the number of randomly assigned groups in the sample is small. It also means that past studies using exogenous variation from a small number of classrooms or grade levels within a school might underestimate the size of peer effects. The bias will be present regardless of whether the assignment of peers is random or not. Thus, there is a mechanical explanation for why past studies looking at grade level peer effects tend to find smaller effects than those looking at classroom level peer effects. Second, I show that when the number of observations varies considerably across randomization tracks, the modified test of random assignment proposed by Guryan et al. (forthcoming) may still reject the null hypothesis of zero correlation. This means that neither the typical test of random assignment nor the modified test of random assignment is generally robust to the negative bias. I argue that testing whether group means (i.e., average test scores across classrooms) jointly differ at baseline is sufficient to verify whether group assignment is exogenous. More importantly, I present a robust solution by including an individual's baseline test score as a regressor, yielding a consistent estimate of the peer effect. This also highlights why including a student's own lagged achievement as a regressor is important.

I find that a 0.1 standard deviation increase in the average baseline math score of classmates leads to a 0.066 standard deviation increase in a student's own math score. This estimated causal effect of peers on math achievement is the first reported for a classroom setting that is closer to a typical western classroom and free of the mechanical negative correlation. The estimate is larger than estimates by previous studies that use a fixed effects estimator or an instrumental variable estimator to address selection bias and the reflection problem. I argue that the larger magnitude of estimated peer effects is attributed to the control for selection bias and the correction for the mechanical negative correlation.

Possible mechanisms for peer effects are through effort and conduct. To assess whether having high-achieving peers induces a student to increase the amount of effort and facilitates better behavior, I examine whether the average baseline test scores of classmates has an effect on the number of classes a student missed and on the incidence of discipline violations cited. Class absences and discipline violations are also important measures of outcomes, given the link between schooling and earnings (Angrist and Krueger 1992) as well as the potential association between discipline and crime (Adams 2000) respectively. I find that a one standard deviation increase in the average baseline scores of classmates leads to roughly 2 days of classes a student missed and a 1.58 standard deviation decrease in the incidence of discipline violations. This finding provides some support that having higher achieving peers can alter a student's effort and conduct.

Since linear-in-means specifications do not inform about the effects of ability grouping and ability mixing on achievement, I use a number of non-linear specifications to investigate whether the effect of peers varies across different types of students. I find that the effects of peers vary non-linearly for different types of students. Students in the top 25 and middle 50 percentiles of the initial test score distribution gain significantly from having better achieving classmates when their classmates have low average baseline math score, while students in the bottom 25 percentile do not gain much from having better achieving classmates. The results imply that ability grouping will likely raise average achievement. The distributional effect of ability grouping is likely small given that the effect of peers on low-achieving students is small.

1.2 Identification and Interpretation of Peer Effects

This study is primarily concerned with identifying the effect of classroom peer achievement on a student's own achievement. Consider the following linear-in-means model:

$$y_{ickt} = \beta_0 + \beta_1 \bar{y}_{-ickt} + \delta_k + \varepsilon_{ickt} \qquad (1.1)$$

The dependent variable, y_{ickt}, is the achievement of student *i*, in classroom *c*, cohort-ability group *k*, in period *t*.[1] β_0 is the intercept term and δ_k is a vector of cohort-ability group dummies. The coefficient of interest, β_1, measures how the average achievement of student *i*'s classmates, $\bar{y}_{-ickt}$, affects student *i*. ε_{ickt} is the error term.

Equation (1.1) suffers from Manski's (1993) reflection problem because researchers cannot effectively isolate the effect of a person's behavior on the behavior of the group from the effect of the behavior of the group on the person.[2] Regression equation (1.1) may yield an estimate of β_1 that is non-zero because: (a) the behavior of peers, as measured by their achievement, influences student *i*; (b) similar students are assigned into the same classroom making the achievement of student *i* and the average achievement peers correlated; (c) student *i* affects her peers; or (d) good (bad) teachers are assigned to the classroom making the achievement of student *i* and the average achievement peers positively (negatively) correlated. (b), (c), and (d) are confounding factors that make it difficult to identify the causal effect of peers on student *i* (i.e., (a)).

If students are randomly assigned into classrooms, an Ordinary Least Squares (OLS) estimate of β_1 in equation (1.1) will not be biased by unobserved selection of students into classrooms. When students are randomly assigned into classrooms, teachers are also effectively randomly assigned into classrooms.[3] Hence, there is no systematic

[1] There are 7 cohorts of students and 2 ability groups for each cohort of students. So k = 13 (excluding the intercept).

[2] The reflection problem is analogous to the problem of interpreting the simultaneous movements of a person and her reflection in a mirror, as it may not be clear to a naïve outside observer whether the mirror image causes the person's movements or simply reflects them.

[3] It is possible that some classrooms have higher average initial achievement than other classrooms do. If teachers are systematically assigned into classrooms according to these differences, then selection bias may still be present.

relationship between the quality of teachers and the initial composition of a classroom. Random assignment of students into classrooms effectively removes what Manski (1993) terms correlated effects (i.e., factor (b) above).

Because student *i* and her classmates simultaneously influence each other, we can also express the achievement of peers as a function of student *i*'s achievement:

$$\bar{y}_{-ickt} = \tilde{\beta}_0 + \tilde{\beta}_1 y_{ickt} + \tilde{\delta}_k + \tilde{\varepsilon}_{ickt} \qquad (1.2)$$

where $\tilde{\beta}_1$ measures the direct effect student *i* on peers. That is, equation (1.2) measures the reverse causation of student *i* on peers or factor (c) listed above. Equation (1.1) and equation (1.2) suggest that student *i* affects her classmates and her classmates affect her at the same time. An OLS estimate of β_1 in equation (1.1) does not show the direct effects of peers on student *i*, but the net effect of their repeated influences. This is essentially the endogeneity problem that Moffitt (2001) discussed. The commonly adopted solution to this simultaneity problem is to replace $\bar{y}_{-ickt}$ in equation (1.2) with the lagged achievement of classmates, $\bar{y}_{-ickt-1}$ (e.g., Betts and Zau 2004; Hanushek et al. 2003; Hoxby and Weingarth 2007; Zimmerman 2003). If students were not repeatedly assigned into the same classrooms in the previous periods, replacing the current average achievement of peers with the lagged average achievement of peers would effectively address the simultaneity problem.[4] Since baseline test scores are determined in elementary school prior to students being randomly assigned into classrooms, this modification effectively removes the feedback effect. Furthermore, because lagged achievement is not affected by the current classroom environment, common shock effect

[4] In the current study, using lagged average achievement of peers is appropriate because students from dozens of different schools were quasi-randomly assigned into different classrooms.

or factor (d) listed above will also be removed by replacing $\bar{y}_{-ickt}$ in equation (1.1) with the lagged achievement of classmates, $\bar{y}_{-ickt-1}$. Regression equation (1.1) is then modified as:

$$y_{ickt} = \beta_0 + \beta_1 \bar{y}_{-ickt-1} + \delta_k + \varepsilon_{ickt} \qquad (1.3)$$

where the estimated β_1 does not capture the reverse causation and common shock effects.

It is important to note that this paper provides a reduced-form estimate of the effect of peers on achievement, as baseline peer achievement is predetermined and may proxy for unobservable peer traits, such as ability, motivation, and effort.[5] In addition, because the response of a teacher may also vary with the effort and ability of students in the classroom, peer achievement will also capture the indirect effects of peers through teachers.

1.3 Mechanical Negative Correlation

With random assignment of students into classrooms, the conventional wisdom indicates that estimating equation (1.3) in OLS should yield a consistent estimate of β_1 (e.g., Lyle 2007). However, Guryan et al. (forthcoming) used Monte Carlo simulations to show that a typical test of random assignment suffers a mechanical negative bias when the sample size is small. Their findings imply that the OLS estimate of β_1 based on equation (1.3) may be underestimated. This section discusses the mechanical negative

[5] Cooley (2009) argues that since peer achievement is a proxy for unobservable peer traits, such as ability, motivation, and effort, equation (1) or (3) does not disentangle what Manski (1993) refers to as endogenous peer effects (e.g., effort) and contextual peer effects (e.g., ability).

bias in a typical test of random assignment and the problem associated with the modified test of random assignment proposed by Guryan et al. (forthcoming). In addition, I show how a mechanical negative correlation may be present when estimating β_1 in equation (1.3) and propose a simple correction method.

1.3.1 A Typical Test of Random Assignment

Guryan et al. (forthcoming) point out that when the sample size N is small a typical test of random assignment of the following form will likely falsely reject the null hypothesis that $\alpha_1 = 0$:

$$y_{ickt-1} = \alpha_0 + \alpha_1 \bar{y}_{-ickt-1} + \delta_{kt} + \varepsilon_{ickt} \qquad (1.4)$$

Guryan et al. demonstrate in Monte Carlo simulations that the OLS estimate of α_1 in equation (1.4) will likely be negative even when peers and outcomes are randomly generated.[6]

To illustrate this problem, consider a simple case when test scores are randomly generated, only one classroom is formed, and we estimate the following model:

$$y_{ic} = \alpha_0 + \alpha_1 \bar{y}_{-i,c} + \varepsilon_{ic} \qquad (1.5)$$

In this case, OLS estimator always yields $\hat{\alpha}_1 = -(N-1)$ and $\hat{\alpha}_0 = N\bar{y}$, because by definition:

$$y_{ic} = N\bar{y} - (N-1)\bar{y}_{-i,c} \qquad (1.6)$$

[6] Lyle (2007) also noted this mechanical negative correlation in a footnote, but he did not investigate its properties. De Giorgi et al. (forthcoming) utilized this mechanical negative correlation to implement an instrumental variable estimator in a different context.

The correlation between y_{ic} and $\bar{y}_{-ic}$ is negative because each y_{ic} in the sample is drawn without replacement. We can also see that the bias does not diminish, but increases with *N*, in this one-classroom problem.

If *C* classrooms of size *n* are formed in the sample of *N* students, the OLS estimator will yield:

$$\hat{\alpha}_1 = \frac{\sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y})(\bar{y}_{-i,c} - \bar{y})}{\sum_{c=1}^{C}\sum_{i=1}^{n}(\bar{y}_{-i,c} - \bar{y})^2} \qquad (1.7)$$

Note that $\bar{y}_{-i,c} = \frac{n\bar{y}_c - y_{ic}}{n-1} = \frac{n}{n-1}\bar{y}_c - \frac{1}{n-1}y_{ic}$, where $\bar{y}_c$ is the mean outcomes of classroom *c* that individual *i* is in, and $N = C \cdot n$.

Table 1.1 illustrates how the mean of $\hat{\alpha}_1$ from regression equation (1.5) is related to various sample sizes (*N*) and class sizes (*n*) based on 10000 repetitions of Monte Carlo simulations where $y_{ic} \sim iid\ N(0,1)$. For example, when *n* is fixed at 2, and *N* increases from 4 to 1024, the average of the estimated slope coefficients decreases from -0.33 to -0.0004 (the first column). When we hold *N* fixed, and increase *n*, the negative relationship gets stronger. When we hold *C* fixed and increase both *N* and *n*, the negative bias also gets more severe. This highlights why in studies that look at college roommates, such as that by Sacerdote (2001), where *n* is 2 and *N* is 1589 students, the test of random assignment fails to reject the null hypothesis that $\alpha_1 = 0$. On the other hand, if a peer effect study uses variation of student characteristics across classrooms within a school that has about 512 to 1024 students and an average class size of 32, the negative bias can be as large as 0.1 to 0.5 standard deviations. This negative bias is comparable to some of the peer effect estimates previously reported. For the negative bias present in a study of classroom peer effects across multiple schools to be as small as the bias present in a study of peer effect looking at 256 pairs of college roommates, the sample size needs to be as large as 100,000 students.[7]

The negative correlation is present because sampling of individuals is done without replacement. For a given sample, the sample mean is known. If an individual's peers have above-average characteristics, the individual must have below-average characteristics. When the number of classrooms in a sample is small, the weight of the average characteristics of an individual's peers on the sample mean is large. For a given sample average, if the peers have above-average characteristics, then the individual's characteristics must be way below the average, leading to a strong negative correlation between the individual's characteristics and the average characteristics of peers. However, as the number of classrooms formed within the sample increases, the weight of the average characteristics of the peers on the sample mean becomes smaller. Indeed, as the number of classrooms approaches infinity, an individual's classmates are similar to a randomly drawn individual in an extremely large sample. Thus, as the number of classrooms increases, the negative correlation between the individual's characteristics and the peers' characteristics becomes weaker and the negative bias diminishes. Appendix 1.1 shows that $\hat{\alpha}_1$ converges in distribution to a random variable with a non-positive mean, as class size $n \to \infty$, while holding the number of classrooms, *C*, fixed. On the other hand, $\hat{\alpha}_1$ converges in probability to zero, as $C \to \infty$, while holding *n* fixed.

---

[7] This means that there are 3125 classrooms with 32 students each. The number of classrooms is similar to that used by Betts and Zau (2004) in their study of classroom peer effects.

The presence of this mechanical negative correlation means that: (i) researchers may find non-positive peer effects when the data used have a small number of groups (classrooms), even if the causal effect of peers is positive; (ii) a typical test of random assignment using baseline test scores will suffer from a negative bias, especially in cases where class size is relatively large or the number of classrooms is small; and (iii) there is a mechanical explanation for why estimated peer effects at the grade level tend to be smaller than estimated peer effects at the classroom level.

1.3.2 A Modified Test of Random Assignment

Guryan et al. (forthcoming) (GKN) propose a modified test of random assignment that does not suffer from this mechanical bias. They argue that the typical test of random assignment used in the study of peer effects can be modified as:

$$y_{ickt-1} = \alpha_0 + \alpha_1 \bar{y}_{-ickt-1} + \alpha_2 \bar{y}_{-ikt-1} + \delta_k + \varepsilon_{ickt-1} \qquad (1.8)$$

where the additional variable $\bar{y}_{-ikt-1}$ is the average test score of all individuals in the same randomization track *k* except individual *i*. They essentially propose including the "one-classroom" measure of peers shown in equation (1.6). I will refer to $\bar{y}_{-ikt-1}$ as "GKN's correction term". GKN's correction method requires that there be multiple tracks of randomization and that the population size differs across tracks (otherwise, OLS estimators will always yield $\hat{\alpha}_1 = 0$ and $\hat{\alpha}_2 = -(N-1)$). They show that by including their correction term in the test of random assignment, the test becomes well-behaved in Monte Carlo simulations, whether the sample size is small or large.

However, when randomization tracks differ considerably in size, GKN's correction term will not fully correct for the negative bias. To illustrate this problem, consider two hypothetical schools, A and B, each of which has two tracks and students are randomly assigned into classrooms (of size 50) within each track. In school A, track 1 has 1000 students and track 2 has 950 students. In school B, track 1 has 1000 students and track 2 has 500 students. Furthermore, assume that the baseline test score of each student y_{it-1} is $iid.\,N(0,1)$. I run the modified test of random assignment, i.e., equation (1.8), in 10000 repetitions of Monte Carlo simulations to evaluate whether the average of $\hat{\alpha}_1$ centers at zero. The averages of $\hat{\alpha}_1$ for school A and school B are reported in column (1) and column (2) of Table 1.2 respectively. Column (1) shows that GKN's correction term successfully removes the negative bias for school A in the test of random assignment and column (2) shows that GKN's correction term fails to remove the negative bias in the test of random assignment for school B. Therefore, the modified test of random assignment proposed by Guryan et al. (forthcoming) is not generally reliable when the randomization tracks do not have similar size.

Since GKN's modified test of random assignment is not always reliable, we need a method that tests whether individuals are randomly assigned into groups and the method will work in small and large samples and whether or not there are multiple randomization tracks. One way to test whether individuals are randomly assigned into groups is to conduct a joint test of significance of average characteristics across groups. If individuals are randomly assigned into groups, then we would not expect systematic differences across groups and the average characteristics of individuals should not be statistically different.

1.3.3 Correction for the Mechanical Negative Bias

Although we may not use the modified test of random assignment to verify whether students were randomly assigned into classrooms, we must still control for the negative bias when estimating peer effects. I demonstrated the mechanical negative correlation between y_{ickt-1} and $\bar{y}_{-ickt-1}$ of equation (1.4) in a previous section. Specifically, when n and N are fixed and finite, it is likely that $\text{cov}(y_{ickt-1}, \bar{y}_{-ickt-1} \mid \delta_{kt}) < 0$. Given this mechanical negative correlation, an OLS estimate of β_1 based on equation (1.3) is also likely to suffer from a negative bias.

Expressing regression equation (1.3) in the following way will help elucidate why the negative correlation must be controlled for:

$$y_{ickt} = \beta_0 + \beta_1 \bar{y}_{-ickt-1} + \delta_{kt} + \beta_2 y_{ickt-1} + v_{ickt} \qquad (1.9)$$

where $\varepsilon_{ickt} = \beta_2 y_{ickt-1} + v_{ickt}$.

Persistence in learning means that $\text{cov}(y_{ickt}, y_{ickt-1} \mid \delta_{kt}) > 0$ and $\beta_2 \neq 0$.[8] When the individual's own baseline test score y_{ickt-1} is omitted, β_1 estimated using regression equation (1.3) will be biased downward as $\text{cov}(\bar{y}_{-ickt-1}, \varepsilon_{ickt} \mid \delta_{kt}) > 0$. Therefore, y_{ickt-1} must be included as a regressor to control for the mechanical negative bias, even in the case when students are randomly assigned into classrooms. Because the random assignment of individuals into classrooms ensures that $E(v_{ickt-1} \mid \bar{y}_{ickt-1}) = 0$, the

[8] For example, see Andrabi et al. (2009) for a discussion on learning persistence.

estimated β_1 is consistent once y_{ickt-1} is controlled for. Note that the estimated β_1 based on equation (1.9) follows the reduced-form interpretation of peer effects.

The following simulation helps demonstrate the effectiveness of an individual's own baseline test score in absorbing the negative bias. Assume that peer effects are absent and that each student's outcome is solely determined by her own baseline test score in the following form :

$$y_{it} = 0.8 y_{it-1} + 0.6 u_{it} \qquad (1.10)$$

where $u_i \sim iid.\ N(0,1)$. Note that parameter values are arbitrarily chosen to ensure $y_{it} \sim N(0,1)$. If an individual's own baseline test score is an effective control variable for the mechanical negative bias, we would expect the following regression:

$$y_{ickt} = \alpha_0 + \alpha_1 \bar{y}_{-ick,t-1} + \alpha_2 y_{ickt-1} + \delta_k + v_{ickt} \qquad (1.11)$$

to yield $\hat{\alpha}_1 = 0$ on average. On the other hand, if peer effects are present such that:

$$y_{it} = 0.5 \bar{y}_{-i,t-1} + v_{it}, \qquad (1.12)$$

then we would expect the OLS estimator to yield $\hat{\alpha}_1 = 0.5$ on average.

Table 1.3 columns (1) and (2) report the average of $\hat{\alpha}_1$ as assumed in equation (1.10) and equation (1.12) respectively based on 10000 repetitions of regression equations (1.11) for a hypothetical school with 1500 students and 30 classrooms. Columns (1) and (2) show that the average of $\hat{\alpha}_1$ corresponds to the values specified in equation (1.10) and equation (1.12), respectively. The Monte Carlo simulation results provide support for using own baseline test scores of student *i* to control for the mechanical negative correlation.

To summarize, this section examined the mechanical negative correlation inherent in a typical linear regression framework used to estimate peer effects. I also show that the typical test of random assignment and the modified test of random assignment used by previous peer effect studies are not always reliable. Nevertheless, we can use a joint test of average characteristics across groups to verify whether individuals are randomly assigned into groups. Because of the mechanical negative correlation, it is important to include a student's own baseline test score as a regressor.

1.4 Institutional Background and Data Description

This paper draws data from a large independent Chinese secondary school in Johor, Malaysia. There are generally two types of secondary schools in Malaysia: public schools and independent Chinese schools.[9] Public secondary schools are run by the government and the language of instruction is primarily Malay.[10] Independent Chinese secondary schools are fee-paying schools run by ethnic Chinese communities. They generally follow curricula set by the United Chinese School Committees Association of Malaysia. The primary language of instruction in these schools is Chinese (Mandarin), although some schools use English textbooks for some subjects. These schools charge relatively low fees, have large class sizes, and rely heavily on donations from local

[9] There are also a small number of private schools that are run by private entities. These so called Private Educational Institutions (PEIs) provide education at preschool, elementary, and secondary levels. Some of these schools follow public school curricula and some follow foreign educational systems.

[10] Public secondary schools are also called national secondary schools. Since 2003, the medium of instruction in math and science has been English, but the government recently announced that the medium of instructions in math and science would be switched to Malay.

Chinese communities.[11] Independent Chinese secondary schools divide grade levels into lower division and upper division. Lower division spans from grade 7 to grade 9 while upper division spans from grade 10 to grade 12.

The sample school admits roughly 1000 ethnic Chinese students each year who attended public Chinese elementary schools in the metropolitan area.[12] The Malaysian government halted the opening of new independent Chinese secondary schools in the 1950s, causing excess demand for Chinese secondary education in Malaysia. The school uses standardized admission examinations to ration student slots. A student's aggregate test score in four subjects, Chinese, English, Malay, and Math, determines whether she is admitted into the school or not. Students scoring below a cutoff point are not admitted, but the school accepts transfer students from other independent secondary schools conditional on the students' past achievement.[13]

This paper analyses the effects of peers in grade 7 for academic years 2002 to 2008 using a sample of 6618 students.[14] Grade 7 students in the secondary school are tracked into one of two groups (group A or group B) on the basis of their admission (aggregate) test scores. Generally, the top 300 students are tracked into group A, and the rest, on average about 700, into group B.[15] The average class size at time of assignment is 50 students, but not all students assigned into classrooms subsequently attend the

[11] These features are similar to the Catholic schools in the United States and other countries. Ten percent of the students in the sample school receive some form of financial aid or scholarship.
[12] Less than 1 percent of the student population at this school is non- Chinese.
[13] It is extremely rare to have transfer students in grade 7.
[14] The school began using admission examinations in 1979, but records of admission examination and test takers before academic year 2002 were lost.
[15] The number of group B classes varies from year to year, ranging from 13 to 19 in the sample.

school.[16] Within each track, ability mixing is enforced through a quasi-randomization process. For example, if the male-female ratio is 2:3 in group A, the school will set the number of males per class to be 20, and the number of females per class to be 30. The school will then assign the top female scorer on the admission exam to class 1, the second-best female to class 2, the 6th female to class 6, the 7th female to class 6, the 8th female to class 5, and so forth. Male students are also assigned in the same way.[17] In group B, there are usually 20 to 40 grade repeaters from the previous academic year and the school uses their test scores in grade 7 as the basis for the quasi-random assignment process. Because the school attempts to equalize average test scores across classrooms, selection bias is controlled for. This means that teacher assignment within each track is also effectively free of selection bias.

Students attend all classes with the same set of peers during the school year.[18] This greatly reduces the opportunities for students in different classrooms, especially those that never met prior to secondary school, to interact.[19] This setup is unusual compared to secondary schools in most other countries where students attend different classes with different sets of peers. The quasi-random assignment of racially homogenous students of the same age into classrooms, where students attend all classes with a fixed set of classmates, provides a unique setting for the study of classroom-level peer effects.

---

[16] According to the school administrators, some students who registered for enrollment did not subsequently attend the school because they were admitted into secondary schools in Singapore with generous scholarship awards or were admitted to elite local public secondary schools. As a result, there are generally larger variations in class size than the initial assignment shows.

[17] The assignment is done by a computer program. This classroom assignment method is similar to the one practiced in South Korea as described by Kang (2007).

[18] Students may participate in extracurricular activities with students in other classrooms and grade levels. Because extracurricular activities are optional, only approximately 60% of students participate in them.

[19] There are roughly 30 feeder elementary schools in the city.

Students are regularly given standardized tests during a school year. The school calculates mid-year and end-of-year test scores based on the weighted average of all test scores for the respective semester. The tests are standardized within each track for some subjects and across tracks for others, but are graded by teachers of the respective classes.[20] However, in Chinese, English, and Malay classes, because test scores include grades in writing assignments administered and graded by respective teachers, test scores in these subjects are less comparable across classrooms. In this study, I focus on math scores as the outcome measure of achievement since math tests are standardized across classrooms within the same track and there is less room for differences in grading standards across teachers.[21] I report estimates of peer effects on Chinese, English, and Malay in Appendix 1.2. The final grade at the end of the school year will determine whether students are promoted to group A, demoted to group B, or forced to repeat the same grade in the following year. Because students are reshuffled across tracks and classrooms in grade 8, I restrict the analysis to grade 7. The school also maintains records of students' attendance and number of school discipline codes violated, which can be used as measures of student outcomes.[22] Descriptive statistics of 6618 grade 7 students in 138 classrooms from academic year 2002 to academic year 2008 are reported in Table 1.4.

---

[20] For example, math tests are standardized within each ability group, but not across ability groups. Typically, there are two parts to a test, where the first part involves multiple choice questions and the second part involves short answers.

[21] The fact that teacher assignment is effectively randomized means that any grading differences will add noise but not bias to estimates.

[22] Teachers take attendance throughout the day for every single class. Academic penalty in the form of final grade point deduction and disciplinary penalty in the form of discipline point deduction may be imposed on students who miss classes. Students may be subject to probation or suspension if their cumulative discipline point deductions reach a certain threshold. Students who never miss a single class during the school year are given honors, which can offset any disciplinary penalties they received.

1.5 Evidence of Quasi-Random Assignment into Classrooms

To demonstrate that students were not selectively assigned into classrooms, I run the following regressions using baseline test scores as the dependent variable:

$$y_{ickt-1} = C_c + \delta_k + \varepsilon_{ickt-1} \qquad (1.13)$$

where C_c is a set of classroom fixed effects. If students were assigned into classrooms in the way that the school claimed, all classroom means should be statistically similar. This is equivalent to testing whether C_c are individually and/or jointly different from zero. This is not the typical test of random assignment that other studies of peer effects used (e.g., Sacerdote 2001), but it is a reasonable alternative as it does not suffer from a negative bias.

Table 1.5 reports the joint tests of whether at least one of the classroom fixed effects is significantly different from zero, as well as the fraction of classroom fixed effects having individual p-values less than certain thresholds, using different baseline test scores as dependent variables. The top panel in Table 1.5 shows that we cannot reject the null hypothesis that all classroom fixed effects are jointly equal to one another. The bottom panel in Table 1.5 demonstrates that the fraction of individual significant classroom effects is close to what the size of the test suggests.

Given that baseline test scores are statistically similar across classrooms, one may be worried that there is not enough cross-classroom variation in baseline test scores for the identification of peer effects. Table 1.6 reports the summary statistics for classroom-level average test scores after conditioning on the set of cohort-ability group fixed effects.

That is, each observation is a classroom's average of the regression residuals across individuals based on the equation $y_{ickt-1} = \hat{\hat{\delta}}_k + \hat{\varepsilon}_{ickt-1}$. The standard deviations here provide some assurance that there is still a reasonable amount of variation in the data.

1.6 Results

1.6.1 Math Achievement: Linear-in-Means Specifications

Columns (1) to (8) in Table 1.7 presents OLS estimates of the effect of peers on a student's own math score based on equation (1.9). Columns (1) to (4) report estimates using math score in semester 1 as the dependent variable, while columns (5) to (8) report estimates using math score in semester 2 as the dependent variable.

Columns (1) and (5) show that ignoring the existence of two ability groups and not controlling for the mechanical negative correlation, the effect of peers on math achievement is statistically positive and large, highlighting the bias associated with selection into ability groups. Columns (2) and (6) show that without controlling for the mechanical negative correlation, but including a set of cohort-ability group fixed effects will still yield an estimated peer effect not statistically different from zero. The cohort-ability group fixed effects control for the selection into different ability groups, but not the mechanical negative correlation. Columns (3), (4), (7) and (8) show that using a student's own baseline test scores as control variables leads to significant larger estimates of peer effects. The results indicate that peer effects are statistically significant in semester 2, but not in semester 1, suggesting that it takes time for peer effects to result in observable changes in own outcomes. According to the preferred specifications, i.e., column (4) and column (8), a student's own math score is predicted to increase by 0.28

standard deviations in semester 1 and 0.66 standard deviations in semester 2 for every one standard deviation increase in the average baseline math scores of peers.

Table 1.8 compares the current estimate of peer effects on math achievement with past estimates of peer effects on math achievement from other studies. Column (1) reports the estimate presented in column (8) of Table 1.7. Column (2), column (3) and column (4) report the estimated effect of peers on math achievement from Duflo et al.'s (forthcoming) sample of Kenyan first graders, Kang's (2007) sample of Korean seventh and eighth graders, Hanushek et al.'s (2003) sample of elementary school students in Texas. The current estimate is much larger than Hanushek et al.'s (2003) and Kang's (2007) estimates, but closer to Duflo et al.'s (forthcoming) estimate.[23] The peer effect estimate in this paper implies that any policy intervention may generate large spillover effects through interaction between students.

In summary, I show that when we effectively control for unobserved selection and the mechanical negative correlation, the estimated effect of peers on an individual's math achievement is significantly positive and large. The size of the current estimate is somewhat larger than most previous estimates.

1.6.2 Class Absences and Discipline Violations

The estimates of peer effects reported above indicate that the effects of peers are strong. Besides learning from peers, possible mechanisms for peer effects are through effort and conduct. To assess whether having better achieving peers induces a student to

[23] Since the variation in average baseline score of classmates within each track is roughly between 0.09 and 0.14, the predicted effect of peers due to a one-standard deviation increase in average baseline score of classmates is likely unrealistic.

increase her effort, I investigate whether the average baseline test scores of peers predicts effort, using student's absences as a proxy for own effort. Similarly, I examine whether having high achieving peers lower a student's total number of discipline violations cited. Measured discipline violations may range from minor incidents, such as classroom disruption and dress code violations, to serious incidents, such as academic dishonesty, fighting, and theft. We would expect that having better achieving peers will lower absence rates and discipline violation rates.

Table 1.9 reports the estimated effects of peers on class absences and discipline violations. Columns (1) and (2) report the effect of peers on the number of classes missed in semester 1; columns (3) and (4) report the effect of peers on the number of classes missed in semester 2; columns (5) and (6) report the effect of peers on the total number of classes missed during the academic year; columns (7) and (8) report the number of discipline violations cited during the academic year. Estimates reported in even numbered columns are based on the specification that includes a set of cohort-ability group fixed effects, but those reported in odd numbered columns are based on the specification that contains a set of cohort fixed effects. The contrast between the estimated peer effects reported in odd numbered columns and even numbered columns shows that using lagged achievement to control for both the selection into different ability groups and the mechanical correlation may still lead to an underestimation of peer effects.

The estimates indicate that high achieving peers have strong negative effects on class absences and discipline violations. For every one standard deviation increase in the average baseline scores of peers, the number of classes missed by a student during a school year is predicted to fall by 14.35. This is equivalent to 2 days of school.[24] This effect seems small, but given that the mean is 10, it is actually quite large. This effect of a one standard deviation increase in the average baseline score of classmates is approximately a 0.88 standard deviation decrease in total number of classes missed per year. The effect of peers on discipline violations is even stronger. A one standard deviation increase in the average baseline scores of peers is associated with a 1.05 decrease in the number of discipline violations cited. This effect is equivalent to roughly 1.58 standard deviation of change in the incidence of discipline violations.

The results presented here imply that having higher achieving peers can alter a student's effort and conduct. Since attendance and discipline are linked to future earnings (Angrist and Krueger 1992) and crime (Adams 2000), respectively, the results also suggest other potential longer term effects of peers.

1.6.3 Differential Peer Effects and Classroom Heterogeneity

The strong positive effect of peers estimated in the previous section suggests that grouping high-achieving students with low-achieving students will benefit the latter at the cost of the former. However, because the effect is estimated using a linear-in-means specification, it does not indicate whether grouping high-achieving students with low-achieving students will increase or decrease the average achievement of a school. For instance, if the peer effect is stronger for low-achieving students than is for high-achieving students, then ability mixing will raise average achievement of a school. In contrast, if classroom heterogeneity hurts student achievement, then ability grouping may

[24] A typical school day has 7 classes.

raise average achievement. I consider a number of non-linear specifications to assess whether peer effects vary for different types of students and whether classroom heterogeneity matters to student achievement.

First, I use the following non-linear specification, which allows the effect of peers to vary for students at different ranges of the initial test score distribution, to estimate whether the effects of peers vary differentially for different types of students:

$$y_{ickt} = \beta_1 I^{Bottom_1}_{ickt-1} + \beta_2 I^{Middle}_{ickt-1} + \beta_3 I^{Top}_{ickt-1} + \gamma_1 I^{Bottom}_{ickt-1}\bar{y}_{-ickt-1} + \gamma_2 I^{Middle}_{ickt-1}\bar{y}_{-ickt-1} + \gamma_3 I^{Top}_{ickt-1}\bar{y}_{-ickt-1}$$

$$+ \pi_1 I^{Bottom}_{ickt-1} y_{ickt-1} + \pi_2 I^{Middle}_{ickt-1} y_{ickt-1} + \pi_3 I^{Top}_{ickt-1} y_{ickt-1} + \delta_k + \varepsilon_{ickt} \qquad (1.14)$$

Note that I^{p}_{ickt-1} is an indicator of whether student *i*'s baseline math score is in the *p* range of the test score distribution and *p* is either bottom 25 percentile, middle 50 percentile, or top 25 percentile. Regression equation (1.14) allows the effect of average past achievement of peers to vary for students at different ranges of the initial test score distribution. Because the school tracks students into two ability groups based on their aggregate baseline test scores in all subjects, there are students who scored in the top 25 percentile of the baseline math score that ended up in the low ability group and students who scored in the bottom 25 percentile of the baseline math score that ended up in the high ability group. Unlike past studies of peer effects using this particular type of non-linear specification, a set of interaction terms $I^{p}_{ickt-1} y_{ickt-1}$ is included to correct for mechanical correlations. Appendix 1.3 shows Monte Carlo simulation results that compare the bias in cases where these correction terms are excluded and included, respectively.

We may also modify equation (1.14) by including the classroom standard deviation of baseline math score, s_{ct-1}, as a measure of classroom heterogeneity:

$$
\begin{aligned}
y_{ickt} = {} & \beta_1 I_{ickt-1}^{Bottom_1} + \beta_2 I_{ickt-1}^{Middle} + \beta_3 I_{ickt-1}^{Top} + \gamma_1 I_{ickt-1}^{Bottom} \bar{y}_{-ickt-1} + \gamma_2 I_{ickt-1}^{Middle} \bar{y}_{-ickt-1} + \gamma_3 I_{ickt-1}^{Top} \bar{y}_{-ickt-1} \\
& + \theta_1 I_{ickt-1}^{Bottom} s_{ckt-1} + \theta_2 I_{ickt-1}^{Middle} s_{ckt-1} + \theta_3 I_{ickt-1}^{Top} s_{ckt-1} \\
& + \pi_1 I_{ickt-1}^{Bottom} y_{ickt-1} + \pi_2 I_{ickt-1}^{Middle} y_{ickt-1} + \pi_3 I_{ickt-1}^{Top} y_{ickt-1} + \delta_k + \varepsilon_{ickt} \qquad (1.15)
\end{aligned}
$$

If classroom heterogeneity impedes achievement, then we would expect θ_p to be negative for $p = Top, Middle,$ or $Bottom$. Summary statistics of test scores and average peers' characteristics by the range of students' initial math score distribution are presented in Appendix 1.4.

Table 1.10 reports the non-linear estimates of peer effects based on equation (1.14) and equation (1.15). The preferred specification shows that for every one standard deviation increase in the average baseline math score of classmates, the math achievement of students whose baseline math scores were in the bottom 25 percentile of the test score distribution is predicted to increase by 0.23 standard deviations in semester 1 (column 2) and 0.66 standard deviations in semester 2 (column 4). For students whose initial math scores were in the middle 50 percentile, the effect of peers on math achievement is 0.36 standard deviations in semester 1 and 0.73 standard deviations in semester 2. For students in the top 25 percentile, the effect of peers on math achievement is estimated to be 0.12 standard deviations in semester 1 and 0.53 standard deviations in semester 2. Note that without controlling for baseline test scores (columns 1 and 3), peer effects are more likely underestimated for students in the middle and top of the initial test score distribution. The joint tests of differential peer effects reported in the bottom panel suggest that the effect of peers differ between top students and middle students in semester 2 and the effect of peers is strongest for students in the middle 50% of the distribution and weakest for students in the top 25% of the distribution. The estimates of classroom heterogeneity effect reported in column (3) and column (6) show that larger variation in the initial math scores of students in the same classroom has no significant effect on a student's subsequent math achievement. However, the variation in average baseline math scores across classrooms may be too small to allow detection of any effect.

One problem with equation (1.14) is that the effect of peers is assumed the same whether the students are in the top ability group or bottom ability group. It is possible that peer effects vary considerably for the same type of students depending on the quality of peers measured by their initial achievement. We can modify equation (1.14) by interacting ability group indicator with whether the student is in the top, middle, or bottom of the initial math score distribution. This modification is equivalent to estimating equation (1.14) separately for students in the high-achieving group (group A) and students in the low-achieving group (group B). This specification is feasible because students were assigned into different ability groups based on their aggregate baseline test scores in all subjects, rather than baseline math score alone. The estimates are reported in Table 1.11.[25]

Column (1) reports the estimates for students in the low-achieving group and column (2) reports the estimates for students in the high-achieving group. Column (3) reports the differences in peer effects between ability groups for each type of students.

[25] Summary statistics of test scores and peer characteristics by ability groups and student types are reported in Append 1.4.

Column (1) shows that among the students in the low-achieving group, the peer effect is strongest for the top students and weakest for the bottom students. The estimates presented in column (2) indicate that among the students in the high-achieving group, peer effects are similar across different types of students. However, because there are fewer classrooms in this track, the estimates are not precisely estimated.[26]

The estimates reported in Table 1.11 show that students in the top of the initial test score distribution will gain significantly from having better achieving peers if they are in the low-achieving track. However, the effect of peers on these top students diminishes considerably if they are placed in the high-achieving group. Similar effects are present for students in the middle of the initial test score distribution, although the gain from peer effects for them is weaker when compared to students in the top of the initial test score distribution. More importantly, top students gain more from having better achieving peers than middle students do in both low-achieving and high-achieving tracks. Comparing the means of the average peers' baseline math scores of top 25% and middle 50% students in the high-achieving group and the low-achieving group reveals that peer effects are strong, especially when peers' initial achievement is significantly lower than their own (Table A1.4.2 and Table A1.4.3). In contrast, students in the bottom of the initial test score distribution experience a small effect of peers regardless of their ability group assignment. Although the estimate is noisier for these students, the small estimated peer effect and the negligible differences between peer effects in high-achieving and low-achieving groups suggest that students at the bottom of the initial test score distribution are less likely to gain from mixing with better achieving students.

[26] There are 42 classrooms for the 7 cohorts of students.

Therefore, it appears that separating all top students into one group will improve their average achievement more than the loss (if any) suffered by students in the bottom of the initial test score distribution. Similarly, separating top students from middle students will also help the former more than it hurts the latter. Since the effect of peer is weak for students in the bottom 25% of the initial test score distribution, ability grouping will not decrease their achievement much and will raise the average achievement of the school.

The results presented in this section show some evidence of non-linear peer effects and indicate that ability grouping may help raise average math achievement. Peer effects are weak for students in the bottom of the initial test score distribution, moderate for students in the middle of the initial test score distribution, and strong for students in the top of the initial test score distribution. The effect of peers is particularly strong when the average baseline math score of classmates is significantly lower than the student's own. Finally, ability grouping will likely increase average achievement, and not hurt low-achieving students much.

1.7 Conclusion

This paper provides two major contributions to the study of peer effects. First, it discusses identification problems associated with estimating the causal effect of peers and examines the mechanical negative correlation inherent in a typical study of peer effects. Second, it exploits the quasi-random assignment of students into classrooms in a large secondary school in Malaysia to estimate the causal effects of peers on math score, class absences, and discipline violations, as well as to investigate whether the effects of peers vary for different types of students. Because the secondary school attempts to equalize average baseline test scores across classrooms and students attend all classes with the same set of peers, the data provide an unusual setting to estimate classroom peer effects that minimizes biases resulting from non-random selection.

I examine the mechanical negative correlation between the characteristic of an individual and the average characteristics of the individual's peers. Specifically, I show that the size of the negative bias is large when class size is large or when the number of classrooms is small. This mechanical negative correlation arises because individuals are sampled without replacement. For a given sample average, if an individual's peers have above-average ability, the individual's ability must be significantly below the sample average by construction. When only a few classrooms are formed in the sample, the individual's peers are weighted heavily in the calculation of the sample average, which means that the individual must deviate more strongly away from the peers. The presence of this mechanical negative correlation means that a typical regression for estimating peer effect and the typical test of random assignment of peers may suffer from a negative bias. I also illustrate that the modified test of random assignment proposed by Guryan et al. (forthcoming) is not reliable when randomization tracks differ considerably in size. Since a typical test of random assignment is not generally reliable, researchers may test whether pre-determined characteristics of individuals significantly differ across classrooms to verify whether students are randomly assigned into classrooms. Most importantly, I demonstrate that including an individual's own baseline test score as a regressor is sufficient to control for the mechanical negative bias. This means that even when students are randomly assigned into classrooms, omitting lagged achievement can lead to the underestimation of peer effects.

I find that a one standard deviation increase in the average baseline math score of classmates leads to a 0.66 standard deviation increase in a student's current math score. The estimated effect of peers on math score is somewhat larger than what Duflo et al. (forthcoming) estimated for Kenyan first graders. The large estimated effect of peers implies that any policy intervention is likely to have strong spillover effects through peers' interaction. The current estimates are the first reported for middle income countries and are also larger than those estimated in most other studies in developed countries. For example, Kang (2007) reported that a one standard deviation increase in average peers' current math score is associated with a 0.31 standard deviation increase in a student's own math score in South Korean middle schools; and Hanushek et al. (2003) reported grade-level peer effects on math score gain of 0.18 standard deviations in Texas elementary schools.

One potential explanation for the strong effect estimated in this paper is the amount of classroom time that students in the current sample spent together. These students attended all classes with the same set of peers for 5 to 6 hours per day, 6 days a week, for about 35 weeks a year. Math classes account for roughly 1/6 of all classes. Second, because students in the sample are racially homogenous and about the same age, estimates are less likely to be confounded by race and age differences. This is consistent with Cooley's (2008) findings that heterogeneity in the racial mix of a classroom can potentially confound the estimated effect of average peers' achievement on a student's own achievement. Third, the use of baseline test scores to control for the mechanical negative correlation and the observation of all individuals in a class may perhaps explain why the estimate is greater than what Kang (2007) found in South Korean middle schools, where the classroom assignment method and institutional setting are similar. Fourth, because of the mechanical negative correlation, I show that including individual's lagged achievement as a regressor to address for selection into different ability groups may even underestimate the effect of peers.

The estimated peer effect is large. Having high-achieving peers may lead to better classroom discussions, make teachers better engage in teaching, induce students to increase effort, and discourage students from mischief. To assess whether having high-achieving peers induces a student to increase effort and facilitates better behavior, I examine whether the average baseline test scores of classmates has an effect on both the number of classes a student missed and on the incidence of discipline violations cited. I find that a one standard deviation increase in the average baseline scores of classmates leads to roughly 2 days of classes a student missed and a 1.58 standard deviation decrease in the incidence of discipline violations. This finding demonstrates having higher achieving peers can alter a student's effort and conduct and provides information on the effects of peers on measures that are related to earnings and criminal behavior.

I also investigate whether the effects of peers vary for different types of students. The results provide evidence that non-linear effects of peers are present and indicate that ability grouping may help raise average math achievement. In particular, peer effects are weak for students in the bottom 25 percentile of the initial test score distribution, moderate for students in the middle 50 percentile of the initial test score distribution, and strong for students in the top 25 percentile of the initial test score distribution. Students in the top 25 and middle 50 percentiles gain significantly from having better achieving classmates when the average baseline achievement of their classmates is low. On the other hand, students in the bottom 25 percentile of the initial test score distribution experience small effect of peers regardless of the average achievement of their peers. Hence, ability grouping will likely increase average achievement, and not hurt low-achieving students much.

CHAPTER 1 – TABLES

Table 1.1: The Mean of the OLS Estimator $\hat{\alpha}_1$ by Different Sizes of *N* and *n*

Sample Size (*N*)	Class size (*n*) 2	4	8	16	32	64	128	256	512
2	-1								
4	-0.33 (0.006)	-3							
8	-0.15 (0.005)	-1.25 (0.012)	-7						
16	-0.07 (0.003)	-0.55 (0.008)	-2.59 (0.025)	-15					
32	-0.03 (0.002)	-0.26 (0.005)	-0.93 (0.014)	-4.51 (0.05)	-31				
64	-0.02 (0.002)	-0.11 (0.003)	-0.37 (0.007)	-1.32 (0.02)	-7.46 (0.09)	-63			
128	-0.01 (0.001)	-0.06 (0.002)	-0.17 (0.004)	-0.47 (0.01)	-1.71 (0.03)	-11.62 (0.17)	-127		
256	-0.003 (0.001)	-0.03 (0.002)	-0.08 (0.003)	-0.20 (0.01)	-0.52 (0.010)	-1.91 (0.04)	-17.00 (0.31)	-255	
512	-0.003 (0.001)	-0.01 (0.001)	-0.04 (0.002)	-0.09 (0.003)	-0.21 (0.01)	-0.57 (0.01)	-2.40 (0.06)	-25.70 (0.54)	-511
1024	-0.0004 (0.0004)	-0.01 (0.0008)	-0.02 (0.001)	-0.04 (0.002)	-0.10 (0.003)	-0.23 (0.01)	-0.59 (0.01)	-2.27 (0.07)	-37.03 (0.93)

Note: $y_{ic} \sim iid\ N(0,1)$. Coefficients are the means of $\hat{\alpha}_1$ over 10,000 repetitions of Monte Carlo simulations. Standard errors are reported in parentheses.

Table 1.2: Bias when GKN's Correction Term is Used

	School A	School B
Mean of $\hat{\alpha}_1$ with GKN's correction included as a regressor	0.00004 (0.0001)	-0.007 (0.001)***
Track 1 size (N_1)	1000	1000
Track 2 size (N_2)	950	500

Note: Regression specification includes an intercept term, average $y_{-i,t-1}$, a track indicator, and GKN's correction term. Class size is 50 students. Coefficients are the average of slope estimates of regression equation (1.8) based on 10000 repetitions of Monte Carlo simulations. Standard errors reported in parentheses. *** significant at 1%

Table 1.3: Mean of Estimated Slope Controlling for Own Baseline Score

	(1)	(2)
Mean of $\hat{\alpha}_1$ with own lagged achievement as a regressor	0.0003 (0.001)	0.50 (0.002)
The true parameter value	0.00	0.50

Note: Regression specification includes an intercept term, average $y_{-i,t-1}$, and own lagged achievement $y_{-i,t-1}$ as a control variable. Coefficients are the average of slope estimates of regression equation (1.11) based on 10000 repetitions of Monte Carlo simulations. Sample size of each regression is 1500 students and class size is 50 students. Standard errors reported in parentheses.

Table 1.4: Descriptive Statistics

Variable		Mean	Std. Dev.	Min	Max
Mathematics	- Baseline	0.556	0.807	-2.501	4.176
	- Semester 1	0.002	0.999	-4.687	2.410
	- Semester 2	0.002	0.999	-3.643	2.249
Chinese	- Baseline	0.534	0.737	-3.265	2.835
	- Semester 1	0.001	0.998	-5.531	2.286
	- Semester 2	0.001	0.999	-6.778	2.243
Malay	- Baseline	0.473	0.755	-2.203	2.823
	- Semester 1	0.002	0.998	-5.067	2.363
	- Semester 2	0.002	0.998	-5.244	2.210
English	- Baseline	0.574	0.687	-1.713	2.592
	- Semester 1	0.002	0.998	-3.983	2.586
	- Semester 2	0.002	0.998	-4.656	2.650
Class Absences	- Semester 1	3.802	9.495	0	217
	- Semester 2	6.144	11.484	0	471
	- Full Year	9.946	16.297	0	514
Discipline Violations	- Full Year	0.175	0.664	0	9
Ave. Peers' Baseline Scores					
Mathematics		0.556	0.437	-0.145	1.485
Chinese		0.534	0.421	-0.041	1.364
Malay		0.473	0.506	-0.164	1.508
English		0.574	0.453	-0.030	1.493
Aggregate		0.621	0.524	0.013	1.641
SD of Ave. Classroom Baseline Math		0.677	0.104	0.462	0.996
Female		0.512	0.500	0	1

Note: Sample includes 6,618 grade 7 students enrolled in academic years 2002 to 2008. There are 138 classrooms in the sample. Students who dropped out are excluded. 14 students without baseline test scores available are also excluded from the sample. Baseline test scores are standardized across all test takers of each academic year. Test scores in semester 1 and semester 2 are standardized across all junior 1 students of each academic year.

Table 1.5: Verification of Quasi-Random Assignment

	Baseline Test Scores			
	Math	Chinese	Malay	English
F-statistics of Classroom Fixed Effects	0.84	0.84	0.72	0.81
(p-value)	(0.90)	(0.90)	(0.99)	(0.94)
Fraction of Significant Classroom Fixed Effects:				
p-value < 0.01	0.00	0.00	0.00	0.01
p-value < 0.05	0.02	0.02	0.02	0.02
p-value < 0.10	0.05	0.07	0.06	0.05
Total Number of Classroom Fixed Effects	124	124	124	124

Note: The sample includes 6618 grade 7 students enrolled in 138 classrooms and all specifications include 14 cohort-track fixed effects. Robust standard errors are reported in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%

Table 1.6: Summary Statistics of Mean Baseline Test Scores across Classrooms

Variable	Obs.	Mean	Std. Dev.	Min	Max
Residual baseline scores					
Mathematics	138	-0.00003	0.0882	-0.233	0.279
Chinese	138	-0.00025	0.0749	-0.205	0.165
Malay	138	-0.00014	0.0647	-0.136	0.157
English	138	-0.00005	0.0627	-0.160	0.177

Note: These statistics were obtained in a few steps. First, regression residuals were calculated using estimates from the equation: $y_{ickt-1} = \delta_k + \varepsilon_{ickt-1}$. Then, classroom-level averages were calculated and used to produce statistics presented above.

Table 1.7: OLS Estimates of Peer Effects on Math Score

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Semester One				Semester Two			
Ave. Peers' Math $_{t-1}$	0.61	0.05	0.26	0.28	0.82	0.50	0.65	0.66
	$(0.06)^{+}$	(0.29)	(0.25)	(0.24)	$(0.06)^{+}$	(0.36)	(0.33)*	(0.32)**
Own Math $_{t-1}$			0.77	0.76			0.57	0.57
			$(0.02)^{+}$	$(0.02)^{+}$			$(0.02)^{+}$	$(0.02)^{+}$
Own Chinese $_{t-1}$				0.20				0.14
				$(0.02)^{+}$				$(0.02)^{+}$
Own Malay $_{t-1}$				0.17				0.18
				$(0.02)^{+}$				$(0.02)^{+}$
Own English $_{t-1}$				0.12				0.10
				$(0.02)^{+}$				$(0.02)^{+}$
Cohort-track F.E	--	Yes	Yes	Yes	--	Yes	Yes	Yes
Cohort F.E	Yes	--	--	--	Yes	--	--	--
Observations	6618	6618	6618	6618	6618	6618	6618	6618
R-squared	0.07	0.09	0.37	0.40	0.13	0.14	0.29	0.31

Note: All specifications include a set of cohort x track fixed effects unless otherwise stated. Robust standard errors clustered at the classroom level are reported in parentheses. The sample includes 138 classrooms. $^{+}$ significant at 1%; ** significant at 5%; * significant at 10%

Table 1.8: A Comparison of Peer Effect Estimates of Various Studies

	(1) Current Study	(2) Duflo et al. (forthcoming)	(3) Kang (2007)	(4) Hanushek et al. (2003)
Peer Effect Estimate	0.66**	0.47***	0.31**	0.18***
Data Type	Quasi-random Assignment	Random Assignment	Quasi-random Assignment	Observational Data
Dependent Variable	Math Score	Math Score	Math Score	Math Score Gain
Peer Measure	Lagged Score	Lagged Score	Current Score	Lagged Score
Control for Own Lagged Score	Yes	Yes	No	No
Estimation Method	OLS	OLS	IV	Student F.E.
Grade Levels	7^{th} Grade	1^{st} Grade	7^{th} & 8^{th} Grade	3^{rd} to 8^{th} Grade
Country	Malaysia	Kenya	Korea	US - Texas

Note: The estimate in column (1) is sourced from column (8) of Table 1.7; the estimate in column (2) is sourced from column (2) of Table 4 in Duflo et al. (forthcoming); the estimate in column (3) is sourced from column (5) of Table 5 in Kang (2007); and the estimate in column (4) is sourced from column (1) of Table II in Hanushek et al. (2003). *** significant at 1%; ** significant at 5%; * significant at 10%

Table 1.9: OLS Estimates of Peer Effects on Absences and Discipline Violations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	---------------------------- Class Absences ----------------------------						Discip. Violations	
	- Semester One -		- Semester Two -		---- Full Year ----		---- Full Year ----	
Peers' Score $_{t-1}$	-0.43	-2.04	-0.22	-12.30	-0.65	-14.35	-0.02	-1.05
	(0.40)	(3.07)	(0.54)	(4.15)$^{+}$	(0.73)	(5.27)$^{+}$	(0.02)	(0.36)$^{+}$
Own Math $_{t-1}$	-0.30	-0.30	-0.07	-0.11	-0.36	-0.40	-0.02	-0.02
	(0.15)**	(0.15)*	(0.18)	(0.19)	(0.26)	(0.26)	(0.01)*	(0.01)**
Own Chinese $_{t-1}$	-0.30	-0.29	-0.44	-0.51	-0.75	-0.81	-0.06	-0.07
	(0.20)	(0.20)	(0.22)*	(0.22)**	(0.34)**	(0.33)**	(0.02)$^{+}$	(0.02)$^{+}$
Own Malay $_{t-1}$	-0.20	-0.20	-0.29	-0.33	-0.48	-0.53	-0.08	-0.09
	(0.27)	(0.27)	(0.26)	(0.26)	(0.42)	(0.43)	(0.02)$^{+}$	(0.02)$^{+}$
Own English $_{t-1}$	0.59	0.55	0.63	0.49	1.22	1.03	-0.02	-0.03
	(0.27)**	(0.27)**	(0.22)$^{+}$	(0.22)**	(0.37)$^{+}$	(0.37)$^{+}$	(0.02)	(0.02)
Cohort-track F.E	--	Yes	--	Yes	--	Yes	--	Yes
Cohort F.E.	Yes	--	Yes	--	Yes	--	Yes	--
Observations	6618	6618	6618	6618	6618	6618	6618	6618
R-squared	0.01	0.01	0.02	0.03	0.02	0.02	0.05	0.06
SD. of Dep. Var.	9.50	9.50	11.48	11.48	16.30	16.30	.66	.66
Change due to Peer Effect	-0.05	-0.21	-0.02	-1.07	-0.04	-0.88	-0.03	-1.58

Note: The bottom panel presents standard deviations of change in the dependent variable due to a one standard deviation increase in the average score of peers. Robust standard errors clustered at the classroom level are reported in parentheses. The sample includes 138 classrooms. $^{+}$ significant at 1%; ** significant at 5%; * significant at 10%

Table 1.10: Differential Effects of Peers on Math Score

	(1)	(2)	(3)	(4)
	---------- Semester One ----------		---------- Semester Two ----------	
Bottom 25%:				
Ave. Peers' Math $_{t-1}$	0.23	0.23	0.56	0.66
	(0.27)	(0.27)	(0.34)	(0.35)*
Middle 50%:				
Ave. Peers' Math $_{t-1}$	0.19	0.36	0.60	0.73
	(0.26)	(0.24)	(0.33)*	(0.33)**
Top 25%:				
Ave. Peers' Math $_{t-1}$	0.13	0.12	0.64	0.53
	(0.25)	(0.24)	(0.33)*	(0.31)*
Own baseline scores x Type	No	Yes	No	Yes
Cohort-track F.E.	Yes	Yes	Yes	Yes
Observations	6618	6618	6618	6618
R-squared	0.30	0.40	0.25	0.31
Joint Tests of Peer Effects:				
Bottom 25% vs. Middle 50%	0.21	1.66	0.20	0.48
(p-value)	(0.65)	(0.20)	(0.65)	(0.49)
Top 25% vs. Middle 50%	1.25	7.60	0.37	5.13
(p-value)	(0.27)	(0.01)***	(0.55)	(0.03)**

Note: All specifications include a set of dummies indicating whether the student is in the top, middle, or bottom of the initial math test score distribution. Own baseline scores x type is a set of interaction terms between baseline scores (math, Chinese, English, and Malay) and dummies indicating whether the student is in the top, middle, or bottom of the initial math score distribution. The bottom panel reports the F stat for the test of whether peer effects differ for different types of students. Robust standard errors clustered at the classroom level are reported in parentheses. The sample includes 138 classrooms. *** significant at 1%; ** significant at 5%; * significant at 10%

Table 1.11: Effects of Classroom Heterogeneity on Math Score

	(1)	(2)	(3)	(4)
	---------- Semester One ----------		---------- Semester Two ----------	
All:				
Ave. Peers' Math $_{t-1}$	0.30		0.65	
	(0.25)		(0.33)**	
S.D. of Classroom Math $_{t-1}$	-0.21		0.08	
	(0.32)		(0.42)	
Bottom 25%:				
Ave. Peers' Math $_{t-1}$		0.25		0.65
		(0.28)		(0.35)*
S.D. of Classroom Math $_{t-1}$		-0.16		0.12
		(0.41)		(0.50)
Middle 50%:				
Ave. Peers' Math $_{t-1}$		0.39		0.72
		(0.25)		(0.33)**
S.D. of Classroom Math $_{t-1}$		-0.12		0.13
		(0.34)		(0.46)
Top 25%:				
Ave. Peers' Math $_{t-1}$		0.15		0.53
		(0.24)		(0.31)*
S.D. of Classroom Math $_{t-1}$		-0.37		-0.10
		(0.39)		(0.43)
Own Baseline Scores	Yes	Yes	Yes	Yes
Cohort-Track F.E.	Yes	Yes	Yes	Yes
Observations	6618	6618	6618	6618
R-squared	0.40	0.40	0.31	0.31
Joint Tests of S.D. of Baseline Classroom Math:				
Bottom 25% vs. Middle 50%		0.01		0.00
(p-value)		(0.91)		(0.98)
Top 25% vs. Middle 50%		1.03		0.73
(p-value)		(0.31)		(0.39)

Note: Own baseline scores in columns (1) and (3) include baseline test scores of math, Chinese, English, and Malay. Own baseline scores in columns (2) and (4) include a set of interaction terms between baseline scores (math, Chinese, English, and Malay) and dummies indicating whether the student is in the top, middle, or bottom of the initial math score distribution. Columns (2) and (4) also include a set of dummies indicating whether the student is in the top, middle, or bottom of the initial math test score distribution. The bottom panel reports the F stat for the test of whether the standard deviations of classroom baseline math scores differ for different types of students. Robust standard errors clustered at the classroom level are reported in parentheses. The sample includes 138 classrooms. *** significant at 1%; ** significant at 5%; * significant at 10%

Table 1.12: Differential Peer Effects on Math Score by Ability Groups

	(1) High-achieving Track	(2) Low-achieving Track	(3) High-achieving Track – Low-achieving Track
Bottom 25%:			
Ave. Peers' Math $_{t-1}$	0.44	0.27	0.17
	(0.77)	(0.44)	(0.88)
Middle 50%:			
Ave. Peers' Math $_{t-1}$	0.31	0.84	-0.53
	(0.49)	(0.43)*	(0.64)
Top 25%			
Ave. Peers' Math $_{t-1}$	0.47	1.66	-1.19
	(0.35)	(0.44)***	(0.56)**
Joint Tests of S.D. of Baseline Classroom Math:			
Bottom 25% vs. Middle 50%	1.27	10.92	
(p-value)	(0.26)	(0.001)***	
Top 25% vs. Middle 50%	2.33	7.14	
(p-value)	(0.13)	(0.01)***	

Note: The sample includes 6618 junior 1 students enrolled in 138 classrooms. Columns (1) and (2) are estimated jointly in one regression specification that also includes a set of dummies indicating whether the student is in the top, middle, or bottom of the initial math score distribution, a set of own baseline test scores interacted with ability group specific indicators of whether the student is in the top, middle, or bottom of the initial math score distribution, and a set of cohort-track specific fixed effects. Column (3) reports the differences in the estimated effects of peers for students at different points of the initial math score distribution but assigned into different ability groups. The regression R-squared is 0.32. *** significant at 1%; ** significant at 5%; * significant at 10%

CHAPTER 1 – APPENDIX 1.1 –

MECHANICAL NEGATIVE CORRELATION

I show mathematically how the mechanical negative correlation exists in a typical study of peer effects, as well as its relationship with class size (*n*) and the number of classrooms (*C*) in this Appendix.

Consider the model: $y_{ic} = \alpha_0 + \alpha_1 \bar{y}_{-i,c} + \varepsilon_{ic}$, where *C* classrooms of size *n* are formed. The OLS estimator is:

$$\hat{\alpha}_1 = \frac{\sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y})(\bar{y}_{-i,c} - \bar{y})}{\sum_{c=1}^{C}\sum_{i=1}^{n}(\bar{y}_{-i,c} - \bar{y})^2} \qquad \text{(A1.1.1)}$$

Note that $\bar{y}_{-i,c} = \dfrac{n\bar{y}_c - y_{ic}}{n-1} = \dfrac{n}{n-1}\bar{y}_c - \dfrac{1}{n-1}y_{ic}$, where $\bar{y}_c$ is the mean outcome of classroom *c* that individual *i* is in.

We can express equation (A1.1.1) as:

$$\hat{\alpha}_1 = (n-1)\frac{\sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y})[n\bar{y}_c - (n-1)\bar{y} - y_{ic}]}{\sum_{c=1}^{C}\sum_{i=1}^{n}[n\bar{y}_c - (n-1)\bar{y} - y_{ic}]^2}$$

$$= -(n-1)\frac{\sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y})[y_{ic} - n\bar{y}_c + (n-1)\bar{y}]}{\sum_{c=1}^{C}\sum_{i=1}^{n}[y_{ic} - n\bar{y}_c + (n-1)\bar{y}]^2}$$

$$= -(n-1)\frac{\sum_{c=1}^{C}\sum_{i=1}^{n}\left[y_{ic} - n\bar{y}_c + (n-1)\bar{y} + n(\bar{y}_c - \bar{y})\right]\left[y_{ic} - n\bar{y}_c + (n-1)\bar{y}\right]}{\sum_{c=1}^{C}\sum_{i=1}^{n}\left[y_{ic} - n\bar{y}_c + (n-1)\bar{y}\right]^2}$$

$$= -(n-1) - (n-1)\frac{\sum_{c=1}^{C}\sum_{i=1}^{n} n(\overset{\varpi}{y}_c - \bar{y})\left[y_{ic} - n\bar{y}_c + (n-1)\bar{y}\right]}{\sum_{c=1}^{C}\sum_{i=1}^{n}\left[y_{ic} - n\bar{y}_c + (n-1)\bar{y}\right]^2}$$

$$= -(n-1) - (n-1)\frac{\sum_{c=1}^{C} n(\overset{\varpi}{y}_c - \bar{y})\left[(n^2 - n)\bar{y}_c + n(n-1)\bar{y}\right]}{\sum_{c=1}^{C}\sum_{i=1}^{n}\left[y_{ic} - n\bar{y}_c + (n-1)\bar{y}\right]^2}$$

$$= -(n-1) + (n-1)\frac{n^2(n-1)\sum_{c=1}^{C}(\overset{\varpi}{y}_c - \bar{y})^2}{\sum_{c=1}^{C}\sum_{i=1}^{n}\left[y_{ic} - n\bar{y}_c + (n-1)\bar{y}\right]^2} \qquad \text{(A1.1.2)}$$

The numerator of (A1.1.2) can be expressed as:

$$\sum_{c=1}^{C}\sum_{i=1}^{n}\left[y_{ic} - n\bar{y}_c + (n-1)\bar{y}\right]^2 = \sum_{c=1}^{C}\sum_{i=1}^{n}\left[y_{ic} - \bar{y} - n(\bar{y}_c - \bar{y})\right]^2$$

$$= \sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y})^2 - 2n\sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y})(\bar{y}_c - \bar{y}) + n^3\sum_{c=1}^{C}(\bar{y}_c - \bar{y})^2$$

$$= \sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y})^2 - 2n^2\sum_{c=1}^{C}(y_{ic} - \bar{y})^2 + n^3\sum_{c=1}^{C}(\bar{y}_c - \bar{y})^2$$

$$= \sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y})^2 + (n^3 - 2n^2)\sum_{c=1}^{C}(\bar{y}_c - \bar{y})^2$$

Thus,

$$\hat{\alpha}_1 \qquad = -(n-1) + (n-1)\frac{n^2(n-1)\sum_{c=1}^{C}(\bar{y}_c - \bar{y})^2}{\sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y})^2 + (n^3 - 2n^2)\sum_{c=1}^{C}(\bar{y}_c - \bar{y})^2}$$

$$= -(n-1) + (n-1)\frac{Cn(n-1)\frac{n}{C}\sum_{c=1}^{C}(\bar{y}_c - \bar{y})^2}{nC\left[\frac{1}{nC}\sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y})^2\right] + n(n-2)C\left[\frac{n}{C}\sum_{c=1}^{C}(\bar{y}_c - \bar{y})^2\right]}$$

$$= -(n-1) + (n-1)\frac{n(n-1)\frac{n}{C}\sum_{c=1}^{C}(\bar{y}_c - \bar{y})^2}{n\left[\frac{1}{nC}\sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y})^2\right] + n(n-2)\left[\frac{n}{C}\sum_{c=1}^{C}(\bar{y}_c - \bar{y})^2\right]}$$

$$= -(n-1) + (n-1)\frac{n(n-1)s_c^2}{ns^2 + n(n-2)s_c^2}$$

$$\hat{\alpha}_1 = (n-1)\left[\frac{n(n-1)s_c^2}{ns^2 + n(n-2)s_c^2} - 1\right]$$

$$= (n-1)\left[\frac{(n-1)s_c^2 - s^2 - (n-2)s_c^2}{s^2 + (n-2)s_c^2}\right]$$

$$= (n-1)\frac{s_c^2 - s^2}{s^2 + (n-2)s_c^2}$$

$$= (s_c^2 - s^2)\frac{n-1}{s^2 + (n-2)s_c^2} \qquad \text{(A1.1.3)}$$

where,

$$s_c^2 = \frac{n}{C}\sum_{c=1}^{C}(\bar{y}_c - \bar{y})^2 \qquad \text{(The between class variance of mean)}$$

and,

$$s^2 = \frac{1}{nC}\sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y})^2$$

$$= \frac{1}{nC}\sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y}_c + \bar{y}_c - \bar{y})^2$$

$$= \frac{1}{nC}\sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y}_c)^2 + \frac{1}{nC}\sum_{c=1}^{C}\sum_{i=1}^{n}(\bar{y}_c - \bar{y})^2$$

$$= \frac{1}{nC}\sum_{c=1}^{C}\sum_{i=1}^{n}(y_{ic} - \bar{y}_c)^2 + \frac{1}{C}\sum_{c=1}^{C}(\bar{y}_c - \bar{y})^2$$

$$= s_u^2 + \frac{1}{n}s_c^2 \qquad \text{(The overall variance)}$$

<u>Case 1</u>: $n \to \infty$ **holding *C* fixed**

When we hold the number of classrooms fixed and increase the number of students assigned into each classroom, we obtain:

$$\hat{\alpha}_1 \to \operatorname*{plim}_{n\to\infty}\left(s_c^2 - s^2\right) \neq 0$$

If we assume that $y_{ic} \sim iid\ \mathrm{N}\left(\mu, \sigma_y^2\right)$, then

$$s_c^2 \xrightarrow{d} \frac{1}{C}\sum_{c=1}^{C}\left(\sqrt{n}\bar{y}_c - \sqrt{n}\bar{y}\right)^2 \overset{d}{=} \frac{1}{C}\sum_{c=1}^{C}\left(\xi_c - \bar{\xi}\right)^2$$

where the distribution of ξ_c is $iid\ \mathrm{N}\left(\mu, \sigma_y^2\right)$. In addition,

$$s^2 \to \operatorname*{plim}_{n\to\infty}\left(s_u^2\right)$$

$$= \frac{1}{C}\sum_{c=1}^{C}\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(y_{ic} - \bar{y}_c)^2$$

$$= \frac{1}{C}\sum_{c=1}^{C}\sigma_y^2$$

$$= \sigma_y^2$$

Thus,

$$\hat{\alpha}_1 \xrightarrow{d} \frac{1}{C}\sum_{c=1}^{C}\left(\xi_c - \bar{\xi}\right)^2 - \sigma_y^2 \equiv \Omega$$

That is, $\hat{\alpha}_1$ converges in distribution to Ω, and:

$$E(\Omega) = \left(\frac{C-1}{C} - 1\right)\sigma_y^2 = -\frac{\sigma_y^2}{C} < 0$$

<u>Case 2</u>: $C \to \infty$, holding *n* fixed.

Under the assumption that $y_{ic} \sim iid\ \mathrm{N}\left(\mu, \sigma_{\mathrm{y}}^2\right)$, $s_c^2 \xrightarrow{p} \sigma_y^2$ and $s^2 \xrightarrow{p} \sigma_y^2$, and hence,

$$\hat{\alpha}_1 \xrightarrow{p} 0$$

The inverse relationship between y_i and *i*'s classmates diminishes as *C* gets larger. As *C* gets larger, the between class variance of mean approaches the overall variance. However, holding *C* fixed, as *n* gets larger, $\hat{\alpha}_1$ converges in distribution to a distribution with a negative mean. In other words, the driving factor for the negative correlation is not the size of sample population (*N*) per se, but the number of classrooms (*C*) formed.

The intuition for the mechanical negative correlation and how it relates to the number of classrooms (*C*) and class size (*n*) can be best understood by looking at equation A.1.1.3. The OLS estimator $\hat{\alpha}_1$ tends to be negative when the between class variance of classroom mean (s_c^2) is smaller than the overall variance (s^2). For any given number of classrooms, the between class variance of classroom is likely to be small if class size (*n*) is large, because the classroom mean ($\bar{y}_c$) is very close to the overall sample mean ($\bar{y}$). As the number of classrooms increases, the between variance of the between classroom means converges to the overall variance, the OLS estimator $\hat{\alpha}_1$ converges to zero.

CHAPTER 1 – APPENDIX 1.2 –

PEER EFFECTS ON ACHIEVEMENT IN CHINESE, ENGLISH, AND MALAY

Table A1.2.1 reports estimates of the effects of peers on Chinese, English, and Malay. Because these test scores include grades from assignments, such as essay writings, designed by the teacher responsible for the class, they potentially suffer from severe measurement errors. The estimates of peer effects are statistically not different from zero for all subjects and in both semesters. The results indicate that either peer effects are absent in these subjects or test scores in these subjects are noisy and not comparable across classrooms. The results echo findings of peer effects on a student's own GPA in foreign language classes by Carrell et al. (2009) at the U.S. Air Force Academy.

Table A1.2.1 OLS Estimates of Peer Effects on Chinese, English, and Malay Achievement

	------------- Semester One -------------			------------- Semester Two -------------		
	Chinese	English	Malay	Chinese	English	Malay
Ave. Peers' Chinese $_{t-1}$	-0.13 (0.36)			0.03 (0.34)		
Ave. Peers' English $_{t-1}$		0.05 (0.37)			-0.15 (0.37)	
Ave. Peers' Malay $_{t-1}$			0.001 (0.31)			-0.05 (0.36)
Own Chinese $_{t-1}$	0.66 (0.02)***	0.15 (0.01)***	0.21 (0.02)***	0.56 (0.02)***	0.15 (0.01)***	0.18 (0.02)***
Own English $_{t-1}$	-0.02 (0.02)	0.89 (0.02)***	0.20 (0.02)***	-0.05 (0.02)**	0.81 (0.02)***	0.14 (0.02)***
Own Malay $_{t-1}$	0.30 (0.02)***	0.32 (0.01)***	0.82 (0.02)***	0.29 (0.02)***	0.30 (0.02)***	0.76 (0.02)***
Own Math $_{t-1}$	0.11 (0.02)***	0.09 (0.01)***	0.09 (0.01)***	0.11 (0.02)***	0.09 (0.01)***	0.07 (0.01)***
Cohort -track F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6618	6618	6618	6618	6618	6618
R-squared	0.41	0.61	0.52	0.34	0.53	0.45

Note: Robust standard errors clustered at the classroom level are reported in parentheses. The sample includes 138 classrooms. *** significant at 1%; ** significant at 5%; * significant at 10%

CHAPTER 1 – APPENDIX 1.3 –

BIAS IN A NON-LINEAR SPECIFICATION AND A CORRECTION METHOD

I present Monte Carlo simulation results showing the bias suffered from a commonly used non-linear model of peer effects in this appendix. In addition, I also show in Monte Carlo simulations that using baseline test score as a control variable can effectively control for the bias.

Consider using the following regression equation to examine whether the effects of peers vary across different types of students:

$$y_{ict} = \beta_1 I_{ict-1}^{Bottom_1} + \beta_2 I_{ict-1}^{Middle} + \beta_3 I_{ict-1}^{Top}$$

$$+ \gamma_1 I_{ict-1}^{Bottom} \bar{y}_{-ict-1} + \gamma_2 I_{ict-1}^{Middle} \bar{y}_{-ict-1} + \gamma_3 I_{ict-1}^{Top} \bar{y}_{-ict-1} + \varepsilon_{ict} \qquad \text{(A1.3.1)}$$

where y_{ict} is the current achievement of student *i*, $\bar{y}_{ict-1}$ is the average baseline achievement of student *i*'s peers, $I_{ict-1}^{Bottom_1}$ is an indicator for whether y_{ict-1} is in the bottom 25 percentile of the initial test score distribution, I_{ict-1}^{Middle} is an indicator for whether y_{ict-1} is in the middle 50 percentile of the initial test score distribution, and I_{ict-1}^{Top} is an indicator for whether y_{ict-1} is in the top 25 percentile of the initial test score distribution. Assume that students are randomly assigned into classrooms, peer effects are absent, and student *i*'s achievement is solely determined by her past achievement:

$$y_{ict} = 0.8 y_{ict-1} + 0.6 u_{ict} \qquad \text{(A1.3.2)}$$

where $y_{ict-1} \sim iid.\, N(0,1)$ and $u_{ict} \sim iid.\, N(0,1)$. Simulations show that OLS estimates based on regression equation (A1.3.2) will produce non-zero $\hat{\gamma}_1$, $\hat{\gamma}_2$, and $\hat{\gamma}_3$ on average.

To illustrate the extent of the bias, consider an example where 1500 students are randomly assigned into 30 classrooms, each with 50 students, and achievement is determined by equation (A1.3.2). Table A1.3.1 column (1) reports the average of $\hat{\gamma}_1$, $\hat{\gamma}_2$, and $\hat{\gamma}_3$, which are statistically different from zero, based on 10,000 Monte Carlo regressions of (A1.3.1). Column (2) shows that even when the sample size is 5000 students and 100 classrooms are formed, the mechanical correlations remain significant.

I propose including a set of interaction terms of baseline test score with the indicator of where a student sat on the initial test score as control variables in equation (A1.3.1):

$$y_{ict} = \beta_1 I_{ict-1}^{Bottom_1} + \beta_2 I_{ict-1}^{Middle} + \beta_3 I_{ict-1}^{Top} + \gamma_1 I_{ict-1}^{Bottom} \overline{y}_{-ict-1} + \gamma_2 I_{ict-1}^{Middle} \overline{y}_{-ict-1} + \gamma_3 I_{ict-1}^{Top} \overline{y}_{-ict-1}$$
$$+ \pi_1 I_{ict-1}^{Bottom} y_{ict-1} + \pi_2 I_{ict-1}^{Middle} y_{ict-1} + \pi_3 I_{ict-1}^{Top} y_{ict-1} + \varepsilon_{ict} \qquad \text{(A1.3.3)}$$

Column (3) and column (4) in Table A1.3.1 show that adding these control variables is effective in absorbing the mechanical correlations. Comparing column (3) to column (1) and column (4) to column (2), it is clear that adding the control variables can effectively reduce the extent of the bias.

Table A1.3.1: Bias with and without Individual Baseline Control Variables

	(1)	*(2)*	*(3)*	*(4)*
Average of γ1 estimates	-0.0048	-0.0031	0.0022	0.0011
	(0.0033)	(0.0017)*	(0.0028)	(0.0015)
Average of γ2 estimates	-0.0103	-0.0025	0.0019	0.0005
	(0.0018)***	(0.0010)***	(0.0016)	(0.0008)
Average of γ3 estimates	-0.0135	-0.0031	-0.0031	-0.0002
	(0.0033)***	(0.0017)*	(0.0028)	(0.0015)
Sample size (*N*)	1500	5000	1500	5000
Class size (*n*)	50	50	50	50
Number of classrooms (*C*)	30	100	30	100
Control variables	No	No	Yes	Yes

Note: Column (1) and column (2) are based on equation (A1.3.1). Column (3) and column (4) are based on equation (A1.3.3). Coefficients are the average of estimated effects of peers when peer effects are assumed zero using 10000 repetitions of Monte Carlo simulations. Standard errors reported in parentheses. *** 1% significant; ** 5% significant; * 10% significant.

CHAPTER 1 – APPENDIX 1.4 –

SUMMARY STATISTICS BY TYPES OF STUDENTS AND ABILITY GROUPS

Table A1.4.1: Descriptive Statistics by Types of Students in All Ability Groups

Variable		*Obs.*	*Mean*	*Std. Dev.*	*Min*	*Max*
Bottom 25%						
Mathematics	- Baseline	1740	-0.417	0.389	-2.501	0.154
	- Semester 1	1740	-0.743	0.934	-4.687	1.738
	- Semester 2	1740	-0.619	0.931	-3.643	1.744
Ave. Peers' Math	- Baseline	1740	0.318	0.246	-0.105	1.451
Ave. SD. Classroom Math	- Baseline	1740	0.684	0.087	0.463	0.996
Middle 50%						
Mathematics	- Baseline	3119	0.530	0.295	-0.069	1.106
	- Semester 1	3119	0.011	0.844	-3.221	2.337
	- Semester 2	3119	-0.028	0.902	-3.275	2.100
Ave. Peers' Math	- Baseline	3119	0.500	0.414	-0.125	1.485
Ave. SD. Classroom Math	- Baseline	3119	0.675	0.098	0.463	0.996
Top 25%						
Mathematics	- Baseline	1759	1.566	0.431	1.032	4.176
	- Semester 1	1759	0.722	0.750	-2.372	2.410
	- Semester 2	1759	0.669	0.791	-2.109	2.249
Ave. Peers' Math	- Baseline	1759	0.892	0.427	-0.145	1.461
Ave. SD. Classroom Math	- Baseline	1759	0.674	0.128	0.463	0.996

Note: The sample includes 6,618 grade 7 students enrolled in 138 classrooms in academic years 2002 to 2008. Students who dropped out are excluded. 14 students without baseline test scores available are also excluded from the sample. Baseline test scores are standardized across all test takers of each academic year. Test scores in semester 1 and semester 2 are standardized across all junior 1 students of each academic year. The type of the student is defined by whether the student scored in the bottom 25%, middle 50%, or top 25% of the baseline math score distribution.

Table A1.4.2: Descriptive Statistics by Types of Students in High Achieving Group

Variable		*Obs.*	*Mean*	*Std. Dev.*	*Min*	*Max*
Bottom 25%						
Mathematics	- Baseline	95	-0.220	0.255	-1.023	0.154
	- Semester 1	95	-0.503	0.794	-2.812	1.330
	- Semester 2	95	-0.182	0.727	-1.979	1.513
Ave. Peers' Math	- Baseline	95	1.169	0.087	1.019	1.451
Ave. SD. Classroom Math	- Baseline	95	0.737	0.139	0.463	0.996
Middle 50%						
Mathematics	- Baseline	791	0.637	0.288	-0.069	1.106
	- Semester 1	791	0.044	0.808	-2.603	1.767
	- Semester 2	791	0.226	0.756	-2.541	1.915
Ave. Peers' Math	- Baseline	791	1.174	0.100	0.998	1.485
Ave. SD. Classroom Math	- Baseline	791	0.680	0.139	0.463	0.996
Top 25%						
Mathematics	- Baseline	1199	1.637	0.460	1.032	4.176
	- Semester 1	1199	0.731	0.755	-1.907	2.410
	- Semester 2	1199	0.788	0.730	-1.999	2.249
Ave. Peers' Math	- Baseline	1199	1.173	0.103	0.962	1.461
Ave. SD. Classroom Math	- Baseline	1199	0.667	0.144	0.463	0.996

Note: The sample includes 2,085 grade 7 students enrolled in 42 high-achieving group classrooms in academic years 2002 to 2008. Students who dropped out and without baseline test scores available are excluded from the sample. Baseline test scores are standardized across all test takers of each academic year. Test scores in semester 1 and semester 2 are standardized across all junior 1 students of each academic year. The type of the student is defined by whether the student scored in the bottom 25%, middle 50%, or top 25% of the baseline math score distribution.

Table A1.4.3: Descriptive Statistics by Types of Students in Low Achieving Group

Variable		*Obs.*	*Mean*	*Std. Dev.*	*Min*	*Max*
Bottom 25%						
Mathematics	- Baseline	1645	-0.429	0.392	-2.501	0.154
	- Semester 1	1645	-0.757	0.939	-4.687	1.738
	- Semester 2	1645	-0.644	0.935	-3.643	1.744
Ave. Peers' Math	- Baseline	1645	0.269	0.140	-0.105	0.548
Ave. SD. Classroom Math	- Baseline	1645	0.680	0.082	0.520	0.871
Middle 50%						
Mathematics	- Baseline	2328	0.494	0.288	-0.069	1.106
	- Semester 1	2328	0.000	0.856	-3.221	2.337
	- Semester 2	2328	-0.114	0.931	-3.275	2.100
Ave. Peers' Math	- Baseline	2328	0.271	0.137	-0.125	0.518
Ave. SD. Classroom Math	- Baseline	2328	0.673	0.079	0.520	0.871
Top 25%						
Mathematics	- Baseline	560	1.414	0.310	1.032	2.858
	- Semester 1	560	0.702	0.739	-2.372	2.268
	- Semester 2	560	0.415	0.854	-2.109	2.193
Ave. Peers' Math	- Baseline	560	0.290	0.133	-0.145	0.501
Ave. SD. Classroom Math	- Baseline	560	0.688	0.083	0.520	0.871

Note: The sample includes 4,533 grade 7 students enrolled in 96 low-achieving group classrooms in academic years 2002 to 2008. Students who dropped out and without baseline test scores available are excluded from the sample. Baseline test scores are standardized across all test takers of each academic year. Test scores in semester 1 and semester 2 are standardized across all junior 1 students of each academic year. The type of the student is defined by whether the student scored in the bottom 25%, middle 50%, or top 25% of the baseline math score distribution.

CHAPTER 1 - REFERENCES

Adams, A. Troy (2000) "The Status of School Discipline and Violence" *Annals of the American Academy of Political and Social Science*, vol. 567, pp.140-156

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, and Tristan Zajonc (2009) "Do Value-Added Estimates Add Value? Accounting for Learning Dynamics," World Bank Policy Research Working Paper No. 5066.

Angrist, Joshua D. and Alan B. Krueger (1991) "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, vol.106 (4), pp.979-1014

Angrist, Joshua D. and Kevin Lang (2004) "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program," *American Economic Review*, vol. 94(5), pp.1613-1634

Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster, and Josh Kinsler (2005) "Estimating Spillovers in the Classroom with Panel Data," Unpublished Manuscript

Betts, Julian and Andrew Zau (2004) "Peer Groups and Academic Achievement: Panel Evidence from Administrative Data", Unpublished manuscript, UCSD

Bishop, John (1989) "Is the Test Score Decline Responsible for the Productivity Growth Decline?" *American Economic Review*, vol.79, pp.178-197

Bound, John and George Johnson (1992) "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations," *American Economic Review*, vol.82, pp.371-392

Burke, Mary A. and Tim R. Sass (2008) "Classroom Peer Effects and Student Achievement," Federal Reserve Bank of Boston Working Paper No. 08-5

Carrell, Scott E., Richard L. Fullerton, and James E. West (2009) "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, vol.27 (3), pp.439-464

Cooley, Jane (2008) "Desegregation and the Achievement Gap: Do Diverse Peers Help?" Unpublished manuscript

Cooley, Jane (2009) "Can Achievement Peer Effect Estimates Inform Policy? A View from Inside the Black Box," WCER Working Paper No. 2009-8, September

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli (forthcoming) “Be as Careful of the Books You Read as of the Company You Keep: Evidence on Peer Effects in Educational Choices,” *American Economic Journal – Applied Economics*

Duflo, Esther, Pascaline Dupas, and Michael Kremer (forthcoming) “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya,” *American Economic Review*.

Foster, Gigi (2006) “It’s Not Your Peers, and It’s Not Your Friends: Some Progress toward Understanding the Educational Peer Effect Mechanism,” *Journal of Public Economics*, vol.90 (8-9), pp.1455-1475

Gamoran, Adam and Robert D. Mare (1989) “Secondary School Tracking and Educational Inequality: Compensation, Reinforcement, or Neutrality?” *American Journal of Sociology*, vol.94 (5), pp.1146-1193

Guryan, Jonathan, Kory Kroft, and Matt Notowidigdo (forthcoming) “Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments,” *American Economic Journal – Applied Economics*

Hanushek, Eric, John F. Kain, John Markman, and Steven G. Rivkin (2003) “Does Peer Ability Affect Student Achievement?” *Journal of Applied Econometrics*, vol.18 (5), pp.527-544

Hoxby, Caroline and Gretchen Weingarth (2007) “Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects,” Unpublished manuscript, Harvard University

Kang, Changhui (2007) “Classroom Peer Effects and Academic Achievement: Quasi-randomization Evidence from South Korea,” *Journal of Urban Economics*, vol.61, pp.458-495

Lyle, David (2007) “Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point,” *Review of Economics and Statistics*, vol.89 (2), pp.289-299

Manski, Charles (1993) “Identification of Endogenous Social Effects: The Reflection Problem”, *Review of Economic Studies*, vol.60 (3), pp.531-542

Moffitt, Robert (2001) “Policy Interventions, Low-Level Equilibria, and Social Interactions,” in *Social Dynamics*, S. Durlauf and H. P. Young eds., Cambridge: MIT Press

Sacerdote, Bruce I. (2001) “Peer Effects with Random Assignment: Results for Dartmouth Roommates,” *Quarterly Journal of Economics*, vol.116 (2), pp.681-704

Vigdor, Jacob and Thomas Nechyba (2007) "Peer Effects in North Carolina Public Schools," in *Schools and the Equal Opportunity Problem*, P. Peterson and L. Woessmann eds., MIT Press

Zimmerman, David J. (2003) "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics*, vol. 85 (1), pp.9-23

CHAPTER 2

RESTRICTING SECULAR EDUCATION – A RELIGIOUS SACRIFICE? EVIDENCE FROM THE AMISH

Abstract

Given the positive returns to education, Amish prohibition of high school education appears puzzling from a rational choice perspective. I extend Iannaccone's (1992) religious club model to explain why the Amish would collectively object to high school education and refuse to comply with compulsory schooling laws. I exploit the surprising 1972 U.S. Supreme Court's decision in *Wisconsin vs. Yoder*, which exempts Amish children from compulsory high school education, as a policy shock to test the predictions of the model. I find that successful restriction on high school education helped the Amish sect exclude individuals who have high labor productivity and would lower the quality of the sect from joining. The evidence supports the idea that the Amish use the restriction on secular education as a religious sacrifice to screen out uncommitted members.

2.1 Introduction

The Amish stirred heated debates in the United States in the mid-twentieth century by stubbornly refusing to comply with compulsory school attendance laws. Their insistence on the eighth grade as the final year of formal schooling, on the basis of their religious belief, frequently led to fines, prosecutions, and even imprisonment by local authorities. After numerous court cases and decades of struggles against the state, the Amish were eventually exempted from compulsory high school education on the grounds of religious liberty by the 1972 U.S. Supreme Court's decision in "State of Wisconsin v. Jonas Yoder et al" (*Wisconsin v. Yoder*). Given the positive returns to education documented in the literature, it is puzzling why the Amish would enforce a ceiling on years of schooling, which seemingly make their members better off.

This paper extends Iannaccone's (1992) religious club model to explain why the Amish would collectively restrict high school education. The model posits that religious activities among Amish members generate a positive externality, which an Amish individual does not take into consideration in maximizing utility. First, by restricting high school education, the Amish can increase the amount of religious activities chosen within the sect and achieve the socially optimal level. Second, in the presence of unobserved heterogeneity, the sect can request that potential members sacrifice their high school education as a signal of their types. This prevents "free-riders" who would otherwise lower the positive externality in the club from joining. The sacrifice is set so that individuals with high labor market productivity and low religious commitment choose not to join the sect. According to the religious club model, the Amish clashed with the government because compulsory schooling laws imposed a level of schooling exceeding the social optimum for the sect.

The religious club model yields a set of testable assumptions and predictions. First, the model predicts that the Supreme Court's ruling should lead to lower educational attainment among the Amish since their restriction on high school education was no longer prevented by the government. Second, the successful prohibition on high school education should reduce the wage rate of affected Amish individuals. Third, the exemption from compulsory schooling laws permits the Amish to request potential entrants to sacrifice high school education. If the restriction on high school education acts as a religious sacrifice for the Amish to screen out uncommitted members, the Supreme Court's decision should lead to (1) Amish individuals with high labor market productivity leaving the sect; and (2) females with high shadow cost of child rearing leaving the sect.

Because the Amish speak Pennsylvania Dutch, I can use the U.S census data to test the model predictions. The census data reveal that Amish individuals on average have lower educational attainment, lower earnings, and larger family size than former Amish and non-Amish individuals. The Supreme Court's decision is estimated to increase Amish high school dropout rates by 15 to 25 percent for males and 6 to 13 percent for females. The Supreme Court's exemption decreased average years of completed schooling by 8 to 10 months for Amish males and 4 to 12 months for Amish females. The exemption lowered hourly earnings by 23 to 34 percent and increased births by 0.16 to 0.34. When I implement a difference-in-difference estimator using non-Amish individuals as a control group to remove cohort differences not affected by the exemption, the estimates are similar for earnings, but a little larger for fertility.

If we attribute the fall in log hourly earnings or the increase in fertility to the decrease in educational attainment driven by the exemption, the implied return to education is estimated to be 23 to 32 percent higher in wages and the implied effect of an additional year of schooling on fertility is estimated to be -0.91 births. These estimates are at least 50 percent greater in magnitude than the estimated causal effects of education on earnings and fertility reported by past studies. They are also more than 3 to 4 times the OLS estimates based on a cross-sectional sample of Amish individuals. The large implied effects of education indicate that the Amish use the restriction on high school education to screen out uncommitted members. That is, exempted Amish youths who joined the sect have much lower labor productivity than non-exempted Amish youths who joined the sect, amplifying the effects of education. The findings support the hypothesis that increased religious sacrifice leads to productive individuals selecting out of religious groups which request a high level of religious participation.

Iannaccone's (1992) religious club model was previously applied and tested in a number of settings, including Israeli Ultra-Orthodox Jews (Berman 2000), radical Muslim groups (Berman and Stepanyan 2005), low European Catholic fertility (Berman, Iannaccone, and Ragusa 2007), and religious terrorists (Berman and Laitin 2008; Berman and Iannaccone 2006). In particular, Berman and Laitin's (2008) explanation for the effectiveness of radical religious groups in conducting acts of terrorism lies crucially on their ability to use religious sacrifice to screen committed operatives. However, previous studies did not empirically test the relationship between increased sacrifice and type selection. This paper provides empirical evidence on how religious sacrifice facilitates effective screening and shows how public policies can influence outcomes of religious sects. The finding that individuals with higher labor productivity are more likely to leave the Amish sect also echoes recent research by Abramitzky (2008; 2009) that shows productive individuals have a higher tendency to exit Israeli kibbutzim, which practice income sharing. Given the increasing tension between religious groups and states around the globe, understanding how public policies can affect a sect's capability to screen potential members may shed lights on options available to government in influencing participation in religious sects, as well as contributes to the debates pertaining to the freedom to exercise religious beliefs.

2.2 Background: Amish Society and Its Educational Conflicts with the State

Founded by Jacob Ammann in Alsace, France in the 1690s, the Amish are a religious sect that split from the Swiss Anabaptist Mennonites when Ammann advocated the shunning of excommunicated members in daily life.[27] Sociopolitical instability and religious persecution in Europe prompted the Amish to migrate to America and settle in Lancaster, Pennsylvania, in the eighteenth and nineteenth centuries (Hostetler 1993, pp. 31-34). In 2000, there were approximately 200,000 Amish residing in the United States, roughly 70 percent of which are in Pennsylvania, Indiana, and Ohio (Kraybill and Hostetter 2001, pp. 75-77).[28] Eighty-five percent of Amish are Old Order Amish, who are most conservative, while other Amish groups, such as New Order Amish, Beachy Amish, and Amish Mennonite, are more progressive (Kraybill and Hostetter 2001, pp. 66-67).

---

[27] Other Anabaptist groups that are similar to the Amish include conservative Mennonites and the Hutterites.

[28] There are also Amish settlements in the Canadian province of Ontario and Latin America.

The Amish and other Anabaptists strive to maintain a simple Christian life that discourages material success, seeking to separate themselves from the world and worldly influences. The emphasis on separation from the world governs many of Amish customs, including dress codes, the use of technology, attitudes towards education, and the choice of school. The conduct of an Amish person is regulated by the *Ordnung* of each congregation, which is a set of standards or expectations for behavior (Kraybill 2001, p.112). Unlike other Christian denominations, the Amish and other Anabaptists practice adult baptism. Starting from age 16, unbaptized Amish participate in *Rumspringa* and may leave their communities for the outside world. After experiencing secular life for a few years, adolescents who decide to be baptized into the Church become full-pledged members.[29] Each Amish community is organized around a church district, which typically consists of 30 families with 60 baptized adults and 75 unbaptized youths. The small size of congregations facilitates both mutual aid provision and social insurance; members help each other with barn raisings, harvesting, quilting, births, weddings, and funerals and assistance in the events of drought, disease, death, injury, bankruptcy, and medical emergency (Kraybill and Bowman 2001, p.113).

The Amish believe that eight years of formal schooling is adequate to equip their children with basic skills necessary to be good famers and citizens and to interact with non-Amish people in general. The Amish object to high school education because it exposes their children to worldly influences in conflict with their beliefs. Typical high school curricula and activities not only unnecessary for successful careers in Amish life

[29] The Amish and other Anabaptists practice adult baptism. The typical age of baptism ranges from sixteen to the early twenties for the Old Order Amish, and roughly sixty percent join the church before they reach twenty-one (Kraybill 2001, p.117).

but also stir aspirations and raise occupational hopes that turn Amish youths away from farm and family (Kraybill 2001, pp.175-176).

The Amish preferred one-room schoolhouses, common in rural America throughout the middle of the twentieth century, because the small scale rural allows the community convenient access and control over multiple facets of their children's education. Parents can unexpectedly visit the classroom, the school board can hire Amish teachers (or otherwise sympathetic) teachers and adjust class schedules when special occasions arise (Meyers 1993). The small local public schools gave the Amish limited contact with non-Amish people and taught the basic skills needed (Huntington 1994). As state authorities consolidated rural public schools and enforced high school attendance, especially during the post-WWII period, the Amish resisted and formed parochial schools to avoid compulsory high school attendance and maintain their traditional education standards.

The first recorded conflict between the Amish and school officials occurred in 1914 in Geauga County, Ohio when Amish fathers were fined for not sending their children under sixteen to public high school (Meyers 1993). Over the next sixty years, Amish people continued to face opposition over schooling related issues from state and local school authorities. Their refusal to comply with compulsory attendance laws frequently led to fines and imprisonments of Amish fathers. Similarly, Amish parochial schools that hired noncertified teachers, who typically had only eight years of education, also faced repeated shut-down attempts from state agents.

After numerous conflicts between the Amish and school authorities, a compromise was finally reached between the Amish and the state of Pennsylvania in

1956. The concession allowed Amish children who were at least fourteen and passed the eighth grade to attend a special vocational school until they were at least fifteen years old. Once a week, the children would meet for a minimum of three hours with an Amish teacher to study English, mathematics, health, and social studies and to report on their week's work at home (Meyers 1993). Classroom learning was supplemented by home projects in agriculture and homemaking (Hostetler and Huntington 1971, p.71). Attendance records were kept and forwarded to the state. In 1958, a similar settlement was reached in Ohio.[30] Nevertheless, Ohio state authorities frequently attempted to shut down "substandard" Amish vocational schools and forced Amish children to attend public high schools throughout the 1960s (Meyers 1993). In 1967, a comparable vocational training program was also established in Indiana for Amish children younger than sixteen (Hostetler and Huntington 1971, p.99).[31] However, conflicts between the Amish and school authorities continued in other states throughout the 1960s.

In 1969, three Amish parents were found guilty of violating Wisconsin's compulsory attendance laws for declining to send their children to public high school after finishing the eighth grade in Green County (Keim 1975, p.151).[32] Subsequent appeals to the circuit court failed. In 1971, the Amish brought the case to the Supreme Court of Wisconsin, which reversed the lower courts' decision. Unsatisfied with the result, the State of Wisconsin pressed on to the Supreme Court of the United States. On May 15, 1972, the U.S. Supreme Court ruled that the Amish had the right to refuse their children a high school education (Meyers 1993). Because of this decision, the Amish

---

[30] In Ohio, students would continue school until the tenth grade (Hostetler and Huntington 1971, p.72).
[31] The vocational school program was never started in other states (Huntington 1994, endnote 13).
[32] The Amish parents are Jonas Yoder, Adin Yutzy, and Wallace Miller. Yoder and Yutzy are Old Order Amish, while Miller is an Amish Mennonite.

were able to enforce the prohibition of high school education without governmental interference.

2.3 The Religious Club Model

In this section, I extend Iannaccone's (1992) religious club model to explain why the Amish would prohibit high school education and refuse to comply with compulsory schooling laws. According to the religious club model, the restriction on high school education allows the Amish to (1) internalize the positive externality generated from the social interaction among sect members; and (2) restrict sect membership only to those who do not "free-ride" in the club.

2.3.1 The Basic Model: Homogenous Type of Amish

Consider a model, where an Amish individual lives for two periods. In period 2, the utility of a baptized adult Amish is:

$$u_{2i} = u(S_i, R_i, Q), \text{ where } Q = \sum R_{-i} / N$$

An adult Amish derives utility from time spent in religious activities, *R*, as well as from the consumption of secular goods, *S*. Religious activities are more satisfying when there are more committed members engaged in them. The average amount of religious time spent by other adult Amish members, *Q*, is a positive externality and can be thought as the quality of the Amish "club". Mutual aid in the form of community members helping one another with barn raisings, quilt making, harvesting, weddings, and so on are typical examples of *Q*. For simplicity, assume the number of other members in the sect, *N*, is exogenously given.

Adults can participate in the labor force, resulting in a budget constraint of the form:

$$w_i T_2 = pS_i + w_i R_i$$

Total time available in the second period is T_2, which is spent on religious activity R and work hours h (i.e., $T_2 = R_i + h_i$). Income is earned at wage rate w_i per hour worked and spent on consumption of the secular good S, at price p.

The wage rate, w_i, is determined by the level of education chosen when the Amish person was young (i.e., period 1):

$$w_i = w(E_i), \text{ where } w'(E) > 0 .$$

The above equation describes labor productivity as a function of education.[33] The assumption that education can only be chosen when young is obviously unrealistic, but it is consistent with the observation that education is usually completed when young. Although it is also common that Amish parents make their children's schooling decisions, Amish youths may pursue higher level of education than the eighth grade during the time of *Rumspringa* and after leaving the sect, or taking a General Educational Development (GED) test after dropping out of school.[34]

In the first period, unbaptized individuals derive utility from leisure only:

$$u_{1i} = u(l_i)$$

The young Amish cannot work and must allocate total time T_1 between leisure l and education E:

---

[33] Alternatively, we may view education as a signal of (secular) labor productivity in the spirit of Spence (1973). This alternative view may be appropriate if we think that Amish education provides no human capital relevant for the secular labor market, but only serves to signal labor productivity.

[34] See McDonnell and Hurst (2006) for a discussion of these cases.

$$T_1 = l_i + E_i$$

Since a rational and forward-looking young Amish maximizes life-time utility subject to the time constraint in period 1, the problem is solved by backward induction. The period 2 problem is:

$$\underset{S_i, R_i}{\text{Max}} \; u_{2i} = u(S_i, R_i, Q),$$

$$\text{Subject to } \; w(E_i)T_2 = pS_i + w(E_i)R_i$$

Note that the adult individual takes the wage rate, *w* and the quality of the club, *Q*, as given in period 2.

Because the Amish individual does not take into consideration the positive externality generated by his religious activities, the chosen level of *R* and *S* will only satisfy the following condition:

$$\frac{w(E_i)}{p} = MRS_{RS}$$

The person ignores the external benefit of his religious participation, MRS_{QS}, that a social planner would consider in the following condition:

$$\frac{w(E_i)}{p} = MRS_{RS} + MRS_{QS}$$

Solving the period 2 problem yields the optimal consumption of the secular good $S_i^*(p, w(E_i); Q)$, the optimal level of religious activities $R_i^*(p, w(E_i); Q)$, and the indirect utility $v_{2i}(p, w(E_i); Q)$. Because the marginal external benefit of religious participation MRS_{QS} is not taken into consideration and that MRS_{RS} is decreasing in *R*, the privately chosen $R_i^*(p, w(E_i); Q)$ and $v_{2i}(p, w(E_i); Q)$ will be lower than the socially desired level.

Assuming no discount factor, the individual's problem in period 1 is:

$$\operatorname*{Max}_{E_i} V_i = u_{1i}(l_i) + v_{2i}(p, w(E_i), Q)$$

$$\text{Subject to } T_1 = l_i + E_i$$

The first order condition yields:

$$\frac{\partial v_{2i}}{\partial w_i}\frac{\partial w_i}{\partial E_i} = -\frac{\partial u_{1i}}{\partial E_i}$$

The left hand side term is the marginal benefit of education and the right hand side term is the marginal cost of education.[35] Since the individual will select *R* and *S* such that the condition $w(E_i)/p = MRS_{RS}$ holds (ignoring the term MRS_{QS}) in period 2, the utility maximizing E^* will be higher than the socially optimal level.

According to the religious prohibition interpretation, by imposing a level of education lower than the privately chosen level, the Amish sect can make labor market participation relatively less attractive and induce the socially optimal level of religious participation.

2.3.2 Unobserved Heterogeneity and Religious Sacrifice

When there are unobserved heterogeneous types of Amish persons, the sect can improve social welfare by requesting a signal from potential members in order to discourage free-riders from joining the club. The signal is interpreted as a religious sacrifice (Iannaccone 1992). Following Berman's (2000) exposition, assume two

[35] The term $\partial v_{2i}/\partial w_i = [\partial v_{2i}/\partial(w_i/p)][\partial(w_i/p)/\partial w_i]$ is non-negative by the property of an indirect utility function, which is non-increasing in (*p/w*) and (*p/w*) is decreasing in *w*. To obtain an interior solution, we need to assume that the Hessian matrix of the objective function is negative semi-definite.

unobserved types of individuals: high-type (H) Amish and low-type (L) Amish. For each birth cohort, *N*, the fraction of high-type Amish, θ_H, and the fraction of low-type Amish, $\theta_L = 1 - \theta_H$, are exogenously determined.[36] High-type Amish enjoy higher return to education in the labor market than low-types:[37]

$$w_H'(E) > w_L'(E)$$

Furthermore, assume that $w_H(0) \geq w_L(0)$, so that without education high-type Amish are more productive than low-type Amish in the secular labor market.

Given that high-types have a higher marginal benefit of education than low-types and that both types of Amish face the same marginal cost of education, high-type youths will optimally select more education than low-types, i.e., $E_H^* > E_L^*$. This means that a high-type adult will earn higher wages and participate less in religious activities than a low-type Amish would, i.e., $w_H > w_L$ and $R_H^* < R_L^*$.[38]

In the absence of an educational restriction, an Amish sect with predominantly low-type Amish persons will not gain from admitting a high-type Amish person because that person will lower the average level of religious participation in the sect and decrease the welfare of existing members. That is, because $V_L(p,w;Q_L) > V_L(p,w;Q_{H+L})$, where $Q_L > Q_{H+L}$, $Q_L \equiv \sum R_L^* / N_L$, $Q_{L+H} \equiv \left(\sum R_L^* + \sum R_H^*\right)/N$, and $N_L \equiv \theta_L(N+1) - 1$, low-type Amish enjoy higher *Q* and utility when high-type Amish are excluded from joining

[36] This assumption is restrictive because each generation of *N* is endogenously affected by Amish fertility through the level of prohibition and sacrifice set and θ_H may also be affected by assortative mating.

[37] Heterogeneity could alternatively be in preferences for religious activities at the margin. I chose heterogeneity in secular returns to education to simplify the exposition, as well as to focus on variables that have data available.

[38] As long as the substitution effect of a change in wage is greater than the income effect of a change in wage, the labor supply curve is upward sloping and religious participation is decreasing in wages.

the sect. If the sect imposes a ceiling on education equivalent to that chosen by low-type Amish and high-type Amish choose not to join, then the signal is not costly to low-types, but serves to exclude high-types.

However, if the level of education that low types optimally choose is not sufficiently low to deter high-type Amish persons from complying with, then the restriction is not effective in discourage free-riders from joining. That is, it is possible that high-type Amish enjoy a higher level of utility by complying with the low level of education and joining the low-type Amish sect than by forming their own group:

$$V_H\left(p, w\left(E_L^*\right); Q_{H+L}\right) > V_H\left(p, w\left(E_H^*\right); Q_H\right),$$

where $Q_{H+L} \equiv \left(\sum R_H\left(p, w\left(E_L^*\right)\right) + \sum R_L^*\right) / N$, and $Q_H \equiv \sum R_H^* / N_H$ and $N_H \equiv \theta_H\left(N+1\right)-1$. If this is the case, then E_L^* is not incentive compatible for the high-type Amish.

When E_L^* is not incentive compatible, the sect has to set the ceiling level of education $\overline{E}$, such that $\overline{E} < E_L^*$ to prevent high-type Amish persons from joining the sect. $\overline{E}$ is incentive compatible, such that:

$$V_H\left(p, w\left(\overline{E}\right); \overline{Q}\right) \leq V_H\left(p, w\left(E_H^*\right); Q_H\right),$$

where $\overline{Q} \equiv \left(\sum \overline{R}_L + \sum \overline{R}_H\right) / N$ and $\overline{R}_i\left(p, w\left(\overline{E}\right)\right)$ for $i = H, L$. Furthermore, $\overline{E}$ needs to satisfy the participation constraint:

$$V_L\left(p, w\left(\overline{E}\right); \overline{Q}_L\right) > V_L\left(p, w\left(E_L^*\right); Q_{H+L}\right),$$

where $\overline{Q}_L \equiv \sum \overline{R}_L / N_L$.

Choosing $\overline{E}$ is costly because if types were fully observable, low-type Amish would have chosen E_L^* and enjoyed a higher level of utility. We may view $\overline{E}$ as grade eight and E_L^* as a high school education. The willingness to sacrifice high school education sends a signal to the sect that a potential entrant is committed to the Amish life. High-type individuals will not comply with the educational restriction $\overline{E}$ and will choose to leave the sect. Since education can only be chosen when young in this simple model, the sacrifice is an “irreversible” act.

2.3.3 Government Enforcement of Compulsory Schooling Laws

When the government enforced compulsory high school attendance, the Amish could not achieve their socially efficient level of education. In the homogenous case, when the government enforced compulsory high school attendance, non-exempted Amish cohorts attended high school and tend to spend less time in religious activities. In the heterogeneous case, when compulsory schooling laws were enforced on the Amish, the Amish sect could not impose the optimal amount of religious sacrifice and admitted members who would lower the average level of religious participation in the sect. These explain why the Amish would refuse to comply with compulsory schooling laws.

2.3.4 Testable Implications

According to the Amish religious club model, the U.S. Supreme Court’s ruling permitted the Amish to enforce their socially optimal level of education. The compulsory schooling exemption switched the Amish from an environment in which they were constrained by the government when setting their optimal level of prohibition and sacrifice to one in which they were unconstrained. Thus, the model predicts that the exemption should have an immediate impact on the educational attainment of Amish individuals.

If the prohibition on high school education is used solely for internalizing the positive externality of religious activities, then we would not expect high-type Amish leaving the sect following the exemption. However, if the restriction on high school education is used as a religious sacrifice to screen out uncommitted members, then the Amish religious club model predicts that (1) Amish individuals tend to have lower labor market wage rates than former Amish or non-Amish individuals; (2) the surprising U.S. Supreme Court's ruling would lead individuals with high labor productivity to leave the sect; (3) the compulsory schooling exemption would encourage women with high shadow cost of child rearing to select out of the sect.

2.4 Data

Data were sourced from the U.S. censuses to test the religious club model predictions. The Census Bureau collected information of the language spoken at home in recent censuses. According to Meyers and Nolt (2005, p.61), the Amish and conservative Mennonites represent almost all of the current speakers of Pennsylvania Dutch, which is a German Dialect.[39] Pollack (1981) reported that as the Amish people shifted to more liberal Mennonite denominations, they ceased to use Pennsylvania Dutch as their primary

---

[39] Dutch comes from "Deutsch", meaning German. There are also a number of Amish who speak a Swiss-German dialect (Meyers and Nolt 2005, p.61). Since the 1980 Census, the Census Bureau began collecting information of the language spoken at home for persons above a certain age.

language, indicating that speaking Pennsylvania Dutch signals attachment to the Amish and conservative Mennonite Church.[40] Since I cannot directly identify the religious denominations of Pennsylvania Dutch speakers in the censuses, the Amish referred to in this paper would include some Conservative Mennonites who speak Pennsylvania Dutch.[41] Specifically, I define a person as an Amish individual when the person resides in a non-single-member household that has at least two Pennsylvania Dutch speakers.[42] For those who report to speak Pennsylvania Dutch, but live in a non-Pennsylvania Dutch household, I define them as former Amish individuals.[43] Since former Amish individuals may no longer speak Pennsylvania Dutch, this method of defining former Amish individuals is likely to lead to severe undercount.[44] Nevertheless, it provides some crude estimates of the characteristics of former Amish individuals.

Table 2.1 compares the distributions of Amish population estimates based on different sources of data. The distributions of Amish population estimates across the United States using the decennial censuses are fairly similar to the distributions of Amish population estimated by Kraybill and Hostetter (2001) and Hostetler (1993) using Amish

---

[40] In 1977, 100% of Old Order Amish families living in Plain City, Ohio, used Pennsylvania Dutch as their primary language, but only 11.8% of Mennonite families used it as their primary language.

[41] Conservative Mennonites (Old Order Mennonites) are similar with the Amish in many aspects, such as their plain clothing, horse-and-buggy mode of transportation, preference for one-room parochial schooling, and prohibition of high school education (Kraybill and Bowman 2001).

[42] In this paper, Pennsylvania Dutch speakers include those who speak Pennsylvania Dutch at home and those with Pennsylvanian German ancestry and speak German at home. I coded a household as a Pennsylvania Dutch household when the household has at least one Pennsylvania Dutch speaker while any other household members speak Pennsylvania Dutch, German, or Dutch. For those who are Dutch or German speaking, they must be native-born to be included. Pennsylvania Dutch speaking people living in single-member households are coded as Amish. If a person is the sole-speaker of Pennsylvania Dutch in a non-single-member household, I code the person as a former Amish person.

[43] A person must communicate with other household members using the same language. Individuals who are the sole-speakers of Pennsylvania Dutch at home are likely identifying Pennsylvania Dutch as their mother tongue, instead of "language spoken at home" per se.

[44] It is also unclear whether persons who are the sole speaker of Pennsylvania Dutch at home constitute a representative sample of all former Amish.

Church membership data, with the decennial censuses tending to undercount the total Amish population.[45] Since 70 percent of the Amish population resides in Pennsylvania, Ohio, and Indiana, I will focus my analysis on individuals living in these three states.

Table 2.2 and Table 2.3 compare the characteristics of the Amish, former Amish, and non-Amish adult population aged 18 to 64 in 1990 and 2000 respectively. Amish are more likely to drop out of school upon finishing grade eight, to be farmers, to have bigger families, and to be employed than former Amish and non-Amish individuals. A high fraction of adult Amish males and females have no more than an eighth-grade education. The fraction of eighth-grade dropouts is less than 5 percent for non-Amish population and around 15 percent for former Amish persons, but as much as two-thirds for the Amish.[46] Furthermore, the trend in educational attainment is decreasing for the Amish population, but increasing for the non-Amish and former Amish population. The low educational attainment reported here for the Amish is consistent with their objection to high school education. The educational attainment of the Amish is also much lower than members of other religious sects in the United States (Iannaccone 1992). Amish individuals tend to have higher employment rates, potentially because they refuse any form of government assistance, including unemployment insurance, and cannot devote as much time to job search when unemployed. It may also be because the Amish have a stronger social network, which makes finding employment easier than for non-Amish

[45] Since both methods provide estimates, it is not clear which one is closer to the truth. It is also not clear whether non-responses will bias the estimates of the characteristics of Amish. The undercount based on censuses may be due to church membership data that include “Swiss Amish” who do not speak Pennsylvania Dutch. Furthermore, the high percentage of children and young adults and the use of non-English language of the Amish are characteristics associated with census undercount. For detailed discussions of census undercount and the extent of undercount, see Edmonston and Schultze (1995) and Edmonston (2002).

[46] In this paper, eighth-grade dropout means having less than a ninth-grade education.

people.[47] The observation that Amish have larger family size is consistent with previous findings regarding the high fertility rates of the Amish.[48]

Table 2.2 and Table 2.3 also show that the Amish participate less in the labor force, have lower earnings, and work fewer hours on average than former Amish persons and non-Amish persons. The wage gap between the Amish and non-Amish populations is similar to the relative differences in incomes between adherents to most Church-like religious groups and sect members in the United States reported by Iannaccone (1992) and between Israeli Ultra-Orthodox Jews and non-Ultra-Orthodox Jews reported by Berman (2000).[49] However, the differences documented here do not imply that the Amish are disadvantaged. Indeed, the Amish eschew material wealth and many Amish activities and mutual aids are non-monetary in nature.

The simple comparison between Amish, former Amish, and non-Amish population shows that Amish individuals have fewer years of completed schooling and earn significantly less. The differences are consistent with the Amish religious club model's predictions. I will exploit the policy shock induced by the U.S. Supreme Court's decision to test the model's predictions in the next section.

2.5 Empirical Evidence

2.5.1 The Impact of Exemption on Dropout and Years of Completed Schooling

[47] Amish work for Amish employers, as well as non-Amish employers. For example, Kraybill (2001, p.247) reports that 11 percent of Old Order Amish adult men aged 21-30 living in Lancaster work for non-Amish employers.

[48] Amish total fertility rates were estimated to be between 6-8 (see Ericksen et al. 1979 and Greksa 2002).

[49] Comparisons based on household incomes reported in Table 1 and Table 2 in Iannaccone (1992) and Table 1 in Berman (2000).

Based on the pooled sample of 1990 and 2000 censuses, Figure 2.1 shows the fraction of eighth-grade dropouts by Amish and non-Amish birth cohorts.[50] The figure clearly reveals that the cohorts born before 1958, who reached age 14 before the Supreme Court's 1972 ruling in *Wisconsin v. Yoder* and were affected by compulsory high school attendance laws, are considerably more likely to have some high school education. In contrast, there is no discernible difference in the fraction of dropouts for non-Amish cohorts who are never exempted by the U.S. Supreme Court's decision.

I estimate the impact of the exemption on the probability of an Amish person not pursuing a high school education using the following linear probability model:

$$Dropout_i = \alpha_0 + \alpha_1 Post_i + X_i'\alpha + u_i$$

where *Dropout* takes the value of 1 if person *i* did not pursue a high school education upon completing grade eight, and 0 otherwise; *Post* indicates if the person was born in 1958 or after (exempted by compulsory high school attendance laws); *X* is a set of control variables, including metropolitan indicator and state dummies; and *u* is the error term. The coefficient α_1 measures the cohort differences in the likelihood of an Amish dropping out of school upon completing grade eight. Using only individuals born between 1956 and 1959 as the sample, we can avoid confounding cohort effects other than that due to the exemption.

Table 2.4 and Table 2.5 report the estimated α_1 for males and females, respectively. Columns (1) to (3) report estimates using Census 1990, columns (4) to (6) report estimates using Census 2000, and column (7) report estimates using pooled censuses. According to the preferred specification (columns 3 and 6) that controls for

[50] The analysis is restricted to adult males living in Pennsylvania, Ohio, and Indiana.

residential location, the effect of the exemption on the probability of an Amish male to drop out of school upon completing grade eight is estimated to be 24 percent based on Census 1990 data and 15 percent using Census 2000 data. On the other hand, the exemption is estimated to increase the likelihood of an Amish female not pursuing a high school education by 9 percent based on Census 1990 data and 13 percent based on Census 2000 data.

To examine how the exemption affected Amish completed years of schooling, I estimate the following regression model:

$$Educ_i = \alpha_0 + \alpha_1 Post_i + X_i'\alpha + u_i$$

where *Educ* is the years of completed schooling; *Post* equals 1 if the person was born in 1958 and after (exempted by compulsory high school attendance laws), and 0 otherwise; *X* is a set of control variables, including metropolitan indicator and state dummies; and *u* is the error term. The coefficient α_1 measures the effect of the exemption on Amish completed years of schooling.

Table 2.6 and Table 2.7 report the estimates for males and females, respectively. Columns (1) to (3) show estimates based on Census 1990 data, and columns (4) to columns (6) are based on Census 2000 data. The preferred specification in columns (3) and (6), the compulsory schooling exemption is estimated to decrease the average years completed schooling for Amish males by 0.7 years based on Census 1990 data and by 0.8 years using Census 2000 data. The estimated effect of the exemption on the average years of completed schooling for Amish females is -0.4 years using Census 1990 data and -1 year using Census 2000 data. The average years of completed schooling fell from roughly 9 years (indicated by the intercept terms) to approximately 8 years (see Figure 2.2).

The results show that the exemption permits the Amish to impose their restriction on high school education. The restriction raised the probability of not pursuing a high school education and reduced the average years of completed schooling for both Amish males and females.

2.5.2 Amish Cohort Differences in Earnings and Fertility

2.5.2.1 Log Hourly Earnings

Since Amish women have low labor force participation rates, the analysis will focus on Amish males only. Table 2.8 reports the estimated cohort differences in the log hourly earnings of Amish males based on the following regression model:

$$Log\left(Earnings_i\right) = \beta_0 + \beta_1 Post_i + X_i'\beta + \varepsilon_i$$

where *Log(Earnings)* is the log hourly earnings; *Post* equals 1 if the person was born in 1958 and after (exempted by compulsory high school attendance laws), and 0 otherwise; *X* is a set of control variables, including metropolitan indicator, state dummies, marital status, potential experience, and potential experience squared; and ε is the error term.

The estimates reported in columns (1) to (3) and columns (4) to (6) are based on Census 1990 data and Census 2000 data, respectively. Column (7) reports estimates based on pooled census data. The estimated cohort differences reported in columns (1) and (2) or (4) and (5) are similar whether or not indicators for metropolitan status and marital status are included as regressors. Estimates based on Census 1990 show that the exempted Amish cohorts earned roughly 23 percent less than non-exempted Amish cohorts. Estimates based on Census 2000 indicate that exempted Amish cohorts earned approximately 34 percent less than non-exempted Amish cohorts. Similarly, the estimate based on pooled censuses (Column 7) shows that exempted Amish cohorts earned 29 percent less than non-exempted Amish cohorts.

Since wage is likely to grow with age and work experience, especially for prime working age males, it is possible that the earnings differences presented in columns (1), (2), (4), and (5) of Table 2.8 are not totally due to the U.S Supreme Court's decision. Column (3) shows that the estimated cohort difference becomes greater when potential work experience is controlled for, while column (6) shows that the estimated cohort difference is not sensitive to the control variables. However, because exempted Amish cohorts are less educated and they started accumulating work experience at younger ages due to the exemption, potential experience is endogenous. Moreover, as the samples cover only four age cohorts, the variation in potential experience is primarily driven by small differences in ages, which are also correlated with the variable *Post*.[51] Hence, including potential work experience as a regressor might actually confound the estimated effect of the exemption on log hourly earnings. Given the problems and the little gain associated with controlling for potential experience, estimates without controlling for potential experience are preferred. I will deal with the problems of age and experience in the next section using a difference-in-difference estimator.

2.5.2.2 Fertility

Table 2.9 presents the estimated cohort differences in Amish fertility using the following regression model:

[51] We may also widen the age window, but that may introduce model specification bias if the effect of age does not follow the specified functional form.

$$Chborn_i = \beta_0 + \beta_1 Post_i + X_i'\beta + \varepsilon_i$$

where *Chborn* is the number of children ever born to a woman; *Post* equals 1 if the woman was born in 1958 and after (exempted by compulsory high school attendance laws), and 0 otherwise; *X* is a set of control variables, including metropolitan indicator, state dummies, and marital status; and ε is the error term. Because Census 2000 did not collect fertility information, only estimates based on Census 1990 data are reported.

Columns (1) to (3) present the estimated cohort differences in fertility without controlling for age. Column (1) and (2) show that exempted Amish women have higher fertility than non-exempted women, who are older. The difference in fertility is roughly 0.35 children. Column (3) shows that controlling for marital status significantly reduces the cohort difference in fertility; exempted Amish women have 0.16 more children than non-exempted women, although the difference is not statistically significant. Given that younger women generally have fewer children than older women, controlling for age may lead to an even greater estimated cohort difference. Column (4) indicates that the estimated cohort difference becomes 0.79 children when age is controlled for. Columns (5) and (6) show that if differential effects of age are allowed for exempted and non-exempted, the cohort difference increases to approximately 0.9 children. However, because the sample covers very few age groups, the estimated cohort differences which have age effects adjusted for are difficult to interpret. Therefore, the estimated cohort difference presented in column (3) is preferred.

2.5.3 Non-Amish Cohort Differences and Difference-in-Difference Estimates

Given the difficulty associated with controlling for age or potential experience in estimating the effects of the exemption on log hourly earnings and fertility, we may use non-Amish individuals as a control group to implement a difference-in-difference estimator to difference out age or work experience specific effect. We can attribute non-Amish cohort differences as differences that are present in the absence of the U.S. Supreme Court's decision.

2.5.3.1 Non-Amish Cohort Differences

Table 2.10 and Table 2.11 report the estimates of non-Amish cohort differences in log hourly earnings and fertility, respectively. Although non-Amish individuals are not exempted from compulsory high school attendance, Table 2.10 shows that cohorts born in 1958 and 1959 earned roughly 1 percent lower in 1990 and 5 percent lower in 2000 than cohorts born in 1956 and 1957. Figure 2.3 contrasts the cohort differences in average log hourly earnings between Amish males and non-Amish males using the pooled census data. Similarly, Table 2.11 shows that there are cohort differences in fertility for non-Amish individuals. Columns (1) to (3) indicate that the younger cohorts have between 0.12 and 0.13 fewer children than older cohorts. On the other hand, columns (4) to (6) show that once controlling for age, younger cohorts are estimated to have more children than older cohorts, although the difference is not greater than 0.03 children. Given the problems associated with controlling for age previously discussed, estimates without controlling for age are preferred. Figure 2.4 compares the cohort differences in average children ever born between Amish and non-Amish females using census 1990 data. In evaluating the effects of the U.S Supreme Court's decision on log hourly earnings and fertility, we must control for the pre-existing cohort differences.

2.5.3.2 Difference-in-Difference Estimates

Table 2.13 presents the difference-in-difference estimates of the effects of the exemption on log hourly earnings and fertility, respectively. Columns (1) and (2) indicate that the U.S Supreme Court's decision led to a fall in hourly earnings of approximately 20 percent based on Census 1990 data. Columns (3) and (4) show that the exemption decreased hourly earnings by 34 percent based on Census 2000 data. These estimates are very similar to the Amish cohort differences presented in Table 2.8; suggesting cohort effects not due to the exemption are small.

According to Column (5) of Table 2.13, the U.S Supreme Court's decision is estimated to increase fertility by 0.28 births. This estimated effect is much larger than the Amish cohort differences presented in column (3) of Table 2.9, because exempted Amish would have been expected to have fewer children if the exemption were not in place according to the non-Amish cohort difference.

2.5.4 Implied Effects of Education on Log Hourly Earnings and Fertility

The estimates presented above show that the compulsory schooling exemption led to lower educational attainment, decreased earnings, and higher fertility. If we attribute the decreased earnings and increased fertility solely to the change in completed years of schooling driven by the U.S Supreme Court's decision, we could estimate the implied returns to education and the implied effect of education on fertility using an instrumental variable estimator. The second-stage instrumental variable regression is:

$$Outcome_i = \pi_0 + \pi_1 Educ_i + \pi_2 Post_i + \pi_3 Amish_i + X_i'\pi + \varepsilon_i$$

and the first-stage instrumental variable regression is:

$$Educ_i = \delta_0 + \delta_1 (Post_i \times Amish_i) + \delta_2 Post_i + \delta_3 Amish_i + X_i'\delta + v_i$$

The dependent variable *Outcome* is *Log(Earnings)* or *Chborn*; the variable *Post* × *Amish* serves as the excluded instrument; *Amish* takes the value of 1 for an Amish person, 0 otherwise; and *X* is a set of Amish specific and non-Amish specific controls. The coefficient of interest is π_1, which measures the return to education when *Outcome* is *Log*(*Earnings*) and the effect of education on fertility when *Outcome* is *Chborn*.

The estimated π_1 does *not* represent the *causal* effect of education, because the instrumental variable does not meet the exogenous condition required for the identification of the causal effect of education. As the Amish religious club model predicts that individuals with high labor market productivity selected out of the Amish sect, while individuals with low labor market productivity selected into the Amish sect following the U.S Supreme Court's decision, we expect the excluded instrument to be correlated with the error term in the outcome equation.

Table 2.14 columns (1) to (5) report the estimated implied returns to education; and column (6) reports the implied effect of education on fertility. The specification that includes a set of controls for metropolitan status, marital status, and state of residence is preferred. The estimated implied return to education is large: 23 percent using Census 1990 data; 32 percent using Census 2000 data; and 30 percent using pooled census data.

Similarly, the implied effect of education on fertility is also large: a one-year decrease in completed years of schooling predicts 0.91 more births.

Past studies estimated that the causal return to an additional year of schooling ranges between 7 percent and 15 percent (Angrist and Krueger 1991; Card 1999; Maluccio 1997). The estimated return to education of the Amish presented in Table 2.14 is at least 50 percent higher than the largest point estimate previously reported. It is difficult to conceive that the majority of non-exempted Amish who attended classes once a week for one additional year could possibly get as much as a 23 percent to 32 percent return on education. Indeed, Table 2.15 columns (1) to (6) show that when a cross-sectional sample of Amish individuals aged 20 to 50 is used to estimate the returns to education, every additional year of schooling is predicted to raise hourly earnings by only 2 to 5 percent. The low estimated returns to education for the Amish are remarkably similar to those of other religious sects as shown by Berman and Stepanyan (2005) and Berman (2000).

The estimated effect of education on fertility reported in columns (6) of Table 2.14 is also significantly larger than past estimates. For example, Osili and Long (2008) estimated that the causal effect of an additional year of education on fertility in Nigeria was between -0.26 and -0.48 births.[52] When we use a cross-sectional sample of Amish women aged 20 to 50 years to estimate the effect of education on children ever born, the estimate ranges between -0.15 to -0.24 depending on specifications (Table 2.15 columns 7 and 8). The implied effect of education on fertility reported in Table 2.14 is roughly

[52] Nigerian average years of schooling of approximately 5 years (Osili and Long 2008) and total fertility rate of 6 (National Population Comission 2000) in 1990 are not too different from those of the Amish.

twice the largest estimate produced by Osili and Long (2008) and almost thrice the largest OLS estimate reported in Table 2.15.

For the implied effect of education on log hourly earnings or fertility to be so large, we would need exempted Amish individuals with high labor market productivity leaving the sect and lowering the average hourly earnings or raising the average fertility more than the causal effect of education suggests. To illustrate how this selection affects the estimates, decompose the instrumental variable estimator into the true causal effect and bias:

$$\hat{\pi}_1^{IV} \xrightarrow{p} \pi_1 + \frac{\operatorname{cov}(z_i, \varepsilon_i \mid X)}{\operatorname{cov}(z_i, Educ_i \mid X)}$$

where z is the excluded instrument *Amish* × *Post*; π_1 is the true causal effect; $\operatorname{cov}(z_i, \varepsilon_i \mid X)/\operatorname{cov}(z_i, Educ_i \mid X)$ is the bias; and X represents all other regressors. According to Table 4 and Table 5, we know that $\operatorname{cov}(z_i, Educ_i \mid X) < 0$. For $\hat{\pi}_1^{IV} > \pi_1$, it must be the case that $\operatorname{cov}(z_i, \varepsilon_i \mid X) < 0$ when estimating the return to education. That is, exempted Amish individuals have unobserved characteristics that are negatively correlated with labor productivity. Similarly, for $\hat{\pi}_1^{IV} < \pi_1$ when estimating the effect of education on fertility, we need $\operatorname{cov}(z_i, \varepsilon_i \mid X) > 0$, which is consistent with exempted Amish females having lower shadow cost of child rearing.

The large estimated effects of education on log hourly earnings and fertility as implied by the surprising U.S Supreme Court's decision provide strong evidence that the Amish use the restrictions on secular high school education as a religious sacrifice to screen committed members.

2.6 Conclusion

Given the positive returns to education, Amish prohibition on high school education appears puzzling from a rational choice perspective. This paper extends Iannaccone's (1992) religious club model to explain why the Amish would collectively restrict education. According to the religious club interpretation, restricting secular education helps the Amish internalize the positive externality of religious participation and prevent less committed individuals from joining the sect. Because the enforcement of compulsory high school attendance by the government interfered with Amish community's socially efficient level of education, the Amish refused to comply.

Interpreting the restriction on secular education as a religious sacrifice is testable. When the government was enforcing compulsory schooling laws on the Amish, Amish born individuals with high labor productivity and low religious participation (high-type Amish) could legitimately attend high school. These high-type individuals would have been excluded from joining the sect if the Amish could effectively request them to sacrifice high school education as a signal of their commitment. The surprising U.S Supreme Court's decision in 1972, which exempts the Amish from compulsory education beyond the eighth grade, permits the Amish to enforce their desired level of religious sacrifice. This increased religious sacrifice predicts that high-type Amish would leave the sect following the exemption.

I use U.S. Census data to test the predictions of the Amish religious club model. First, I find that former Amish persons are more educated and enjoy relatively higher earnings than Amish persons. Second, exempted Amish cohorts have significantly lower educational attainment than non-exempted Amish cohorts. Third, the exemption led to lower earnings and higher births. The estimated effect of each additional year of education on log hourly earnings (between 0.23 and 0.32) implied by the exemption is at least half times greater than past causal estimates. Similarly the estimated effect of education on fertility (-0.91 births) is also more than twice the magnitude of past estimates. The large implied effects of education provide strong evidence that individuals with high labor productivity and high shadow cost of child rearing select out of the sect following the U.S. Supreme Court's ruling.

CHAPTER 2 - FIGURES

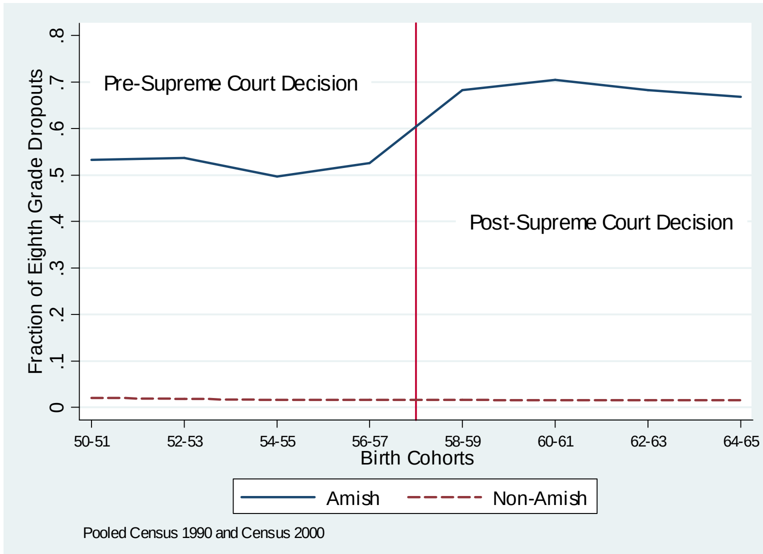

Figure 2.1: Cohort Differences in Eighth Grade Dropout

Notes: Author's own calculation based on pooled Census data sourced from Ruggles et al. (2004). Sample includes Amish and non-Amish residing in Pennsylvania, Indiana, and Ohio. Non-Amish are native born white population. Eighth grade dropout means having no more than an eighth grade education.

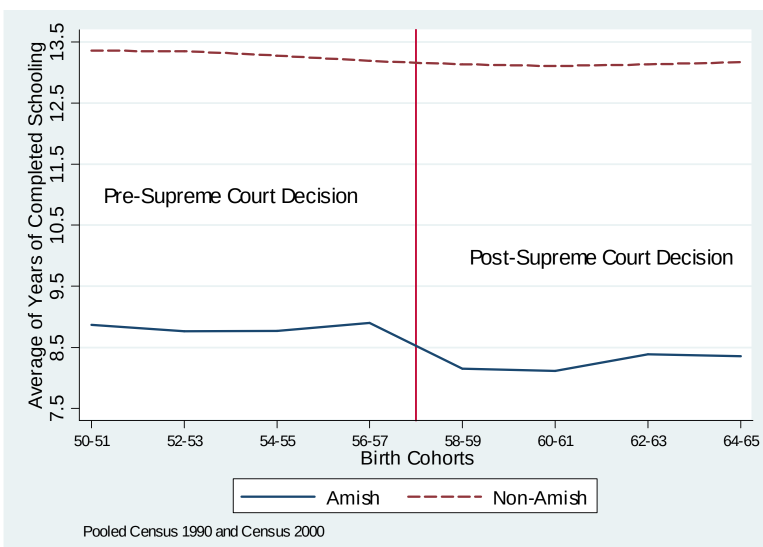

Figure 2.2: Cohort Differences in Average Years of Completed Schooling

Notes: Author's own calculation based on pooled Census data sourced from Ruggles et al. (2004). Sample includes Amish and non-Amish residing in Pennsylvania, Indiana, and Ohio. Non-Amish are native born white population.

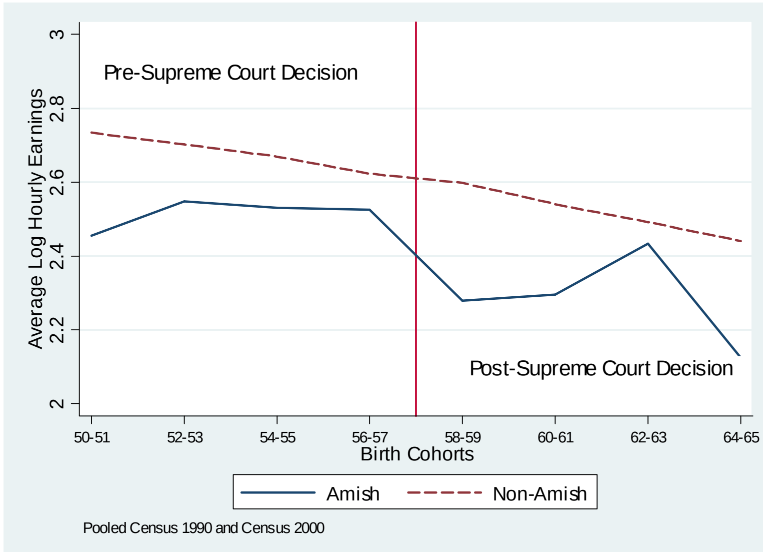

Figure 2.3: Cohort Differences in Average Log Hourly Earnings

Notes: Author's own calculation based on pooled Census data sourced from Ruggles et al. (2004). Sample includes Amish and non-Amish males residing in Pennsylvania, Indiana, and Ohio. Self-employed individuals with non-positive earnings are excluded from the sample. Non-Amish are native born white men.

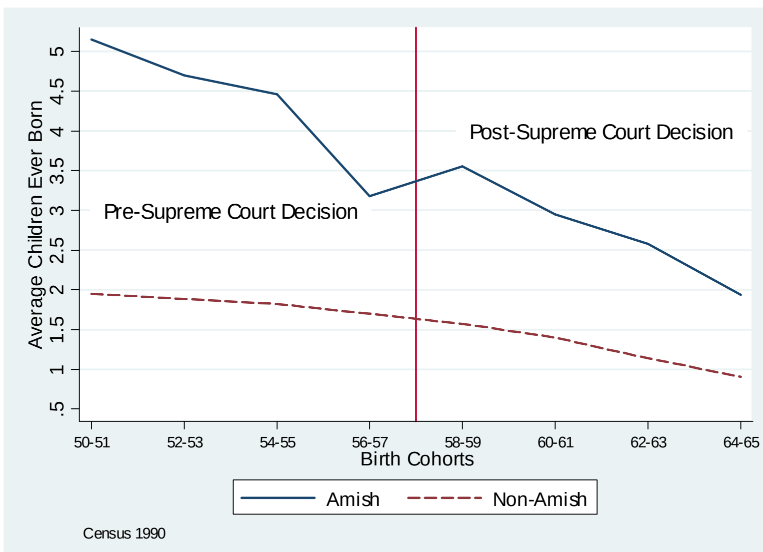

Figure 2.4: Cohort Differences in Average Children Ever Born

Notes: Author's own calculation based on Census 1990 data sourced from Ruggles et al. (2004). Sample includes Amish and non-Amish females residing in Pennsylvania, Indiana, and Ohio. Non-Amish are native born white women.

CHAPTER 2 - TABLES

Table 2.1: Pennsylvania Dutch Speakers and Amish Population Estimates by States

States	(1) All Amish Groups 2000	(2) Penn. Dutch Speakers 2000	(3) Old Order Amish 1992	(4) Penn. Dutch Speakers 1990
Ohio	51,302	22,321	43,200	16,705
Pennsylvania	47,860	47,137	35,200	51,394
Indiana	34,786	11,081	25,200	10,118
Wisconsin	9,561	4,994	7,800	1,583
Michigan	8,591	2,698	6,500	1,595
Missouri	6,701	3,230	5,200	2,474
Kentucky	6,042	2,306	1,500	1,207
Illinois	4,849	1,749	3,200	1,002
Iowa	4,775	1,683	3,700	1,299
New York	4,748	3,694	4,700	2,477
Tennessee	2,248	755	800	882
Kansas	1,599	478	800	848
Minnesota	1,574	490	1,500	691
Virginia	1,390	265	0	675
Maryland	1,127	1,097	1,000	1,740
Other states	5,199	4,590	1,600	3,606
Total	192,352	108,568	141,900	98,296

Note: (1) Kraybill and Hostetter's (2001) estimates of Old Order Amish, New Order Amish, Beachy Amish, and Amish Mennonites; (2) Pennsylvania Dutch speaking households in Census 2000; (3) Hostetler's (1993) estimates of Old Order Amish; (4) Pennsylvania Dutch speaking households in Census 1990.

Table 2.2: Descriptive Statistics by Groups – Census 1990

	Amish		Former Amish		Non-Amish		Amish – Former Amish
	Obs.	Mean	Obs.	Mean	Obs.	Mean	Differences
Male	34,773	0.48	6,460	0.54	16,726,052	0.49	-0.06
		(0.50)		(0.50)		(0.50)	(0.01)***
Metropolitan	34,773	0.59	6,460	0.72	16,726,052	0.74	-0.13
		(0.49)		(0.45)		(0.44)	(0.01)***
Married	34,773	0.70	6,460	0.81	16,726,052	0.61	-0.11
		(0.46)		(0.40)		(0.49)	(0.01)***
Family size	34,773	5.31	6,460	3.11	16,726,052	2.99	2.20
		(2.97)		(1.34)		(1.50)	(0.02)***
8th grade dropout	34,773	0.62	6,460	0.17	16,726,052	0.04	0.44
		(0.49)		(0.38)		(0.19)	(0.01)***
Years of education	34,773	8.54	6,460	11.40	16,726,052	12.76	-2.86
		(2.62)		(2.76)		(2.38)	(0.04)***
Lab. force participation	34,773	0.68	6,460	0.77	16,726,052	0.76	-0.09
		(0.47)		(0.42)		(0.43)	(0.01)***
Employed	23,732	0.99	4,989	0.98	12,753,372	0.94	0.01
		(0.12)		(0.15)		(0.24)	(0.002)***
Farmer	34,773	0.10	6,460	0.02	16,726,052	0.004	0.08
		(0.30)		(0.15)		(0.06)	(0.003)***
Weekly earnings	23,494	435.66	5,241	467.09	13,481,540	470.92	-31.43
		(1015.43)		(450.15)		(726.13)	(9.09)***
Hourly earnings	23,494	11.30	5,241	11.23	13,481,540	12.27	0.07
		(27.89)		(10.11)		(60.99)	(0.23)
Log hourly earnings	23,494	1.93	5,241	2.22	13,481,540	2.18	-0.29
		(0.95)		(0.61)		(0.75)	(0.01)***
Weeks worked yearly	34,773	32.03	6,460	37.76	16,726,052	35.81	-5.73
		(23.37)		(20.83)		(21.19)	(0.29)***
Hours worked weekly	34,773	30.56	6,460	33.48	16,726,052	31.67	-2.92
		(24.85)		(19.10)		(18.64)	(0.27)***

Note: Native-born adult population aged 18 to 64 living in Pennsylvania, Indiana, and Ohio. Eighth grade dropout means having no more than an eighth-grade education. Years of education was coded according to Park's (1994) method. Former Amish are sole speakers of Pennsylvania Dutch in non-single-member households. Non-positive earnings were dropped. Robust standard errors are reported in parentheses. *** significant 1% ** significant 5% * significant 10%

Table 2.3: Descriptive Statistics by Groups – Census 2000

	Amish		Former Amish		Non-Amish		Amish – Former Amish
	Obs.	Mean	Obs.	Mean	Obs.	Mean	Difference
Male	35,617	0.50	5,587	0.49	17,333,458	0.49	0.01
		(0.50)		(0.50)		(0.50)	(0.01)
Metropolitan	35,617	0.52	5,587	0.70	17,333,458	0.78	-0.17
		(0.50)		(0.46)		(0.41)	(0.01)***
Married	35,617	0.70	5,587	0.81	17,333,458	0.58	-0.11
		(0.46)		(0.39)		(0.49)	(0.01)***
Family size	35,617	5.53	5,587	2.94	17,333,458	2.80	2.59
		(2.90)		(1.43)		(1.49)	(0.02)***
8th grade dropout	35,617	0.65	5,587	0.11	17,333,458	0.02	0.54
		(0.48)		(0.32)		(0.14)	(0.005)***
Years of education	35,617	8.36	5,587	11.98	17,333,458	13.12	-3.63
		(2.42)		(2.54)		(2.30)	(0.04)***
Lab. force participation	35,617	0.66	5,587	0.78	17,333,458	0.77	-0.12
		(0.47)		(0.42)		(0.42)	(0.01)***
Employed	23,459	0.98	4,331	0.99	13,354,908	0.95	-0.02
		(0.15)		(0.08)		(0.22)	(0.002)***
Farmer	35,617	0.09	5,587	0.02	17,333,458	0.003	0.07
		(0.28)		(0.13)		(0.05)	(0.002)***
Weekly earnings	23,710	606.33	4,660	727.16	14,358,417	703.73	-120.83
		(951.25)		(1249.3)		(1289.24)	(19.31)***
Hourly earnings	23,710	15.98	4,660	16.56	14,358,417	17.61	-0.58
		(41.18)		(28.15)		(79.81)	(0.49)
Log hourly earnings	23,710	2.30	4,660	2.54	14,358,417	2.54	-0.24
		(0.92)		(0.68)		(0.74)	(0.01)***
Weeks worked yearly	35,617	31.19	5,587	39.70	17,333,458	37.76	-8.51
		(23.61)		(19.99)		(20.52)	(0.30)***
Hours worked weekly	35,617	28.46	5,587	35.15	17,333,458	33.10	-6.69
		(24.19)		(19.61)		(18.46)	(0.29)***

Note: Native-born adult population aged 18 to 64 living in Pennsylvania, Indiana, and Ohio. Eighth grade dropout means having no more than an eighth-grade education. Years of education was coded according to Park's (1994) method. Former Amish are sole speakers of Pennsylvania Dutch in non-single-member households. Non-positive earnings were dropped. Robust standard errors are reported in parentheses. *** significant 1% ** significant 5% * significant 10%

Table 2.4: Amish Male Cohort Differences in High School Dropout Likelihood

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	---------- Census 1990 ----------			---------- Census 2000 ----------			Pooled
Post (=1)	0.25	0.25	0.24	0.15	0.15	0.15	0.20
	(0.02)***	(0.02)***	(0.02)***	(0.03)***	(0.03)***	(0.03)***	(0.02)***
Metropolitan (=1)		-0.04	-0.06		0.01	-0.01	-0.04
		(0.02)*	(0.02)**		(0.03)	(0.03)	(0.02)**
Indiana (=1)			0.12			0.06	0.10
			(0.03)***			(0.04)	(0.02)***
Ohio (=1)			-0.12			-0.06	-0.09
			(0.03)***			(0.03)*	(0.02)***
Yr. 2000 (=1)							-0.04
							(0.02)**
Constant	0.50	0.52	0.54	0.49	0.49	0.51	0.54
	(0.02)***	(0.02)***	(0.03)***	(0.02)***	(0.02)***	(0.03)***	(0.02)***
Observations	1834	1834	1834	1275	1275	1275	3109
R-squared	0.06	0.07	0.09	0.02	0.02	0.03	0.06

Note: Author's estimates using Census data sourced from Ruggles et al. (2004). The dependent variable, *Dropout*, means having no more than an eighth grade education. The omitted state is Pennsylvania. Cohorts born between 1956 and 1959 are included in the sample. Exempted cohorts were born in 1958 and 1959. Robust standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.5: Amish Female Cohort Differences in High School Dropout Likelihood

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	---------- Census 1990 ----------			---------- Census 2000 ----------			Pooled
Post (=1)	0.08	0.06	0.09	0.14	0.11	0.13	0.11
	(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***
Metropolitan (=1)		-0.24	-0.20		-0.20	-0.21	-0.20
		(0.02)***	(0.02)***		(0.02)***	(0.03)***	(0.02)***
Indiana (=1)			0.32			0.21	0.28
			(0.02)***			(0.04)***	(0.02)***
Ohio (=1)			-0.06			-0.05	-0.05
			(0.03)*			(0.03)*	(0.02)**
Yr. 2000 (=1)							-0.001
							(0.02)
Constant	0.57	0.73	0.65	0.54	0.65	0.64	0.65
	(0.02)***	(0.02)***	(0.03)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***
Observations	1760	1760	1760	1770	1770	1770	3530
R-squared	0.01	0.06	0.13	0.02	0.06	0.07	0.10

Note: Author's estimates using Census data sourced from Ruggles et al. (2004). The dependent variable, *Dropout*, means having no more than an eighth grade education. The omitted state is Pennsylvania. Cohorts born between 1956 and 1959 are included in the sample. Exempted cohorts were born in 1958 and 1959. Robust standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.6: Amish Male Cohort Differences in Mean Years of Completed Education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	---------- Census 1990 ----------			---------- Census 2000 ----------			Pooled
Post (=1)	-0.87	-0.78	-0.70	-0.75	-0.76	-0.76	-0.77
	(0.14)***	(0.14)***	(0.14)***	(0.14)***	(0.14)***	(0.14)***	(0.10)***
Metropolitan (=1)		0.68	0.43		-0.23	-0.08	0.23
		(0.14)***	(0.18)**		(0.14)	(0.15)	(0.12)*
Indiana (=1)			-0.85			0.28	-0.42
			(0.14)***			(0.19)	(0.11)***
Ohio (=1)			-0.56			0.37	-0.12
			(0.24)**			(0.17)**	(0.15)
Yr. 2000 (=1)							0.62
							(0.10)***
Constant	8.64	8.20	8.58	9.20	9.33	9.08	8.56
	(0.10)***	(0.12)***	(0.18)***	(0.10)***	(0.14)***	(0.16)***	(0.13)***
Observations	1834	1834	1834	1275	1275	1275	3109
R-squared	0.02	0.03	0.04	0.02	0.02	0.03	0.04

Note: Author’s estimates using Census 1990 data sourced from Ruggles et al. (2004). The dependent variable, *Educ*, is years of completed education based on Park’s (1994) code. The omitted state is Pennsylvania. Cohorts born between 1956 and 1959 are included in the sample. Exempted cohorts were born in 1958 and 1959. Robust standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.7: Amish Female Cohort Differences in Mean Years of Completed Education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	---------- Census 1990 ----------			---------- Census 2000 ----------			Pooled
Post (=1)	-0.45	-0.31	-0.42	-0.96	-0.84	-0.99	-0.68
	(0.12)***	(0.12)***	(0.11)***	(0.11)***	(0.12)***	(0.13)***	(0.09)***
Metropolitan (=1)		1.57	1.35		0.74	0.85	1.09
		(0.11)***	(0.14)***		(0.12)***	(0.12)***	(0.09)***
Indiana (=1)			-1.39			-1.00	-1.31
			(0.11)***			(0.12)***	(0.08)***
Ohio (=1)			0.07			0.41	0.24
			(0.16)			(0.15)***	(0.11)**
Yr. 2000 (=1)							-0.02
							(0.08)
Constant	8.81	7.75	8.17	9.04	8.62	8.58	8.41
	(0.09)***	(0.10)***	(0.15)***	(0.10)***	(0.12)***	(0.12)***	(0.11)***
Observations	1760	1760	1760	1770	1770	1770	3530
R-squared	0.01	0.09	0.13	0.04	0.06	0.07	0.10

Note: Author’s estimates using Census data sourced from Ruggles et al. (2004). The dependent variable, *Educ*, is years of completed education based on Park’s (1994) code. The omitted state is Pennsylvania. Cohorts born between 1956 and 1959 are included in the sample. Exempted cohorts were born in 1958 and 1959. Robust standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.8: Amish Male Cohort Differences in Log Hourly Earnings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	---------- Census 1990 ----------			---------- Census 2000 ----------			Pooled
Post (=1)	-0.22	-0.23	-0.30	-0.35	-0.34	-0.34	-0.29
	(0.05)***	(0.05)***	(0.05)***	(0.04)***	(0.05)***	(0.05)***	(0.03)***
Metropolitan (=1)	0.23	0.22	0.21	-0.02	-0.01	-0.02	0.13
	(0.05)***	(0.05)***	(0.05)***	(0.04)	(0.04)	(0.04)	(0.03)***
Married (=1)		0.21	0.30		0.03	-0.03	0.18
		(0.06)***	(0.05)***		(0.05)	(0.05)	(0.04)***
Exp.			-0.01			0.19	
			(0.07)			(0.12)	
Exp. squared			-0.00			-0.00	
			(0.00)			(0.00)	
Indiana (=1)	0.26	0.24	0.28	0.38	0.38	0.38	0.29
	(0.07)***	(0.07)***	(0.07)***	(0.04)***	(0.04)***	(0.04)***	(0.05)***
Ohio (=1)	-0.06	-0.08	-0.10	0.44	0.43	0.45	0.17
	(0.05)	(0.06)	(0.06)*	(0.05)***	(0.05)***	(0.05)***	(0.04)***
Yr. 2000 (=1)							0.33
							(0.03)***
Constant	2.20	2.04	2.69	2.57	2.54	-0.14	2.08
	(0.05)***	(0.07)***	(0.58)***	(0.04)***	(0.07)***	(1.61)	(0.05)***
Observations	1650	1650	1650	1172	1172	1172	2822
R-squared	0.04	0.04	0.07	0.12	0.12	0.12	0.07

Note: Author's estimates using Census data sourced from Ruggles et al. (2004). The dependent variable, *Log(Earnings)*, is the log hourly wage salary and business or farm income. Self employed individuals with non-positive earnings are excluded from the sample. *Experience* = *Age* – *Educ* – 6. The omitted state is Pennsylvania. Cohorts born between 1956 and 1959 are included in the sample. Exempted cohorts were born in 1958 and 1959. Robust standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.9: Amish Female Cohort Differences in Fertility – Census 1990

	(1)	(2)	(3)	(4)	(5)	(6)
Post (=1)	0.35	0.34	0.16	0.79	0.90	0.87
	(0.12)***	(0.12)***	(0.10)	(0.20)***	(0.17)***	(0.13)***
Metropolitan (=1)		-0.35	-0.30	-0.35	-0.18	-0.37
		(0.14)**	(0.11)***	(0.11)***	(0.09)*	(0.07)***
Married (=1)			3.45	3.45	3.42	3.80
			(0.09)***	(0.09)***	(0.07)***	(0.05)***
Age (scaled)				0.36	0.45	0.49
				(0.10)***	(0.08)***	(0.04)***
Post X Age					-0.10	-0.30
					(0.10)	(0.04)***
Indiana (=1)	0.71	0.57	0.30	0.33	0.30	0.34
	(0.15)***	(0.16)***	(0.12)**	(0.12)***	(0.10)***	(0.08)***
Ohio (=1)	0.53	0.34	-0.25	-0.28	-0.16	-0.30
	(0.17)***	(0.18)*	(0.16)	(0.16)*	(0.13)	(0.08)***
Constant	2.97	3.26	0.67	0.21	0.01	-0.21
	(0.09)***	(0.14)***	(0.12)***	(0.15)	(0.17)	(0.12)*
Observations	1760	1760	1760	1760	2561	4268
R-squared	0.02	0.02	0.32	0.33	0.32	0.43

Note: Author's estimates using Census 1990 data sourced from Ruggles et al. (2004). The dependent variable, *Chborn*, is the number of children ever born to a woman. The omitted state is Pennsylvania. Columns (1) to (4) use cohorts born between 1956 and 1959 as the sample; column (5) uses cohorts born between 1955 and 1960 as the sample; and column (6) uses cohorts born between 1953 and 1962 as the sample. The variable age is scaled to zero for individuals aged 32 years old (born in 1958). Exempted cohorts were born in 1958 and after. Robust standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.10: Non-Amish Male Cohort Differences in Log Hourly Earnings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	---------- Census 1990 ----------			---------- Census 2000 ----------			Pooled
Post (=1)	-0.05	-0.04	-0.18	-0.01	-0.01	-0.20	-0.02
	(0.001)***	(0.001)***	(0.001)***	(0.002)***	(0.002)***	(0.002)***	(0.001)***
Metro. (=1)	0.14	0.16	0.10	0.17	0.18	0.10	0.17
	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.001)***
Married (=1)		0.28	0.26		0.32	0.26	0.30
		(0.002)***	(0.001)***		(0.002)***	(0.002)***	(0.001)***
Exp.			-0.15			-0.31	
			(0.002)***			(0.003)***	
Exp. Sq.			0.003			0.005	
			(0.0001)***			(0.0001)***	
Indiana (=1)	-0.03	-0.03	-0.02	0.01	0.01	0.02	-0.01
	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.001)***
Ohio (=1)	-0.02	-0.03	-0.02	-0.004	0.001	-0.002	-0.01
	(0.002)***	(0.002)***	(0.001)***	(0.002)**	(0.002)	(0.002)	(0.001)***
Yr. 2000 (=1)							0.43
							(0.001)***
Constant	2.33	2.11	3.74	2.71	2.46	7.43	2.07
	(0.002)***	(0.002)***	(0.013)***	(0.002)***	(0.002)***	(0.038)***	(0.002)***
Observations	843,102	843,102	843,102	839,652	839,652	839,652	1,682,754
R-squared	0.011	0.053	0.123	0.010	0.052	0.154	0.146

Note: Author's estimates using Census data sourced from Ruggles et al. (2004). The dependent variable, *Log(Earnings)*, is the log hourly wage salary and business or farm income. Self employed individuals with non-positive earnings are excluded from the sample. *Experience* = *Age* – *Educ* – 6. The omitted state is Pennsylvania. Cohorts born between 1956 and 1959 are included in the sample. Robust standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.11: Non-Amish Female Cohort Differences in Fertility - Census 1990

	(1)	(2)	(3)	(4)	(5)	(6)
Post (=1)	-0.13	-0.13	-0.12	0.03	0.001	0.01
	(0.003)***	(0.003)***	(0.003)***	(0.006)***	(0.004)	(0.003)**
Metropolitan (=1)		-0.23	-0.16	-0.16	-0.16	-0.17
		(0.003)***	(0.003)***	(0.003)***	(0.002)***	(0.002)***
Married (=1)			0.70	0.70	0.68	0.66
			(0.003)***	(0.003)***	(0.002)***	(0.002)***
Age (scaled)				0.07	0.05	0.05
				(0.003)***	(0.002)***	(0.001)***
Post x Age					0.01	0.04
					(0.003)***	(0.001)***
Indiana (=1)	0.21	0.17	0.16	0.16	0.16	0.14
	(0.004)***	(0.004)***	(0.004)***	(0.003)***	(0.003)***	(0.002)***
Ohio (=1)	0.15	0.14	0.13	0.13	0.12	0.11
	(0.003)***	(0.003)***	(0.003)***	(0.003)***	(0.002)***	(0.002)***
Constant	1.60	1.78	1.25	1.14	1.19	1.23
	(0.002)***	(0.003)***	(0.004)***	(0.005)***	(0.005)***	(0.003)***
Observations	950,976	950,976	950,976	950,976	1,423,566	2,315,262
R-squared	0.01	0.01	0.08	0.08	0.08	0.09

Note: Author's estimates using Census data sourced from Ruggles et al. (2004). The dependent variable, *Chborn*, is the number of children ever born to a woman. The omitted state is Pennsylvania. Cohorts born between 1956 and 1959 are included in the sample. The control group is non-Amish white individuals who were native born. Robust standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.12: Dif-in-Diff Estimates of Exemption on Earnings and Fertility

	(1)	(2)	(3)	(4)	(5)	(6)
	------------------------- Log Hourly Earnings -------------------------					Fertility
	---- Census 1990 ----		---- Census 2000 ----		Pooled	Census 1990
Amish x Post (=1)	-0.17	-0.20	-0.34	-0.34	-0.27	0.28
	(0.05)***	(0.05)***	(0.04)***	(0.05)***	(0.03)***	(0.10)***
Post (=1)	-0.05	-0.04	-0.01	-0.01	-0.02	-0.12
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Amish (=1)	-0.12	-0.07	-0.13	0.08	0.01	-0.58
	(0.05)**	(0.07)	(0.04)***	(0.07)	(0.05)	(0.12)***
Metropolitan (=1)	0.14	0.16	0.17	0.18	0.17	-0.16
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Amish x Metro	0.09	0.06	-0.19	-0.19	-0.04	-0.14
	(0.05)*	(0.05)	(0.04)***	(0.04)***	(0.03)	(0.11)
Married (=1)		0.28		0.32	0.30	0.70
		(0.00)***		(0.00)***	(0.00)***	(0.00)***
Amish x Married		-0.07		-0.29	-0.13	2.75
		(0.06)		(0.05)***	(0.04)***	(0.09)***
Indiana (=1)	-0.03	-0.03	0.01	0.01	-0.01	0.16
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Amish x Indiana	0.29	0.27	0.38	0.37	0.30	0.14
	(0.07)***	(0.07)***	(0.04)***	(0.04)***	(0.05)***	(0.12)
Ohio (=1)	-0.02	-0.03	-0.00	0.00	-0.01	0.13
	(0.00)***	(0.00)***	(0.00)**	(0.00)	(0.00)***	(0.00)***
Amish x Ohio	-0.04	-0.06	0.44	0.43	0.18	-0.38
	(0.05)	(0.06)	(0.05)***	(0.05)***	(0.04)***	(0.16)**
Yr. 2000 (=1)					0.43	
					(0.00)***	
Amish x Yr. 2000					-0.10	
					(0.03)***	
Observations	844,752	844,752	840,824	840,824	1,685,576	952,736
R-squared	0.01	0.05	0.01	0.05	0.15	0.08

Note: Author's estimates using Census data sourced from Ruggles et al. (2004). The dependent variable is *Log(Earnings)* for estimating the returns to education and *Chborn* for estimating the effect of education on fertility, respectively. The omitted state is Pennsylvania. All specifications include an intercept term. Cohorts born between 1956 and 1959 are included in the sample. Self employed individuals with non-positive earnings are excluded from the earnings sample. The control group is non-Amish white individuals who were native born. Robust standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.13: The Implied Returns to Education and Effect of Education on Fertility

	(1)	(2)	(3)	(4)	(5)	(6)
	----------------------- Log Hourly Earnings -----------------------					Fertility
	---- Census 1990 ----		---- Census 2000 ----		Pooled	Census 1990
Educ (years)	0.18	0.23	0.52	0.32	0.30	-0.91
	(0.05)***	(0.06)***	(0.14)***	(0.06)***	(0.05)***	(0.43)**
Post (=1)	-0.03	-0.02	0.01	0.00	-0.01	-0.17
	(0.01)***	(0.01)***	(0.00)	(0.00)	(0.00)**	(0.02)***
Amish (=1)	0.48	0.35	1.74	0.28	0.43	-3.71
	(0.19)**	(0.13)***	(0.56)***	(0.13)**	(0.10)***	(1.55)**
Metropolitan (=1)	0.01	-0.01	-0.21	-0.06	-0.05	0.40
	(0.04)	(0.04)	(0.10)**	(0.05)	(0.03)	(0.26)
Amish x Metro	0.20	0.18	0.29	0.16	0.17	0.52
	(0.07)***	(0.07)**	(0.14)**	(0.08)*	(0.05)***	(0.35)
Married (=1)		0.20		0.12	0.16	1.03
		(0.02)***		(0.04)***	(0.02)***	(0.15)***
Amish x Married		0.38		0.73	0.60	1.30
		(0.15)***		(0.20)***	(0.12)***	(0.73)*
State x Amish F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Census x Amish F.E.	-	-	-	-	Yes	-
First-Stage F Stat						
Amish observations	1650	1650	1172	1172	2822	1760
Observations	844,752	844,752	840,824	840,824	1,685,576	952,736
R-squared	-0.01	-0.13	-1.50	-0.28	-0.16	-1.56

Note: Author's estimates using Census data sourced from Ruggles et al. (2004). The dependent variable is *Log(Earnings)* for estimating the returns to education and *Chborn* for estimating the effect of education on fertility, respectively. The omitted state is Pennsylvania. All specifications include an intercept term. Cohorts born between 1956 and 1959 are included in the sample. The control group is non-Amish white individuals who were native born. The instrumental variable for *Educ* is (*Amish* x *Post*), implying that the effect of the exemption on log hourly earnings or fertility is channeled through education. Robust standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.14: OLS Returns to Education and Effect of Education on Fertility

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	------------------------ Log Hourly Earnings ------------------------						----- Fertility -----	
	-- Census 1990 --		-- Census 2000 --		-- Pooled Census --		-- Census 1990 --	
Educ.	0.02	0.04	0.05	0.05	0.04	0.04	-0.15	-0.24
	(0.003)***	(0.003)***	(0.005)***	(0.005)***	(0.003)***	(0.003)***	(0.01)***	(0.01)***
Metro.	-0.002	0.02	0.19	0.19	0.09	0.10	-0.13	-0.15
	(0.02)	(0.02)	(0.02)***	(0.02)***	(0.01)***	(0.01)***	(0.05)***	(0.04)***
Married	0.43	0.19	0.22	0.20	0.38	0.20	3.48	2.41
	(0.02)***	(0.02)***	(0.03)***	(0.03)***	(0.02)***	(0.02)***	(0.04)***	(0.04)***
St. FE.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Yr. FE.	-	-	-	-	Yes	Yes	-	-
Exp	-	Yes	-	Yes	-	Yes	-	-
Exp. sq.	-	Yes	-	Yes	-	Yes	-	-
Age	-	-	-	-	-	-	-	Yes
Age sq.	-	-	-	-	-	-	-	Yes
Obs.	11376	11376	8046	8046	19422	19422	12495	12495
R-sq.	0.06	0.09	0.07	0.08	0.10	0.12	0.30	0.49

Note: Author's estimates using Census data sourced from Ruggles et al. (2004). The dependent variable is *Log(Earnings)* for estimating the returns to education and *Chborn* for estimating the effect of education on fertility, respectively. The omitted state is Pennsylvania. Amish aged 20 to 50 are included in the sample. Robust standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

CHAPTER 2 - REFERENCES

Abramitzky, Ran (2008) "The Limits of Equality: Insights from the Israeli Kibbutz," *Quarterly Journal of Economics*, vol. 123 (3), pp. 1111-1159.

Abramitzky, Ran (2009) "The Effects of Redistribution on Migration: Evidence from the Israeli Kibbutz," *Journal of Public Economics*, vol. 93, pp. 498-511.

Angrist, Joshua D. and Alan B. Krueger (1991) "Does Compulsory Schooling Attendance Affect Schooling and Earnings," *Quarterly Journal of Economics*, vol. 106(40), pp. 979-1014.

Berman, Eli (2000) "Sect, Subsidy, and Sacrifice: An Economist's View of Ultra-Orthodox Jews," *Quarterly Journal of Economics*, vol.115 (3), pp. 905-953.

Berman, Eli and Laurence R. Iannaccone (2006) "Religious Extremists: The Good, the Bad, and the Deadly," *Public Choice*, vol.128 (1-2), pp. 109-129.

Berman, Eli, Laurence R. Iannaccone, and Giuseppe Ragusa (2007) "From Empty Pews to Empty Cradles: Fertility Decline Among European Catholics," UCSD Mimeo.

Berman, Eli and David Laitin (2008) "Religion, Terrorism and Public Goods: Testing the Club Model," *Journal of Public Economics*, vol.92 (10-11), pp.1942-1967.

Berman, Eli and Ara Stepanyan (2005) "How Many Radical Islamists? Indirect Evidence from Five Countries," UCSD Mimeo.

Card, David (1999) "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics*, vol.3 Part A, Orley Ashenfelter and David Card eds., North Holland, Amsterdam.

Edmonston, Barry (2002) "The Undercount in the 2000 Census," The Annie E. Casey Foundation and the Population Reference Bureau, Baltimore, Maryland.

Edmonston, Barry and Charles Schultze (1995) *Modernizing the U.S. Census: Panel on Census Requirements in the Year 2000 and Beyond*, National Academy Press, Washington, D.C.

Ericksen, Julia A., Eugene P. Ericksen, John A. Hostetler, and Gertrude E. Huntington (1979) "Fertility Patterns and Trends among the Old Order Amish," *Population Studies*, vol.33 (2), pp. 255-276.

Greksa, Lawrence P. (2002) "Population Growth and Fertility Patterns in an Old Order Amish Settlement," *Annals of Human Biology*, vol.29 (2), pp. 192-201.

Greksa, Lawrence P. and Jill E. Korbin (2002) “Key Decisions in the Lives of the Old Order Amish: Joining the Church and Migrating to Another Settlement,” *Mennonite Quarterly Review*, vol.76 (4), pp. 373-378.

Hostetler, John A. (1993) *Amish Society*. The Johns Hopkins University Press, Baltimore.

Hostetler, John A. and Gertrude E. Huntington (1971) *Children in Amish Society: Socialization and Community Education*, Holt, Rinehart and Winston, New York.

Huntington, Gertrude E. (1994) “Persistence and Change in Amish Education,” in *The Amish Struggle with Modernity*, Donald B. Kraybill and Marc A. Olshan eds., University Press of New England, Hanover.

Iannaccone, Laurence R. (1992) “Sacrifice and Stigma: Reducing Free-Riding in Cults, Communes, and Other Collectives,” *Journal of Political Economy*, vol.100 (2), pp. 271-291.

Keim, Albert N. (1975) *Compulsory Education and the Amish: The Right Not to be Modern*. Beacon Press, Boston.

Kraybill, Donald B. (2001) *The Riddle of Amish Culture*. The Johns Hopkins University Press, Baltimore.

Kraybill, Donald B. and Carl F. Bowman (2001) *On the Backroad to Heaven: Old Order Hutterites, Mennonites, Amish, and Brethren*. The Johns Hopkins University Press, Baltimore.

Kraybill, Donald B. and C. Nelson Hostetter (2001) *Anabaptist World USA*. Herald Press, Scottdale.

Maluccio, John (1997) “Endogeneity of Schooling in the Wage Equation,” Unpublished Manuscript, Department of Economics, Yale University.

McConnell, David L. and Charles E. Hurst (2006) “No ‘Rip Van Winkles’ Here: Amish Education since Wisconsin v. Yoder,” *Anthropology and Education Quarterly*, vol.37 (3), pp. 236-254.

Meyers, Thomas J. and Steven M. Nolt (2005) *An Amish Patchwork*. Indiana University Press,

Meyers, Thomas J. (1993) “Education and Schooling.” in *The Amish and the State*, Donald B. Kraybill eds., The Johns Hopkins University Press, Baltimore.

National Population Commission (2000) *Nigeria Demographic and Health Survey 1999*, Calverton, Maryland: National Population Commission and ORC/Macro.

Osili, Una Okonkwo and Bridget Terry Long (2008) "Does female schooling reduce fertility? Evidence from Nigeria," *Journal of Development Economics*, vol.87 (1), pp. 57-75.

Park, Jin Heum (1994) "Estimation of Sheepskin Effects and Returns to Schooling Using the Old and the New CPS Measures of Educational Attainment," Working Paper No. 335, Industrial Relations Section, Princeton University.

Pollack, Randy Beth (1981) "Culture Change in an Amish Community," *Central Issues in Anthropology*, vol.3, pp. 51-67.

Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander (2004) *Integrated Public Use Microdata Series: Version 4.0* [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor].

Spence, Michael. (1973) "Job Market Signaling," *Quarterly Journal of Economics*, vol.87 (3), pp. 355-374.

CHAPTER 3

THE IMPACT OF IMMIGRATION ON NATIVE FERTILITY

Abstract

The interaction between immigration and native fertility has been overlooked in the literature. Previous research shows that immigration affects wages, income, and the cost of child rearing, while standard fertility model predicts that changes in wages, income, and the cost of child rearing would affect fertility. Using the cross-state variation in the total fertility rates of native-born American women and the share of immigrants in the population between 1970 and 2005, this paper estimates that for every one percentage point increase in the share of immigrants in the population, native total fertility rate is predicted to increase by roughly 0.01 children. The negative effect of immigration on wages is the most likely explanation, because the fertility of less educated women and women who resided in their states of birth is most affected.

3.1 Introduction

The large number of immigrants coming from Latin American countries has been offered as an explanation for the low old-age dependency ratio and the high fertility rates in the United States, as these immigrants increase working age population and new births (Caldwell and Schindlmayr 2003). Given the adverse effects that an aging and shrinking population may have on an economy, especially its impact on the solvency of the “pay as you go” social security system, some have advocated raising or maintaining the level of immigration as a partial solution. Since immigrant influx affects wages and prices, and fertility behavior is also responsive to wage and price changes, it is possible that immigration has an indirect effect on native fertility. We may miss an important feedback effect of higher immigration on native population growth if the interaction between immigrants and native fertility is not taken into consideration.

The argument that links native births to immigration can be dated back to Walker’s (1891) controversial hypothesis that immigrants led to the substitution of foreign born for native born and kept native population growth in place. Given Becker’s (1991) theory that ties rising wages and the costs of child rearing to falling fertility, immigrant influx may influence native fertility behavior through its effects on income and prices. However, the sign of the effect of immigration on native fertility can be ambiguous, because the empirical evidence of the effects of immigration on prices and wages is quite mixed and the effect of wages and prices on fertility is also difficult to sign.

It has been shown that the large increase in low-skilled immigrants in the labor force over the past few decades has lowered the wages earned by low-skilled native workers (Altonji and Card 1991; Borjas 1987, 2003, 2004; Borjas, Freeman, and Katz 1996, 1997; Card 1990, 2001, 2006). A fall in wages may increase fertility as it lowers the time cost of child rearing, but it may also decrease fertility as household income shrinks. Yet, there is also evidence that the increase in immigrants leads to an increase the average wages of native-born workers (Ottaviano and Peri 2006, 2008). Similarly, low-skilled immigration leads to lower prices of household services (Cortés 2008) and lower wages that private household workers earn (Khananusapkul 2004), but the overall immigrant influx also increases housing and rental prices across U.S. cities (Saiz 2003, 2006). Lower prices of goods and services may lower the cost of child rearing, but the higher cost of housing and rental prices mean that raising children becomes more costly. It is also possible that lower prices of household services induce women to increase labor supply, rather than fertility. Thus, the direction of the effect of immigration on native fertility is not as clear as Walker's (1891) claim and it needs to be assessed empirically.

A number of studies used simple cross-city, cross-state, or time series analyses to examine the relationship between immigration and native births in the late 19th century and early 20th century United States (Walker 1896; Goldenweiser 1912; Rollins 1930; Yasuba 1962; and Shergold 1974). Because the waves of immigration in the past 40 to 50 years, the roles of women, as well as the U.S. economy differed considerably to those in the late 19th century and early 20th century, the older findings are unlikely applicable to the modern U.S. economy. More importantly, because immigrants consider the economics and living conditions of their destinations and may time their immigration decision accordingly, findings based on simple cross-sectional or time series regressions are likely suffer from endogeneity bias and selection bias.

Using data from the U.S. Censuses and American Community Survey (ACS), this paper exploits the cross-state variations in the total fertility rates of native-born American women and the share of immigrants in the population between 1970 and 2005 to investigate the impact of immigration on native fertility. First, I show in Ordinary Least Squares (OLS) regressions that places with higher share of immigrants in the population also tend to have lower native total fertility rate. This is consistent with Walker's (1896) observation in the late 19th century. However, using the fixed effects (FE) model (with state and time fixed effects), I show that greater share of immigrants in the population predicts higher native fertility rates. The different results suggest that timing and location of immigration decision can bias simple cross-sectional or time series estimates.

Since immigrants may be attracted to places with higher level of wage rates or lower growth in the costs of child rearing, the negative or positive relationship may reflect immigrants' location choice. To address this form of bias, I exploit the social network of immigrants to implement an instrumental variable (IV) estimation strategy introduced by Card (2001). Specifically, the instrumental variable for the share of immigrants in the population is constructed by assigning the total number of immigrants from different countries in the U.S. based on their historical distribution across different states in 1960 as the weights to impute the share of immigrants in the population in each state of each period. This IV strategy allows us to identify the exogenous increase in immigrants in each state and hence their causal effect on native fertility.

In contrast to Walker's (1891) hypothesis that immigration lowers native-born fertility, the IV estimates show that the increased immigration during 1970 to 2005 actually increased fertility of native-born women. For every one percentage point increase in the share of immigrants in the population, native total fertility rate is predicted to increase by roughly 0.01 children. I also estimate the effects of immigration on native fertility for different groups of native-born women. The effect is particularly strong for high-school educated women and white women, less so for low-educated, but not for black women, Hispanic women, and women with some college degree. The results are fairly robust to restricting the samples of women who resided in their states of births and to potential outlier observations. Given that births to low and middle educated women are most responsive to increased immigration, and the strong effects of immigration on births to women who did not move from their states of births suggest that the negative effect of immigration on wages may most likely be the explanation. The positive effect of immigration on native fertility may provide an explanation for Camarota's (2005) and Frejka's (2004) observation that fertility rates remain high in the United States even after excluding births to immigrants and Hispanic Americans.

The rest of the paper is organized as follows. Section 3.2 presents a simple fertility model to highlight the mechanisms through which immigrant influx may affect native fertility. Section 3.3 describes the data. Section 3.4 presents the empirical methodology. Section 3.5 reports the results. Section 3.6 discusses the potential mechanisms explaining the observed effects. Section 3.7 concludes.

3.2 A Simple Model of Fertility Choice

In this section, I present a simple economic model to highlight how the influx of immigrants may affect native fertility and why the effects of immigration on native fertility can be difficult to sign. Consider a family maximizing a joint utility function for two adults and *f* children:

$$\max U\left(\frac{C}{2+f}, f\right)$$

where *C* is consumption, so that $C/(2+f)$ is consumption per family member, and *f* is the number of children.

The family is subject to a budget constraint where total time available, *T*, can be spent in either labor market (*H*), leisure (*L*), or raising children, λf :

$$T = H + L + \lambda f$$

Income earned through labor market activities can be spent on household consumption:

$$pC = wH = w(T - L - \lambda f),$$

where *p* is unit price of consumption. The family has a choice over *H* and *f*.

Immigrants can affect the budget constraint in a number of ways. Let *m* measure the share of immigrants in the population. First, according to the studies on the effects of immigration on prices, *m* will affect the prices of goods and services (*p*) and the time cost of child rearing (λ). For example, according to Cortés's (2008) and Khananusapkul's (2004) findings, low-skilled immigration keeps *p* and λ as low-skilled immigrants lower prices of food and household services. On the other hand, increased immigration may also increase *p* and λ as immigrant influx leads to a greater demand for all goods and services and pushes up prices of housing and rental housing. Thus, it is difficult to sign $\partial p/\partial m$ and $\partial \lambda/\partial m$.

Second, immigrants can also influence the household's budget constraint through their effect on wages. It is also unclear whether the effect of immigration (*m*) on wages (*w*) is positive or negative. For example, Borjas (1987, 2003, 2004) and Card (1990, 2001, 2006) show that low-skilled immigration dampens the wages of native-skilled native workers, but Ottaviano and Peri (2008) show that the effect of immigration on the average wages of native-workers is positive in the long run. Similarly, Lopez (2003) finds a positive effect of high-skilled immigration on high-skilled native workers, but Borjas (2005) reports a negative effect of high-skilled immigration on high-skilled native workers.

Solving for the optimal choice of hours and fertility yields a standard labor supply equation and a derived demand for children:

$$f\left(\frac{w}{p},\frac{\lambda}{p}\right)$$

Assuming that children are normal good, we would expect $\partial f/\partial(\lambda/p) > 0$, but $\partial f/\partial(w/p)$ has an ambiguous sign as it depends on the relative size of the income and substitution effects. Studies, such as those by De Tray (1973), Willis (1973), Schultz (1985), and Heckman and Walker (1990), have shown that the number of children is negatively related to the wage rate or other measures of the value of time of women, but positively related to male earnings.[53] Lindo (2010) shows that a negative income shock due to a husband's job displacement predicts an immediate rise in a woman's fertility, but an eventual reduction in her fertility. Hence, it is unclear whether higher wages necessarily imply lower fertility or higher fertility.

[53] See Butz and Ward (1979) and Macunovich (1995) for discussions on the counter-cyclicality of fertility.

Since the effects of immigration on prices and wages and the effect of wages on fertility are theoretically ambiguous to determine, it remains an empirical question whether higher immigration leads to an increase or a decrease in native fertility.

3.3 Data Description

This paper draws Census 1970, 1980, 1990, and 2000 data and ACS 2005 data from the Integrated Public Use Microdata Series (IPUMS) to study the relationship between immigration and native fertility across 50 states and the District of Columbia during the period 1970 to 2005. Following the definition commonly used in the literature, a native is referred to as a person born in the United States or abroad of American parents, while an immigrant is defined as a person who is a naturalized citizen or non-US citizen. Summary statistics of all variables used are reported in Table 3.1.

3.3.1 Fertility Data

The total fertility rate of native-born women aged 15 to 49 is used as the measure of native fertility in this study. The total fertility rate is best interpreted as the expected number of births to a woman during her child bearing years. It is a reasonable measure of fertility especially for fertility comparison because it is standardized for differences in age distributions. However, because the natality database does not report birthplaces of women giving births, I cannot link births to native-born women. Instead, I use children aged 0 residing with their mothers reported in the censuses and ACS as a proxy for births to compute total fertility rates. Infant mortality and living arrangement of mothers and infants mean that there is likely a measurement error in the total fertility rate.

Nevertheless, since the measurement error is in the dependent variable, as long as it is random, it will not bias the estimated effect of immigration on native fertility.

Table 3.2 shows the total fertility rates of native-born American women age 15-49 for the 50 states and the District of Columbia between 1970 and 2005. We can see that the total fertility rates vary across states and over time. Most states did not experience a steady decline or increase in total fertility rates during the sampled period. Utah has had the highest average total native fertility rates and the District of Columbia has had the lowest average total native fertility rates during the sampled period.

3.3.2 Immigration Data

The regressor of interest is the share of immigrants in the population. Table 3.3 reports the share of immigrants in the population by state and the District of Columbia between 1970 and 2005. California, New York, and Hawaii tend to have the largest share of immigrants in the population, while West Virginia and Mississippi have the smallest share of immigrants in the population. Table 3.4 shows the 10 most common birthplaces of foreign-born individuals between 1960 and 2005. I report the common birthplaces of foreign-born individuals in 1960 because the instrumental variable relies on the 1960 distribution of immigrants from different nationals across the U.S. In 1960 and 1970, immigrants tend to come from European countries, Canada, and Mexico. Since 1980, Mexico, Latin-American countries, and Asian countries have become the major immigrant source countries for the U.S.

3.4 Empirical Methodology

3.4.1 Basic Empirical Model

I estimate the effect of immigration on the TFR of the following groups of native-born women: (1) all women; (2) white women; (3) black women; (4) Hispanic women; (5) women with some college education (high educated); (6) women with a high school diploma (middle educated); and (7) women having less than 12 years of education (low educated). The empirical model takes the following form:

$$TFR_{jt} = \alpha_j + \delta_t + \beta(m_{jt}) + u_{jt},$$

where TFR_{jt} is the total fertility rates of native-born women residing in state *j* at time *t*; α_j denotes a set of state fixed effects; δ_t denotes a set of time effects; m_{jt} is the share of immigrants in the population in state *j* at time *t*; and u_{jt} is the error term. The time fixed effect will remove the aggregate shock to total fertility rates and immigration at time *t* and the state fixed effects will remove time invariant state-specific unobserved influences of fertility and immigration.

The FE model will provide a consistent estimate of the causal effect of immigration on fertility, β, if the share of immigrants in the population is not correlated with the state-specific time varying unobserved influences of native fertility. However, it is difficult to rule out the possibility that changes in local economic condition simultaneously influence immigrant influx and the fertility decision of native-born women. Therefore, we need an instrumental variable strategy to remove this form of endogeneity bias and to identify the causal relation between immigration and native fertility.

3.4.2 Instrumental Variable

Following Card and DiNardo's (2000) and Card's (2001) approach, I construct an instrumental variable for the share of immigrants in the population by exploiting the tendency of immigrants to reside in places with well established immigrant networks from the same region or country. Since the social network of immigrants in a city and the concentration of early immigrants from the same country greatly influence immigrants' location choice, this instrumental variable will have a strong predictive power (Bartel 1989, Munshi 2003).

The instrumental variable approach employed here allocates the total number of immigrants by nationalities according to their historical distribution across states to form an imputed number of immigrants by states. Specifically, I use the share of each foreign-born group of the total number of that foreign-born group in state *j* in 1960 as the weight to assign the total number of that immigrant group in year *t* in the whole United States and sum up the weighted number of immigrant groups.[54] The imputed number of immigrants, $\tilde{Z}_{jt}$, can be expressed as:

$$\tilde{Z}_{jt} = \sum_k \phi_{kj,1960} \cdot M_{kt},$$

where $\phi_{kj,1960}$ is the share of immigrants from country or region *k* included residing in state *j* in 1960;[55] and M_{kt} is the total number of immigrants from country or region *k* in time *t*. *k* includes 54 countries or regions of foreign birthplaces reported in the 1% public

---

[54] Census 1960 does not separately identify foreign-born citizens, native-born citizens, and citizens born abroad to American parents. Therefore, all foreign-born individuals are treated as immigrants in computing the historical distribution of immigrants.

[55] That is, $\phi_{kj,1960} = M_{kj,1960} / M_{k,1960}$

use micro-sample of Census 1960.[56] Then, the imputed number of immigrants in state *j* time *t* is divided by the number of persons in state *j* time *t* to obtain the instrumental variable for the share of immigrants in the population.

The instrument will predict the share of immigrants in the population if there is large number of immigrants from a country in 1960 to attract the location choice of future immigrants. Furthermore, for the instrument to satisfy the exogeneity condition, we need: (1) the distribution of immigrants in 1960 in each state to not directly affect the future native fertility in the state, and; (2) the national immigrant stocks to be exogenous to the local economic conditions of immigrant states.[57]

3.5 Empirical Results

3.5.1 OLS, FE, and IV Estimates: All Native-born Women

Table 3.5 presents the Ordinary Least Squares (OLS), fixed effects (FE), and instrumental variable (IV) estimates of the effect of immigration on native fertility. Column (1) and (2) report the pooled OLS estimates and demonstrate Walter's (1891) argument that higher immigration leads to lower native fertility. The inverse relationship may simply indicate that immigrants are more likely to locate into more urbanized places, where native fertility is also low.

Column (3) shows that the inverse relationship between immigration and native fertility remains strong in the FE model where time fixed effects are excluded. However,

[56] The countries or regions selected are based on the variable BPL in the IPUMS. The full list of country and region names is available at http://usa.ipums.org/usa-action/codes.do?mnemonic=BPL

[57] We can relax the second assumption by using the predicted total national inflows of immigrants from various sending economies on the basis of a regression of total national inflows of immigrants against a set of country-specific variables that affect emigration but are exogenous to changes in US city-specific amenities. Saiz (2006) shows that the results are similar whether or not one uses this alternative approach.

once a set of time fixed effects are included in the FE model to account for aggregate time shock to immigration and native fertility, higher immigration is predicted to increase native total fertility rates (Column 4). For every one percentage point increase in the share of immigrants in the population, the native total fertility rate is predicted to increase by 0.016 children. Given that the national average native total fertility rate is 1.72 children in 2005, the effect size is approximately 1%.

Column (5) presents the IV estimates. The first-stage partial F of 11.5 indicates that the instrumental variable is strong. The second-stage IV results show that for a one percentage point increase in the share of immigrants in the population, the native total fertility rate will increase by roughly 0.01 children. This effect size is approximately 0.5% of the national average total fertility rate in 2005. The IV estimate is smaller than the FE estimate reported in column (4), perhaps because immigrants are drawn to places in which the costs of living and child rearing are not growing fast.

3.5.2 FE and IV Estimates: White, Black, and Hispanic Women

Table 3.6 shows the estimated effects of immigration on the fertility of different native-born racial groups. The estimates indicate that greater immigration leads to higher fertility among white and black women, but not Hispanic women. According to the IV estimates, for every one percentage point increase in the share of immigrants in the population, white native fertility is expected to increase by 0.02 children and black native fertility is expected to increase by 0.03 children. These are equivalent to a 1.1% increase and a 1.6% increase from the mean total fertility rates, respectively. However, the large effect size for black women is not precisely estimated. Therefore, the evidence only suggests that immigration tends to increase births of native-born white women.

3.5.3 FE and IV Estimates: High, Middle, and Low Educated Women

Table 3.7 reports the estimated effect of immigration on the total fertility rate of native-born women by educational attainment. The IV estimates show that the fertility of low educated and middle educated women tend to be more responsive to increased immigration than the fertility of high educated women does. For every one percentage point increase in the share of immigrants in the population, the fertility of low educated women is expected to increase by 0.021 children, the fertility of middle educated women is expected to increase by 0.015 children, and the fertility of high educated women is expected to increase by 0.007 children. Nonetheless, only the effect of immigration on the fertility of middle educated women is precisely estimated.

3.6 Some Robustness Checks

3.6.1 Native Mobility

It is possible that native-born women are driven out to states with fewer immigrants to avoid labor market competition and costs of child rearing, leading to selection bias in the estimated effect of immigration on native fertility.

The empirical findings regarding the mobility response of natives to the inflow of immigrants are mixed. Earlier studies, such as those by Frey (1995, 1996) showed a positive relation between immigrant inflows and native outflows in metropolitan areas. Borjas, Freeman, and Katz (1996) show that the negative effects of immigration on native labor market outcomes increase in magnitude when the geographical area used as the unit of observation expands. Presumably, native out-migration tends to undo the labor supply shocks observed at the local labor market. On the other hand, Card (2001, 2006) and Card and DiNardo (2000) show that immigrants increase the labor force size proportionately, suggesting little labor-market native flight. Similarly, Saiz's (2003, 2006) findings on the rise in housing and rental prices with higher immigration also do not support for the native flight hypothesis.

The use of state-level data in this paper, following the argument made in Borjas et al. (1996), should avoid the bias in estimates associated with cross-city within-state native flight. To further test whether estimates are sensitive to potential cross-state native flight, I limit the analysis to native-born women who resided in their states of birth. Table 3.8 reports the IV estimates using total fertility rates of non-movers of various native-born groups as the dependent variable. Column (1) reports the IV estimate using the sample of all native-born non-movers. The estimate is more significant and positive than the sample that includes women that moved. IV estimates for white women, black women, Hispanic women, low educated women, middle educated women, and high educated women are reported in columns (2), (3), (4), (5), (6), and (7), respectively. All of the estimates are more positive than those presented in previous sections. This means native mobility tends to reduce the observed effects of immigration on native fertility. Thus, ignoring native mobility provides a lower bound on the positive effect of immigration on native fertility.

3.6.2 Excluding Potential Outliers

There may be a concern that the results presented in previous sections are driven by potential outliers. To assess whether estimates are senstive to potential outliers, Figures 3.1 to 3.7 plot the instrumental variable regression residuals of total fertility rate against the share of immigrants after conditioning on year and state fixed effects for various groups of native-born women. A number of states appear to produce potential outlier observations in various years. Table 3.9 lists state-year observations with TFR or the (predicted) share of immigrants after conditioning on a set of state and year fixed effects that are 2 standard deviations above their mean and are dropped as potential outliers. IV estimates using samples that exclude these potential outlier observations are reported in Table 3.10. It appears that the size and significance of the IV estimates for all native-born women, white women, middle educated women, and low educated women remain similar to those presented in previous sections. Thus, the results are fairly robust to potential outliers.

3.7 Discussion of the Effects and Channels

I estimated positive effects of increased immigration on fertility of native-born white women and women without any college education, but no significant effect of increased immigration on high educated women and Hispanic women. The estimated effect of immigration on native-born black women is positive but appears noisier as the size of the coefficient varies considerably depending on the sample used.

The lack of effect on high-educated women, who tend to benefit from lowered prices of household services, suggests that lower prices of household services are unlikely an explanation for the fertility response. This explanation is consistent with

Cortés and Tessada's (2009) finding that lower prices of household services due to low-skilled immigration allows highly educated women to increase their work hours. Similarly, the strong positive fertility response of less-educated women to increased immigration implies that higher prices of rental housing due to immigrant influx is likely not a reason for the observed positive relationship between immigration and native fertility. The positive effects of immigration on women without any college education and black women suggest that the effect of immigration on wages is more likely the mechanism underlying the relationship between immigration and fertility. Moreover, the wage effect as an explanation is also consistent with the strong positive effect of immigration on women who resided in their states of birth, because they absorb the labor market impact of immigrants. This explanation requires that the substitution effect of wages on fertility is greater than the income effect of wages on fertility. Increased immigration dampens the wages of these women, lowering their opportunity cost of child rearing and leading to increased fertility.

3.8 Conclusion

This paper investigates the causal link between immigration and native fertility. Specifically, I argue that increased immigration may influence native fertility through its effects on prices and wages that native-born women face. Because it is theoretically ambiguous whether increased immigration leads to higher or lower native fertility, the relationship needs to be examined empirically.

I employ Census and American Community Survey data between 1970 and 2005 to assess whether changes in immigration influence changes in native fertility. I find that immigration leads to higher overall native fertility and the positive effect is strong among women with low and middle educational attainment and white women. The estimates are stronger when restricting to the sample of non-moving individuals. The estimated effect of immigration on black women is also positive and large, but it is noisier and somewhat sensitive to the sample used. The estimated effect of immigration on highly educated women and Hispanic women is statistically insignificant and small in magnitude.

Some possible explanations for the effect of immigration on native fertility are discussed. The most likely channel through which immigration affects native fertility is through wages, given that the fertility of less educated women and women who resided in their states of birth is most responsive to increased immigration.

CHAPTER 3 – FIGURES

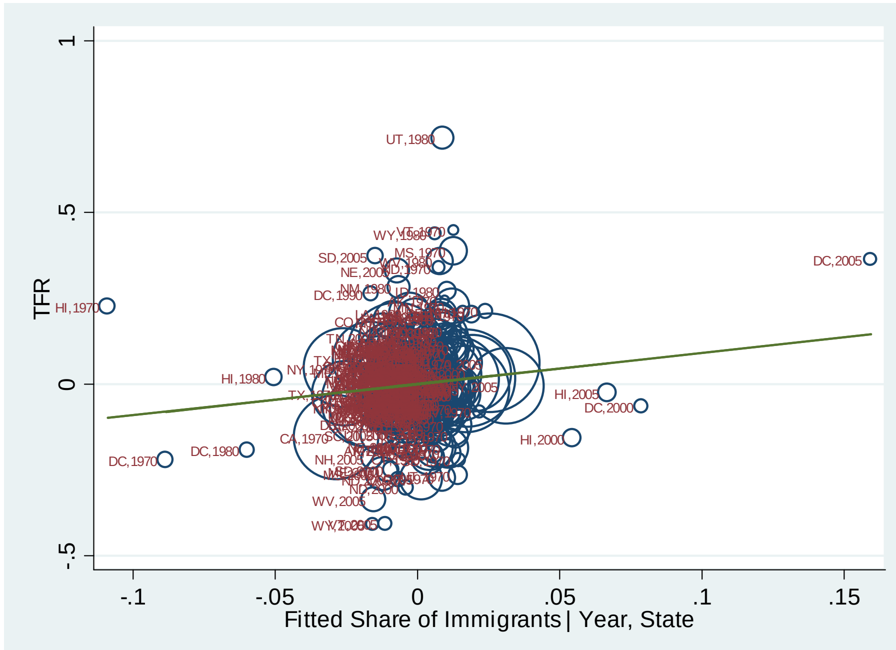

Figure 3.1: TFR of All Native- Born Women vs. Share of Immigrants

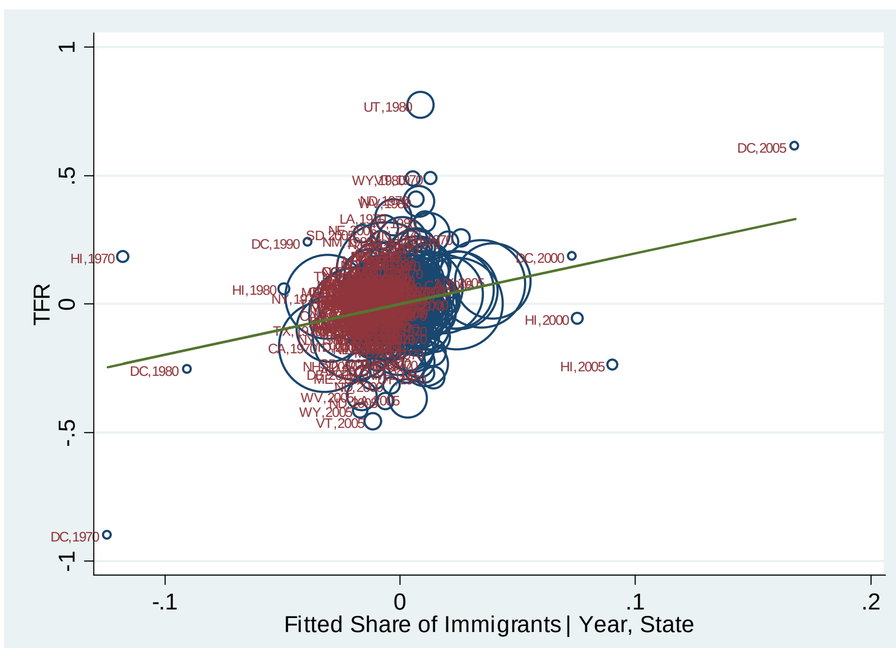

Figure 3.2: TFR of Native-Born White Women vs. Share of Immigrants

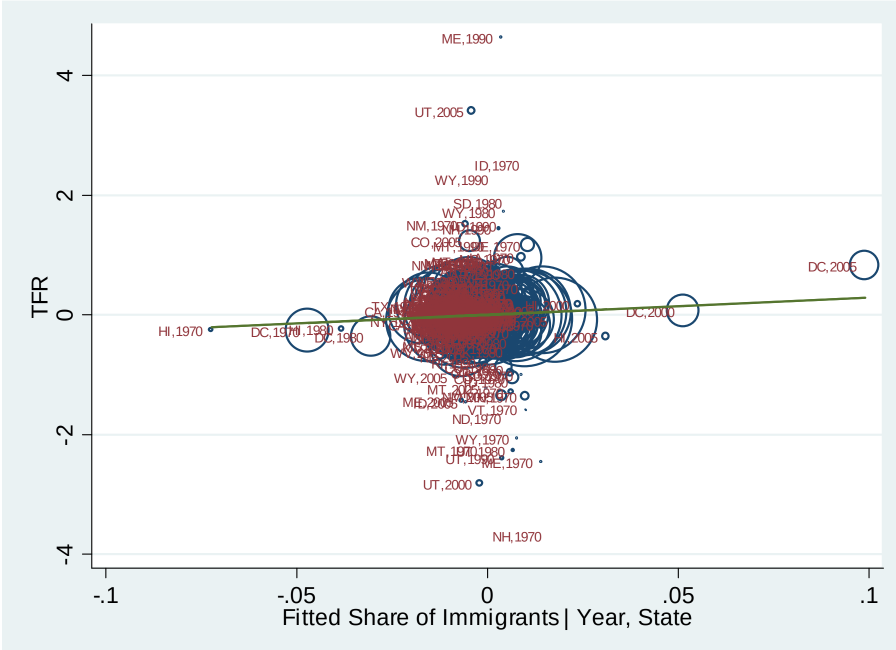

Figure 3.3: TFR of Native- Born Black Women vs. Share of Immigrants

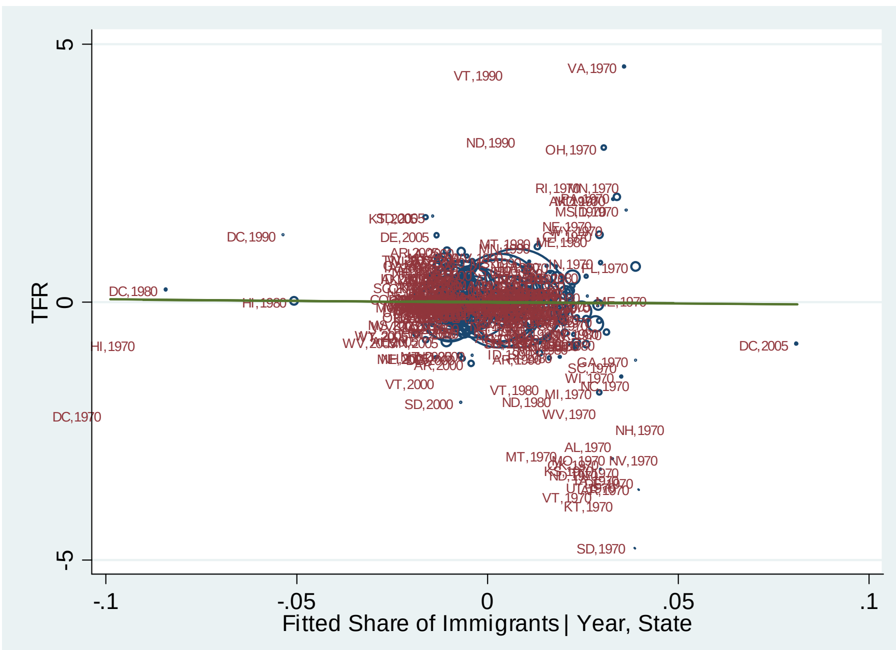

Figure 3.4: TFR of Native- Born Hispanic Women vs. Share of Immigrants

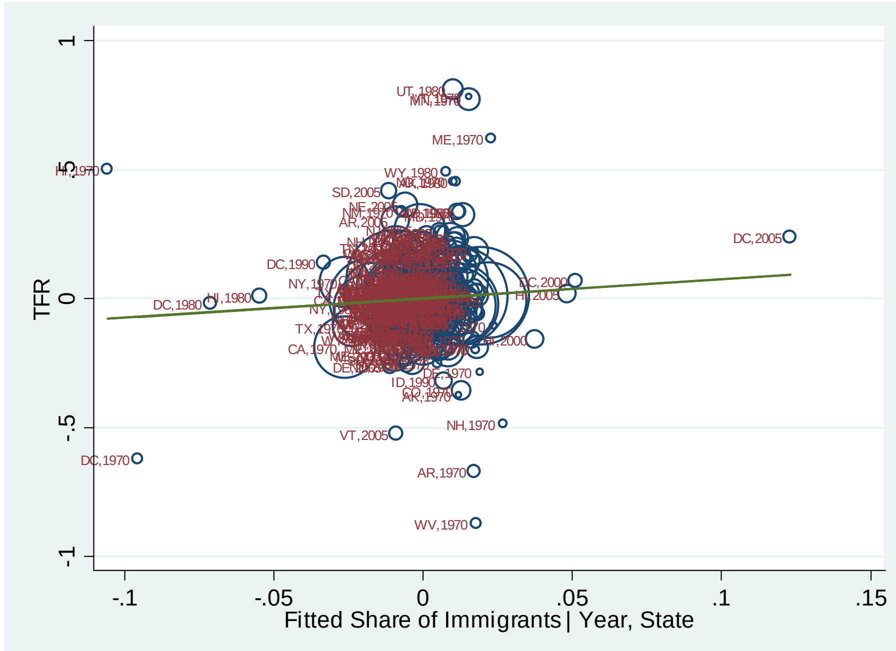

Figure 3.5: TFR of Native- Born High Educated Women vs. Share of Immigrants

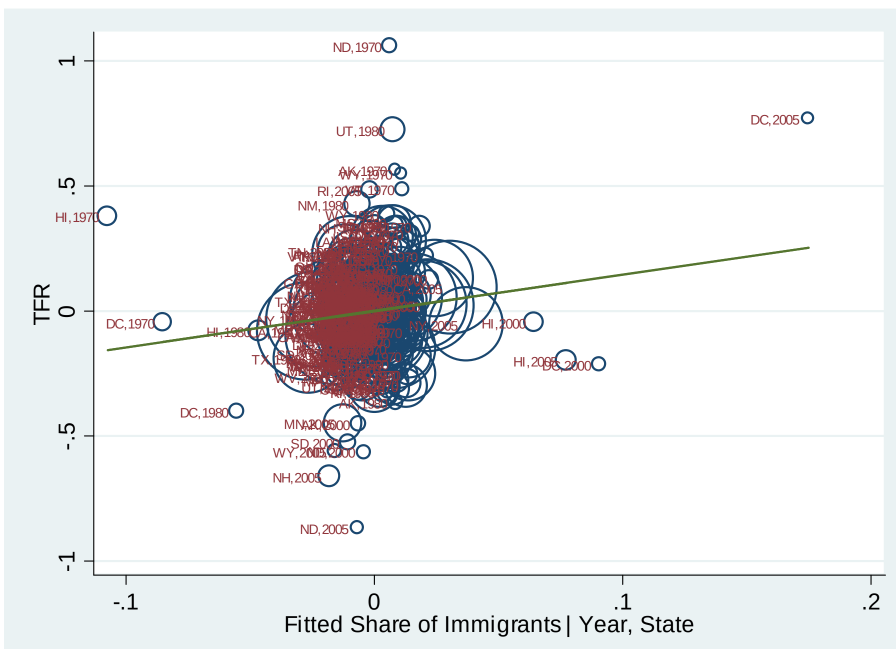

Figure 3.6: TFR of Native- Born Middle Educated Women vs. Share of Immigrants

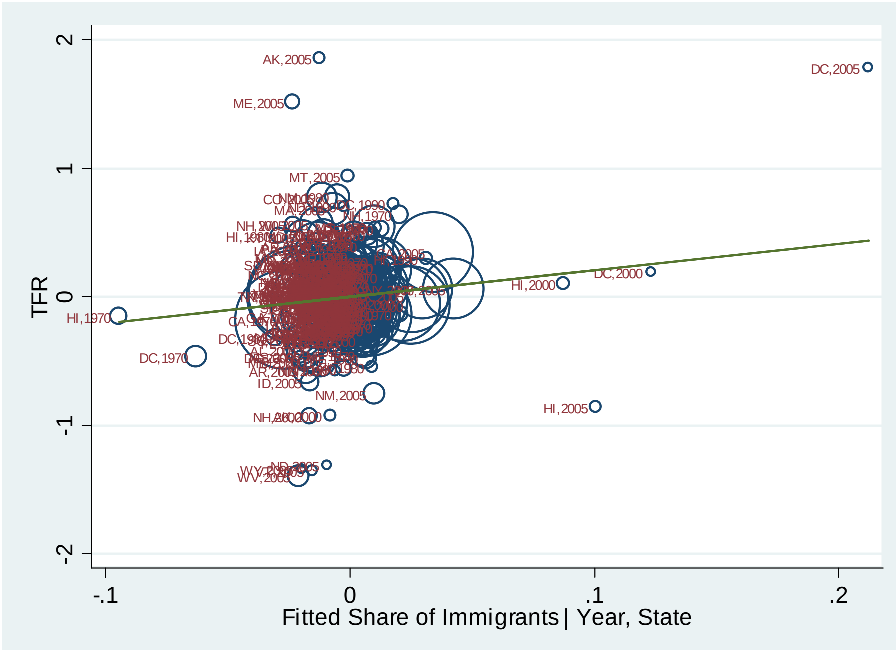

Figure 3.7: TFR of Native- Born Low Educated Women vs. Share of Immigrants

CHAPTER 3 – TABLES

Table 3.1: Descriptive Statistics

Variable	*Obs.*	*Weight*	*Mean*	*Std. Dev.*	*Min*	*Max*
All Native-Born						
Total Fertility Rate	255	277455052	1.776614	0.31067	1.084404	3.3049
Immigrant share	255	277455052	0.081976	0.068501	0.004933	0.272994
Instrumental variable	255	277455052	0.076629	0.074714	0.003129	0.679427
Female pop. (15-49)	255	277455052	2186932	1521202	67800	5790964
White						
Total Fertility Rate	255	217664525	1.759688	0.311016	0.592904	3.337876
Immigrant share	255	217664525	0.076731	0.065082	0.004933	0.272994
Instrumental variable	255	217664525	0.071109	0.067728	0.003129	0.679427
Female pop. (15-49)	255	217664525	1593634	996145.2	36400	4111684
Black						
Total Fertility Rate	254	37294425	1.775549	0.431465	0	6.934704
Immigrant share	254	37294425	0.078557	0.06629	0.004933	0.272994
Instrumental variable	254	37294425	0.070937	0.079027	0.003129	0.679427
Female pop. (15-49)	254	37294425	355979.9	169426.7	160	719876
Hispanic						
Total Fertility Rate	253	17431151	1.94587	0.505191	0	7.666667
Immigrant share	253	17431151	0.142467	0.076261	0.004933	0.272994
Instrumental variable	253	17431151	0.137626	0.083074	0.003129	0.679427
Female pop. (15-49)	253	17431151	593064.1	503309.9	100	1514288
High Educated						
Total Fertility Rate	255	125128245	1.641346	0.281477	0.746181	3.308182
Immigrant share	255	125128245	0.092781	0.072034	0.004933	0.272994
Instrumental variable	255	125128245	0.08608	0.080413	0.003129	0.679427
Female pop. (15-49)	255	125128245	1158122	926182.1	17800	3470048
Middle Educated						
Total Fertility Rate	255	89053016	1.943483	0.429278	0.634104	4.151695
Immigrant share	255	89053016	0.074976	0.064515	0.004933	0.272994
Instrumental variable	255	89053016	0.070586	0.069728	0.003129	0.679427
Female pop. (15-49)	255	89053016	678753.4	429254.4	26052	1532938
Low Educated						
Total Fertility Rate	255	63273791	1.9964	0.407334	0.281643	4.590182
Immigrant share	255	63273791	0.070461	0.063421	0.004933	0.272994
Instrumental variable	255	63273791	0.066446	0.067188	0.003129	0.679427
Female pop. (15-49)	255	63273791	525637.8	370183.2	16024	1426200

Note: Data sourced from Census 1960, 1970, 1980, 1990, and 2000, and American Community Survey 2005 (Ruggles et al. 2004). Total fertility rates (women aged 15-49) are calculated using infants residing with mothers (not living in group quarters) as the proxy of births. Immigrant share is the share of immigrants in the population and the unweighted value is the same for all groups of women. Immigrants are individuals who are either naturalized citizens or non-US citizens. Persons living in institutions are excluded. See construction of the instrumental variable in text and note that its construction requires Census 1960 data. All statistics are weighted by the female population of the respective group.

Table 3.2: Total Fertility Rates of Native- Born Women by State and Year

State	1970	1980	1990	2000	2005	State	1970	1980	1990	2000	2005
Alabama	2.47	1.81	1.47	1.75	1.65	Montana	2.34	2.03	1.69	1.79	2.05
Alaska	2.98	2.19	2.10	1.90	1.95	Nebraska	2.62	2.16	1.74	1.87	2.36
Arizona	2.52	2.04	1.71	1.78	1.80	Nevada	2.23	1.70	1.55	1.78	1.55
Arkansas	2.29	2.04	1.58	1.79	1.74	N. Hampsh.	2.51	1.68	1.50	1.75	1.41
California	2.06	1.58	1.47	1.59	1.61	New Jersey	2.29	1.59	1.40	1.68	1.74
Colorado	2.20	1.73	1.51	1.79	1.92	New Mexico	2.59	2.20	1.64	1.72	1.74
Connecticut	2.51	1.52	1.42	1.77	1.68	New York	2.26	1.56	1.39	1.60	1.55
Delaware	2.53	1.68	1.58	1.62	1.54	Nth Carolina	2.25	1.62	1.36	1.72	1.61
D.C	1.64	1.08	1.29	1.16	1.56	Nth Dakota	2.93	2.15	1.91	1.64	1.64
Florida	2.26	1.55	1.38	1.56	1.53	Ohio	2.37	1.82	1.59	1.77	1.72
Georgia	2.35	1.78	1.52	1.74	1.77	Oklahoma	2.45	1.94	1.62	1.94	1.86
Hawaii	2.62	1.83	1.57	1.59	1.70	Oregon	2.07	1.79	1.60	1.65	1.79
Idaho	2.66	2.41	1.69	2.19	1.97	Pennsylvania	2.32	1.67	1.45	1.60	1.71
Illinois	2.34	1.79	1.50	1.65	1.78	Rhode Island	2.22	1.54	1.38	1.57	1.65
Indiana	2.41	1.88	1.60	1.85	1.71	Sth Carolina	2.26	1.81	1.45	1.68	1.46
Iowa	2.63	2.05	1.65	1.82	1.84	Sth Dakota	2.62	2.41	1.90	1.95	2.55
Kansas	2.34	2.00	1.71	1.98	1.74	Tennessee	2.36	1.67	1.45	1.73	1.83
Kentucky	2.35	1.92	1.43	1.74	1.62	Texas	2.42	1.94	1.59	1.82	1.77
Louisiana	2.64	2.07	1.56	1.79	1.50	Utah	2.91	3.30	2.21	2.39	2.42
Maine	2.51	1.72	1.55	1.41	1.67	Vermont	2.73	1.69	1.55	1.68	1.21
Maryland	2.35	1.48	1.45	1.57	1.69	Virginia	2.11	1.56	1.46	1.69	1.56
Massachu.	2.29	1.39	1.36	1.57	1.49	Washington	2.27	1.83	1.57	1.68	1.68
Michigan	2.54	1.76	1.59	1.76	1.66	W. Virginia	2.20	2.00	1.32	1.64	1.23
Minnesota	2.66	1.89	1.55	1.72	1.68	Wisconsin	2.57	1.89	1.58	1.74	1.78
Mississippi	2.90	2.13	1.59	1.77	1.58	Wyoming	2.81	2.51	1.77	1.94	1.58
Missouri	2.22	1.89	1.62	1.81	1.75	State Ave.	2.43	1.87	1.57	1.74	1.72

Note: Data sourced from Census 1970, 1980, 1990, and 2000, and American Community Survey 2005 (Ruggles et al. 2004). Total fertility rates (age 15-49) are calculated using infants residing with mothers (not living in group quarters) as the proxy of births. Persons living in institutions are excluded.

Table 3.3: Share of Immigrants in the Population by State and Year

State	1970	1980	1990	2000	2005	State	1970	1980	1990	2000	2005
Alabama	0.01	0.01	0.01	0.02	0.03	Montana	0.03	0.02	0.02	0.02	0.02
Alaska	0.03	0.04	0.04	0.06	0.05	Nebraska	0.02	0.02	0.02	0.04	0.06
Arizona	0.05	0.06	0.08	0.13	0.15	Nevada	0.04	0.07	0.09	0.16	0.17
Arkansas	0.01	0.01	0.01	0.03	0.04	N. Hampsh.	0.05	0.04	0.04	0.04	0.06
California	0.09	0.15	0.22	0.26	0.27	New Jersey	0.09	0.10	0.13	0.18	0.19
Colorado	0.03	0.04	0.04	0.09	0.10	New Mexico	0.02	0.04	0.05	0.08	0.09
Connecticut	0.08	0.09	0.08	0.11	0.12	New York	0.12	0.14	0.16	0.21	0.21
Delaware	0.03	0.03	0.03	0.06	0.07	Nth Carolina	0.01	0.01	0.02	0.05	0.07
DC	0.05	0.06	0.10	0.13	0.13	Nth Dakota	0.03	0.02	0.01	0.02	0.02
Florida	0.08	0.11	0.13	0.17	0.19	Ohio	0.03	0.03	0.02	0.03	0.04
Georgia	0.01	0.02	0.03	0.07	0.09	Oklahoma	0.01	0.02	0.02	0.04	0.04
Hawaii	0.10	0.14	0.15	0.18	0.17	Oregon	0.03	0.04	0.05	0.09	0.10
Idaho	0.02	0.02	0.03	0.05	0.06	Pennsylvania	0.04	0.03	0.03	0.04	0.05
Illinois	0.05	0.07	0.08	0.12	0.14	Rhode Island	0.08	0.09	0.09	0.11	0.12
Indiana	0.02	0.02	0.02	0.03	0.04	Sth Carolina	0.01	0.02	0.01	0.03	0.04
Iowa	0.01	0.02	0.02	0.03	0.03	Sth Dakota	0.01	0.01	0.01	0.02	0.03
Kansas	0.01	0.02	0.03	0.05	0.05	Tennessee	0.01	0.01	0.01	0.03	0.04
Kentucky	0.01	0.01	0.01	0.02	0.03	Texas	0.03	0.06	0.09	0.14	0.16
Louisiana	0.01	0.02	0.02	0.03	0.03	Utah	0.03	0.04	0.03	0.07	0.08
Maine	0.05	0.04	0.03	0.03	0.03	Vermont	0.05	0.04	0.03	0.04	0.03
Maryland	0.03	0.05	0.07	0.10	0.12	Virginia	0.02	0.03	0.05	0.08	0.10
Massachu.	0.09	0.09	0.10	0.12	0.15	Washington	0.05	0.06	0.07	0.11	0.12
Michigan	0.05	0.04	0.04	0.05	0.06	W. Virginia	0.01	0.01	0.01	0.01	0.01
Minnesota	0.03	0.02	0.03	0.05	0.07	Wisconsin	0.03	0.03	0.02	0.04	0.04
Mississippi	0.00	0.01	0.01	0.01	0.02	Wyoming	0.03	0.02	0.02	0.02	0.02
Missouri	0.01	0.02	0.02	0.03	0.04	State Ave.	0.04	0.04	0.05	0.07	0.08

Note: Data sourced from Census 1970, 1980, 1990, and 2000, and American Community Survey 2005 (Ruggles et al. 2004). Author's own calculation of the share of immigrants in the population. Immigrants are individuals who are either naturalized citizens or non-US citizens. Persons living in institutions are excluded.

Table 3.4: Distribution of the 10 Most Common Foreign Birthplaces

	-------- Census 1960 --------		-------- Census 1970 --------		-------- Census 1980 --------	
Rank	Birthplaces	Share	Birthplaces	Share	Birthplaces	Share
1	Italy	0.125	Italy	0.099	Mexico	0.157
2	Germany	0.099	Canada	0.094	Canada	0.060
3	Canada	0.096	Germany	0.091	Germany	0.060
4	Poland	0.072	Mexico	0.082	Italy	0.059
5	Russia	0.067	Poland	0.054	West Indies	0.047
6	Mexico	0.058	England	0.051	Cuba	0.044
7	England	0.050	Cuba	0.046	South America	0.041
8	Ireland	0.040	Russia	0.044	Philippines	0.037
9	Austria	0.029	Ireland	0.028	China	0.032
10	Hungary	0.024	South America	0.027	England	0.031
	-------- Census 1990 --------		-------- Census 2000 --------		-------- ACS 2005 --------	
Rank	Birthplaces	Share	Birthplaces	Share	Birthplaces	Share
1	Mexico	0.218	Mexico	0.295	Mexico	0.307
2	West Indies	0.061	West Indies	0.067	Central America	0.070
3	Central America	0.057	Central America	0.065	South America	0.068
4	South America	0.053	South America	0.062	West Indies	0.063
5	China	0.048	China	0.049	India	0.052
6	Philippines	0.047	India	0.045	China	0.049
7	Cuba	0.037	Philippines	0.044	Philippines	0.045
8	Canada	0.037	Vietnam	0.032	Africa	0.034
9	Germany	0.036	Korea	0.028	Vietnam	0.030
10	India	0.031	Cuba	0.028	Korea	0.028

Note: Author's own calculation based on Census and ACS data sourced from Ruggles et al. (2004). Data from Census 1960 includes individuals born abroad of American parents as immigrants. Data from all other years include non-citizen individuals and naturalized citizens as immigrants.

Table 3.5: OLS, FE, and IV Estimates – All Native- Born Women

	(1) OLS	(2) OLS	(3) FE	(4) FE	(5) IV
Immigrant Share	-1.54	-0.88	-2.90	1.56	0.91
	(0.18)***	(0.17)***	(0.45)***	(0.31)***	(0.49)*
Year fixed effects	No	Yes	No	Yes	Yes
State fixed effects	No	No	Yes	Yes	Yes
Observations	255	255	255	255	255
R-squared	0.12	0.75	0.28	0.92	0.90
First-stage F stat					11.52***

Note: Author's own estimates based on data sourced from Ruggles et al. (2004). Observations are weighted by the size of female population aged 15 to 49. Robust standard errors clustered by state reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3.6: FE and IV Estimates – White, Black, and Hispanic Women

	(1) FE	(2) IV	(3) FE	(4) IV	(5) FE	(6) IV
	--------- White ---------		--------- Black ---------		-------- Hispanic --------	
Immigrant Share	2.20	1.98	1.27	2.85	2.93	-0.51
	(0.30)***	(0.46)***	(1.19)	(3.43)	(1.62)*	(3.04)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	255	255	254	254	253	253
R-squared	0.89	0.89	0.75	0.74	0.72	0.70
First-stage F stat		15.63***		4.14**		13.37***

Note: Author's own estimates based on data sourced from Ruggles et al. (2004). Observations are weighted by the size native-born female population aged 15 to 49 of each racial group. The sample size is 254 for column (3) and (4) because Vermont has no native born child bearing age black women in the 1970 sample and hence received a weight of zero. Arkansas and South Dakota are dropped in columns (5) and (6) because there are no native born child bearing age Hispanic women in the 1970 sample. Whites and blacks exclude Hispanic individuals. Robust standard errors clustered by state reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3.7: FE and IV Estimates – High, Middle, and Low Educated Women

	(1) FE	(2) IV	(3) FE	(4) IV	(5) FE	(6) IV
	---- High Educated ----		--- Middle Educated ---		---- Low Educated ----	
Immigrant Share	1.39	0.74	1.77	1.45	1.14	2.08
	(0.34)***	(0.59)	(0.54)***	(0.65)**	(1.04)	(1.34)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	255	255	255	255	255	255
R-squared	0.78	0.78	0.89	0.89	0.67	0.67
First-stage F stat		10.64***		11.66***		12.89***

Note: Author's own estimates based on data sourced from Ruggles et al. (2004). Observations are weighted by the size of native-born female population aged 15 to 49 of each education level group. Low educated individuals have at most 11 years of education; middle educated individuals have 12 years of education and no post-school education; high educated individuals have at least some college education. Robust standard errors clustered by state reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3.8: IV Estimates –Non-Movers

	(1) All	(2) White	(3) Black	(4) Hispanic	(5) High Ed.	(6) Mid. Ed.	(7) Low Ed.
Immigrant Share	1.44	2.24	4.52	-0.52	0.54	1.54	3.78
	(0.60)**	(0.66)***	(4.83)	(2.67)	(0.56)	(0.77)**	(1.92)**
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	255	255	247	246	255	255	255
R-squared	0.86	0.85	0.63	0.59	0.67	0.85	0.59
Number of states	255	255	247	246	255	255	255
First-stage F stat	15.87***	20.35***	4.64**	15.16***	16.13***	15.30***	15.62***

Note: Author's own estimates based on data sourced from Ruggles et al. (2004). Observations are weighted by the size of non-moving female population aged 15 to 49. Non-movers are individuals residing in their states of birth. Column (3) has 8 observations dropped because Alaska, Maine, Montana, New Hampshire, North Dakota, South Dakota, and Vermont have no native-born TFR for non-moving blacks in some years. Column (4) has 9 observations dropped because Alaska, Arkansas, DC, Maine, North Carolina, North Dakota, Rhode Island, South Dakota, and Vermont have no native-born TFR for non-moving Hispanics in some years. Robust standard errors clustered by state reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3.9: Potential Outlier Observations

(1) All	(2) White	(3) Black	(4) Hispanic	(5) High Ed.	(6) Mid. Ed.	(7) Low Ed.
DC,2000	DC,2000	DC,2000	DC,2005	AK,1980	AK,1970	AK,2005
DC,2005	DC,2005	DC,2005	ME,1970	DC,2000	DC,2000	DC,2000
HI,2000	HI,2000	HI,2000	ND,1990	DC,2005	DC,2005	DC,2005
HI,2005	HI,2005	HI,2005	NH,1970	HI,1970	HI,2000	HI,2000
MS,1970	LA,1970	ID,1970	NV,1970	HI,2000	HI,2005	HI,2005
ND,1970	ND,1970	ME,1990	OH,1970	HI,2005	ND,1970	ME,2005
NE,2005	UT,1980	SD,1980	VA,1970	ME,1970	RI,2005	MT,2005
SD,2005	VT,1970	UT,2005	VT,1990	MN,1970	UT,1980	NM,1980
UT,1980	WV,1980	WY,1980		ND,1970	VT,1970	
VT,1970	WY,1980	WY,1990		SD,2005	WY,1970	
WV,1980				UT,1980		
WY,1980				VT,1970		
				WY,1980		

Note: State-year observations listed above have conditional TFR and conditional share of immigrants in the second stage of IV regression greater than 2 standard deviations from the mean.

Table 3.10: IV Estimates – Excluding Outlier Observations

	(1) All	(2) White	(3) Black	(4) Hispanic	(5) High Ed.	(6) Mid. Ed	(7) Low Ed.
Immigrant Share	0.71	1.73	0.32	-0.96	0.41	1.26	1.93
	(0.44)+	(0.34)***	(1.62)	(3.11)	(0.40)	(0.54)**	(1.14)*
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	243	245	244	245	242	245	247
R-squared	0.92	0.91	0.78	0.72	0.81	0.90	0.70
Number of states	51	51	51	51	51	51	51
First-stage F stat	18.62***	18.10***	14.76***	13.71***	19.61***	18.40***	18.66***

Note: Author's own estimates based on data sourced from Ruggles et al. (2004). Observations are weighted by the size of native-born female population aged 15 to 49 of each group. White and black individuals do not include Hispanic individuals. Low educated individuals have at most 11 years of education, middle educated individuals have 12 years of education, and high educated individuals have some college education. Outlier observations listed in Table 3.9 are dropped. Robust standard errors clustered by state reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1, + p<0.11

CHAPTER 3 - REFERENCES

Altonji, Joseph G. and David Card (1991) "The Effects of Immigration in the Labor Market Outcomes of Less-skilled Natives," in *Immigration, Trade and Labor*, John M. Abowd and Richard B. Freeman, eds., University of Chicago Press, Chicago

Becker, Gary (1991) *A Treatise on the Family*. Harvard University Press, Cambridge, MA.

Borjas, George (2005) "The Labor Market Impact of High-Skill Immigration," NBER Working Paper No.11217.

Borjas, George. (2004) "Increasing the Supply of Labor through Immigration: Measuring the Impact on Native-Born Workers," Center for Immigration Studies Backgrounder, May.

Borjas, George (2003) "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *Quarterly Journal of Economics*, vol.118 (4), pp. 1335-1374

Borjas, George (1987) "Immigrants, Minorities, and Labor Market Competition," *Industrial and Labor Relations Review,* vol.40 (3), pp. 382-393

Borjas, George, Richard Freeman and Lawrence Katz (1997) "How Much Do Immigration and Trade Affect Labor Market Outcomes?" *Brookings Papers on Economic Activity*, vol.0 (1), pp. 1-67

Borjas, George, Richard Freeman and Lawrence Katz (1996) "Searching for the Effect of Immigration on the Labor Market," *American Economic Review*, vol.86 (2), pp. 246-251

Butz, William P. and Michael P. Ward (1979) "The Emergence of Countercyclical U.S. Fertility," *American Economic Review*, vol. 69 (3), pp. 318-328

Caldwell, John C. and Thomas Schindlmayr (2003) "Explanations of the Fertility Crisis in Modern Societies: A Search for Commonalities," *Population Studies*, vol.57 (3), pp. 241-263

Camarota, Steven A (2005) "Immigration in an Aging Society: Workers, Birth Rates, and Social Security," Center of Immigration Studies Backgrounder, April.

Bartel, Ann P. (1989) "Where do the New U.S. Immigrants Live," *Journal of Labor Economics,* vol.7 (4), pp. 371-391

Butz, William P. and Michael P. Ward (1979) “The Emergence of Countercyclical U.S. Fertility,” *American Economic Review*, vol. 69 (3), pp. 318-328

Card, David (2006) “Is the New Immigration Really So Bad?,” *Economic Journal*, vol.115, pp. F300-F323

Card, David (2001) “Immigrant Flows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration,” *Journal of Labor Economics*, vol.19 (1), pp. 22-64

Card, David (1990) “The Impact of the Mariel Boatlift on the Miami Labor Market,” *Industrial and Labor Relations Review*, vol.43 (1), pp. 245-257

Card, David and John DiNardo (2000) “Do Immigrant Inflows Lead to Native Outflows?,” *American Economic Review*, vol.90 (2) pp. 360-367

Cortés, Patricia (2008) “The Effect of Low-skilled Immigration on U.S. Prices: Evidence from CPI Data,” *Journal of Political Economy*, vol.116 (3), pp. 381-422

De Tray, Dennis N (1973) “Child Quality and the Demand for Children,” *Journal of Political Economy*, vol.81 (2) Part 2, pp. S70-S95

Frejka, Tomas (2004) “The ‘Curiously High’ Fertility of the USA,” *Population Studies*, vol.58 (1), pp. 88-92

Frey, William H (1996) “Immigration, Domestic Migration, and Demographic Balkanization in America: New Evidence for the 1990s,” *Population and Development Review*, vol.22 (4), pp. 741-763.

Frey, William H (1995) “Immigration and Internal Migration ‘Flight’ from US Metropolitan Areas: Toward a New Demographic Balkanization,” *Urban Studies*, vol.32 (4), pp. 733-757

Goldenweisier, E. A. (1912) “Walker’s Theory of Immigration,” *American Journal of Sociology*, vol.18 (3), pp. 342-351

Heckman, James and James Walker (1990) “The Relationship between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data,” *Econometrica*, vol.56 (6).

Khananusapkul, Phanwadee (2004) “Do Low-skilled Female Immigrants Increase the Labor Supply of Skilled Native Women?,” Ph.D. Dissertation, Harvard University.

Lindo, Jason M. (2010) “Are Children Really Inferior Goods? Evidence from Displacement-Driven Income Shocks,” *Journal of Human Resources*, vol.45 (2), pp. 301-327

Lopez, Mary Jane (2003) "High-Skill Immigrants in the United States Economy: Do High-Skill Immigrants Substitute or Complement Native-Born Workers?" mimeo, University of Notre Dame.

Macunovich, Diane J. (1995) "The Butz-Ward Fertility Model in the Light of More Recent Data," *Journal of Human Resources*, vol.30 (2), pp. 229-255

Munshi, Kaivan (2003) "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," *Quarterly Journal of Economics*, vol.118 (2), pp. 549-599

Ottaviano, Gianmarco I. P. and Giovanni Peri (2008) "Immigration and National Wages: Clarifying the Theory and the Empirics," Unpublished paper.

Ottaviano, Gianmarco I. P. and Giovanni Peri (2006) "Rethinking the Effect of Immigration on Wages," NBER Working Paper No. 12496, Cambridge, MA.

Rollins, Weld A. (1930) "The Effect of Immigration on the Birth Rate of the Natives," *Journal of Heredity*, vol.21 (9), pp. 387-402

Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander (2004) *Integrated Public Use Microdata Series: Version 4.0* [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor].

Saiz, Albert (2003) "Room in the Kitchen for the Melting Pot," *Review of Economics and Statistics*, vol.85 (3), pp. 502-521

Saiz, Albert (2006) "Immigration and Housing Rents in American Cities," IZA Discussion Paper No.2189.

Schulz, T. (1985) "Changing World Prices, Women's Wages and the Fertility Transition: Sweden 1860-1910," *Journal of Political Economy*, vol.93, pp. 1126-1154

Shergold, Peter R. (1974) "The Walker Thesis Revisited: Immigration and White American Fertility, 1800-60," *Australian Economic History Review*, vol. 14, pp. 168-189

Walker, Francis A. (1896) "Restriction of Immigration," *Atlantic Monthly*, vol.77 (464), pp. 822-829

Walker, Francis A. (1891) "Immigration and Degradation," *Forum*, vol. 11, pp. 634-643

Willis, Robert J. (1973) "A New Approach to the Economic Theory of Fertility Behavior," *Journal of Political Economy*, vol.81 (2), pp. S14-S64

Yasuba, Yasukichi (1962) *Birth Rates of the White Population in the United States, 1800-1860: An Economic Study*, Johns Hopkins Press, Baltimore.