HOW EFFECTIVE ARE PUBLIC HEALTH EDUCATION PROGRAMS, UNFETTERED FARM MARKETS AND SINGLE SEX SCHOOLS?

by

Jonathan Franklin Fox

Creative Commons 3.0 Attribution-Share Alike License

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ECONOMICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2010

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation

prepared by Jonathan Franklin Fox

entitled How Effective are Public Health Education Programs, Unfettered Farm Markets and Single Sex Schools?

and recommend that it be accepted as fulfilling the dissertation requirement for the

Degree of Doctor of Philosophy

____ Date: 28 May 2010
Price V. Fishback

____ Date: 28 May 2010
Paul W. Rhode

____ Date: 28 May 2010
Ron L. Oaxaca

____ Date: 28 May 2010
Manuela Angelucci

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

____ Date: 28 May 2010
Dissertation Director: Price V. Fishback

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. This work is licensed under the Creative Commons Attribution-Share Alike 3.0 United States License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/3.0/us/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

SIGNED: Jonathan Franklin Fox

ACKNOWLEDGEMENTS

My first thanks goes to Price Fishback, a better dissertation chair than I ever could have asked for. Always willing to help, always open to questions, making sure my work was on schedule, that it was focused and interesting, and that I was presenting it in a number of different conferences, I cannot say enough about how important his guidance was. The other members of my dissertation committee were also incredible. Ron Oaxaca, Manuela Angelucci and Paul Rhode, thank you. I am very appreciative for the experiences that I had with the rest of the faculty at the University of Arizona as well. Whether I was in a class or just dropping by to ask a question, I always felt like they all were there to help. Price, Paul, Ron, Manuela and the rest of the Economics faculty at Arizona, thank you all so much for giving me your time, energy and ideas to help improve my work.

I wish to also thank my parents, Frank and Mary Fox. I say this with no exaggeration, their support made it possible for me to finish graduate school. Emotionally or financially, they were always there to give me whatever I needed to be successful. I am truly lucky to have such an awesome mom and dad.

To all of my friends that helped me get through graduate school and kept me grounded, I am very grateful. Ben, Chris, Jan, Mike, Oscar, and everybody else, thank you. Arthur Emery, it takes a true friend to proofread those rough drafts of research papers and grant proposals about early public health programs and historical agricultural markets. Thank you. Trevor, Brendan, Dina, Matt, Alex, Carl and all of the other graduate students that I worked and studied with (especially those in my cohort), you guys are awesome. You made graduate school fun.

Lastly, I want to thank the Economic History Association and National Science Foundation for funding that allowed my dissertation to be completed. I am extremely grateful for their support.

DEDICATION

This is dedicated to my mom and dad. You guys are great.

TABLE OF CONTENTS

TABLE OF CONTENTS – *Continued*

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES – *Continued*

ABSTRACT

My dissertation examines the effectiveness of three policy choices in meeting socio-economic goals. The first analyzes the impact of public health education and poverty relief on child mortality in the early twentieth century, when infant and child mortality rates in the United States were startlingly high. During the 1920s, the rates dropped significantly and only part of the declines can be attributed to major sanitation and water projects in cities. Using a fixed effects identification strategy and adjusting to 2007 dollars, about $29,000 in public health education spending and about $781,000 in poverty relief spending were each associated with an infant death avoided. In comparisons with many modern programs, these costs associated with saving infant lives in the early 1920s were low. After controlling for city-specific trends in mortality, the effect of public health education programs is attenuated. This potentially suggests that with public health education, it is the stock of knowledge that is important. The second part of my dissertation examines the sensitivity of agricultural prices and output to local and non-local weather fluctuations in the United States prior to 1932, when markets were relatively open and largely unfettered by modern farm programs. The price sensitivity to these local and non-local weather fluctuations is estimated for the crops cotton and wheat, which have relatively low transport costs and are primarily exported to non-local markets, as well as for corn and hay, crops with high transport costs and used in local productive activities. For cotton and wheat, changes in local weather seem to have little effect on farm-gate prices, while changes in weather affecting the aggregate market play an important role. Corn and hay prices are much more sensitive to changes in state-level temperature, precipitation, and drought conditions. The third study examines the returns to education for women who attended a college with a predominantly female population. Using the program evaluation framework and matching tech-

niques, I find that attending a female-dominated school yields positive labor market effects on the order of about 15 percent upon first entry into the labor market but that these effects seem to diminish over time.

CHAPTER 1

PUBLIC HEALTH MOVEMENTS, LOCAL POOR RELIEF AND CHILD MORTALITY IN AMERICAN CITIES: 1923-1932

1.1 Introduction

During the early 20th century, infant mortality and child mortality declined substantially both in urban and rural areas. Prior to 1910, for every 1,000 babies born in the United States, 165 died, and the rate was even worse in the countryside (Newmayer 1911). However, by 1920 the infant mortality rate was cut in half to about 85 deaths per 1,000 live births, and by 1930 it dropped to 65. Child mortality experienced similar declines. Only part of the declines can be attributed to major sanitation and water projects in cities. Public health historians suggest that one reason for the decline was improvements in the education of the population about simple health procedures like hand washing and boiling water. This paper investigates these early 20th century public health education programs in the large municipalities and estimates their influence on declining mortality among children and infants. The educational programs of the state, municipal, and county health departments were much more cheaply implemented than the large scale public works projects occurring in the cities during the same time. Evaluating the effectiveness of such policies allows better understanding about how such low-cost policies might work in reducing child and infant mortality in countries facing conditions similar to those faced by industrializing cities in the early 1900s.

In the first few decades of the 20th century, the U.S. experienced great reductions in infant and child mortality rates. But the adjustments were uneven, and there existed substantial variation across locations in the mortality trends, as well as in the type and extent of government and public health programs. Before the New Deal of the 1930s, few federal welfare or public health programs existed, and those that did were either investigative bodies or mandated states to distribute benefits to certain classes of people. As was the case for decisions about poverty assistance, most health spending decisions were made at the state, city or county level. For this reason, this paper focuses on the period prior to the introduction of the New Deal in 1932. By choosing the decade of the 1920s, I can analyze the effectiveness of state and local public health and poverty assistance programs at saving the lives of children without them being confounded by large-scale changes associated with the New Deal.[1]

Given the concurrence of the declines in child and infant mortality with the growth of public health work, it is natural to think the two related. Many believe that the public health education work played an important role in improving outcomes (Ferrell et al 1932, Rockefeller 1921, Blackburn 1927, Ewbank and Preston 1990), but as yet no one has tested their impact with modern econometric methods. Without data on the state and local health and education programs, researchers studying other determinants of the mortality decline have controlled for the influence of the public health education movement using techniques such as difference-in-differences (Cutler and Miller 2005) or year and geography fixed effects (Troesken 2004). With new data on municipal health education spending, I can separately evaluate its effects from those of other city spending.

It is important to evaluate the success of these programs because they are inexpensive relative to the large sums needed for sanitation and water filtration works. In settings where locales have limited access to the necessary capital to build the sanitation and filtration infrastructure, these programs are potentially low cost ways to reduce mortality while waiting to build the larger public works. They also may be low cost and complementary ways to lower mortality even after the works are built.

[1] The analysis is constrained to those years after 1923 because the level of financial detail necessary is not available between the years 1920-1922

1.2 Public health education and poverty relief in the early 20th Century

Public health and poverty assistance programs first started gaining support in the early 20th century as birth and death registration areas grew and the data collected on births and deaths were collected on a more consistent basis, and became more reliable for comparisons. Demographers began to have a clearer picture of the how poorly children fared in the United States compared with other developed countries, and public health advocates began to question if perhaps the U.S. could possibly do better (see Newmayer, 1911). Research by scientists such as Louis Pasteur in the late 1800s on the relationship between sanitation and health, and by Paul Karrer on the importance of vitamins and nutrients offered ways in which these issues could be addressed. Several public programs were designed with the goal of reducing child mortality rates. Child and woman advocacy groups (Skocpol et. al, 1993) influenced politicians to pass legislation such as mothers' pensions, form organizations such as the Children's Bureau, and encourage state and city departments of health to form child hygiene divisions and distribute information about how to improve health outcomes.

The Children's Bureau, formed in 1912, was charged with investigating and reporting on all matters pertaining to the welfare of children and child life. Through its publications and political presence, the Children's Bureau helped bring attention to the exceptionally high mortality rates in some U.S. cities and for certain classes of people. Although its mandate included investigation of the "whole child," a limited budget and reluctance to duplicate work done by other federal agencies induced the bureau to limit its initial focus to the causes and potential solutions of the high infant mortality rates. An inquiry into these causes and solutions in the city of Johnstown, PA was the first field study done by the Children's Bureau and for the first two years absorbed almost its entire attention (Department of Labor, 1915). Other cities were chosen for case studies to isolate factors associated with different types of industrialization.[2]

[2]By 1918, the field studies included a steel city (Johnstown, PA), two textile cities (Manchester,

The studies led the Children's Bureau to conclude that high infant mortality rates were not only the result of poor hospital care or ignorance among birthing mothers but also the result of a range of socioeconomic factors related to poverty. In 1916, Julia Lathrop, chief of the Children's Bureau, mentions the "coincidence of a high infant mortality rate with low earnings, poor housing, mother's work and large families (Department of Labor, 1916)." She expanded on these ideas in her contribution to the 1920 Report of the Department of Labor:

> From the findings in Baltimore certain facts stand forth to which we as a Nation can no longer close our eyes. Without qualification - regardless of color, race or nationality - the infant death rate varies inversely with the father's income. When the father's income represented the ability to insure care and comfort ($1,850 a year or more) the infant death rate was one-fourth as high as when the father's earnings fell into the lowest wage group.

The bureau's findings stressed the importance of socioeconomic conditions and emphasized a middle class family ideal, for the most part ignoring the impact of medical causes. Since the Children's Bureau did not have a physician on staff for the initial field studies and wanted to avoid stepping on the toes of the American Medical Association, factors like the importance of proper medical care, clean milk and other sanitation related variables were left up to the Public Health Service to study (Lindenmeyer 1995).

The Children's Bureau did, however, encourage the development of maternal and child hygiene divisions within city and state health departments, and also lobbied strongly for the Sheppard-Towner Act, which passed Congress in 1921. This act constituted the first federal public health program and had its primary focus on

---

N.H. and New Bedford, MA), a center for the manufacture of high grade shoes (Brockton, MA), a manufacturing city with no one dominant industry (Saginaw, MI), a city with the production of brass as its dominant industry (Waterbury CT), a rubber manufacturing center (Akron, OH) and lastly, a large cosmopolitan area (Baltimore, MD) (Lathrop, 1918).

health education. Federal matching grants were distributed to states with specific instruction in their use. Recipients were prohibited from using the money for the "purchase, erection, or repair of any building or equipment, or for the purchase or rental of any buildings or lands, or for any maternity or infancy stipend, gratuity or pension (Abbott 1922)." Instead, the money was intended to pay for the operation of health centers to instruct mothers in hygienic ways and to distribute pamphlets to new mothers about how best to care for their baby (Thompson 1921).

The Sheppard-Towner grants consisted of an initial \$5,000 sum, and additional money in the form of matching grants up to some specified maximum based on a state's population. If a state chose to accept the Sheppard-Towner appropriations, it was required to have a maternity and infant hygiene division within their health department. For this reason and others, Connecticut, Illinois, and Massachusetts chose not to accept the Federal grants.[3] Additionally, localities within participating states had to be prosperous enough to be able to match the grants, although counties and cities that did not directly receive Sheppard-Towner funds still felt its effects. Many of the maternal and child hygiene divisions arose in the state and municipal health departments shortly after the Act's passage (Ferrell et al. 1932, U.S. Public Health Service 1923), and in at least one state, North Carolina, the funds were the primary support for its Bureau of Maternity and Infancy.

The Children's Bureau studied the reasons behind the high child mortality rates because they knew that information was essential in designing policy to combat them. This perspective still holds. If poverty was the primary cause, then welfare-type social programs would be most effective at reducing the number of child deaths. Alternatively, if a general absence of nutritional and birthing information was the

[3]Connecticut did not engage in county health work, but sent out literature on baby hygiene to new mothers and held "well-child" conferences in various towns throughout the state. Illinois had established 6 county health organizations between 1922 and 1929 (although these all had been discontinued by 1930), subsidized clinics for treatment of indigents, and promoted maternal and infant hygiene through distribution of prenatal literature and sending out nurses for personal instruction to mothers. Massachusetts set up "well-child" conferences and had a law requiring the medical examination of all school children.

issue, then spending on health and mother's education would be more effective. Because I have data on two different types of social programs, one educational and another directed towards transfer payments, I am able to separately estimate how effective each of these different types of programs were, and give an empirically rigorous answer to the question pondered by the Children's Bureau. To accomplish this, I exploit a panel data set that allows identification of the effects of the programs within cities over time while controlling for nationwide shocks and mortality time trends within the cities. Using that approach, I can help answer which types of government spending were most effective for different age groups.

1.3 The data and basic correlations

The panel data set is composed of annual information from 67 cities with populations over 100,000 during the period 1923-1932. Those dates were chosen both for data availability reasons, and to eliminate the effect of any New Deal programs enacted after 1932.[4] City financial data, including spending on sanitation, health, mothers' pensions, and other forms of poverty relief were collected from the *Financial Statistics of Cities* volumes, published by the Department of Commerce. Per capita summary statistics adjusted to 2007 dollars for each of the spending variables are given in the top panel of Table 1.1. Population data were also collected from the *Financial Statistics of Cities* volumes, and when missing, estimated between the nearest two years.[5]

The two financial variables of primary interest are the spending on public health education in a city, and the spending on poverty assistance. Spending on public

[4]Fishback, Haines, and Kantor (2007) examine the time period from 1929 through 1940 to examine the role of the New Deal in influencing infant deaths, noninfant deaths and births.

[5]The cities interpolated were: Los Angeles, CA, 1924-1927; Seattle, WA, 1924-1927; Portland, OR 1925-1927; Akron, OH 1924-1927; Bridgeport CT, 1924-1927; New Bedford, MA, 1926-1927; Norfolk, VA , 1924-1925; Lowell, MA, 1926-1927; Lawrence, MA, 1926-1927; Elizabeth, NJ, 1924-1927; Erie, PA, 1924-1927; Waterbury CT, 1924-1927; Jackson, FL, 1926-1927; Hoboken, NJ, 1923-1925; Brockton, MA, 1926-1927; Davenport, IO, 1926-1927; Haverhill, MA, 1926-1927; Wheeling, WV, 1923-1927; Superior, WI, 1923-1927; Auburn, NY, 1926-1927; Newport, VA, 1923-1924.

health education includes spending on the medical inspection of school children and spending for education about proper hygiene, milk preparation techniques and other things that could be done to conserve child life. Money distributed under the "medical inspection for school children" category helped pay for physician and nurse visits to distribute information and perform physical examinations. School children were not treated, but their parents were informed if any defects were found. Spending on poverty assistance includes spending on mothers' pensions, funding for almshouses and orphanages and other charitable spending for children. "Outdoor care"[6] of the poor who lived outside almshouses generally comprised the largest portion of poverty assistance, especially for cities with populations between 100 and 300 thousand, since many of those did not provide aid in the form of mothers' pensions.[7] Adjusted to 2007 dollars, an average city in the dataset spent about $3.15 per person on health programs for children and about $16.23 per person on poverty assistance.

The city spending data were matched with city mortality data entered from the Mortality Statistics volumes, published by the Department of Commerce. Figure 1.1 plots the death rates for infants, as well as for children aged 1 to 4, and 5 to 9. These death rates are defined as the number of deaths occurring in a city per 1,000 people. Although some age groups experienced greater drops in mortality than others, every child age group experienced mortality declines between 1923 and 1932. This is particularly interesting given that, except for Milwaukee, every city in the sample had developed their water and sewer systems prior to the start of the panel. The mortality rates for infants for cities in the sample dropped from about 1.8 infant deaths per 1,000 people in 1923 to about 1 infant death per 1,000 people in 1932. Note that this mortality rate differs from the standard infant mortality rate, calculated as infant deaths per 1,000 live births. I do not use the more conventional infant

[6] This typically involved relief to individuals or families that due to unemployment, illness, accident, or for perhaps some other reason, were temporarily dependent. It also sometimes involved the giving of aid more or less permanently, when it seemed desirable to keep a family together instead of scattering its members among institutions.

[7] 24 out of the 67 cities in the panel did not provide aid in the form of mothers' pensions.

mortality rate in my analysis because the size of the birth registration area in 1923 was much smaller than the size of the death registration area. Specifically, usage of the more conventional infant mortality rate requires dropping cities in Alabama, Georgia, Louisiana, and Texas that are currently in the sample. For the sample cities that were also part of the birth registration area, the infant mortality rate did decline substantially, from about 78.5 in 1923 to about 55.9 in 1932. Mortality rates for children aged 1-4 also decreased considerably over this period, dropping nearly 60 percent from their level in 1923. Meanwhile, mortality rates for children aged 5-9 decreased only slightly. Both in absolute terms, and relative to the other child age groups, mortality rates for infants experienced the greatest improvement. In 1923 the mortality rate for infants was at least twice as large as the mortality rate for any of the other child age groups, but by 1932, the gap had fallen significantly.

Figure 1.2 plots the annual mean-differenced number of infant deaths per thousand people within cities for groups of cities in the bottom and top quartiles of aggregate health education spending between 1923 and 1932. The trends show that cities that spent a relatively large amount on health education generally had infant death rates much greater than other large cities in the early 1920s. However, by 1932 these cities were performing better than the average.[8] Figure 1.3 plots the annual mean-differenced death rates for children aged 1 to 4 for the same sets of cities, and a similar story occurs. Cities that spent more on health education between 1923 and 1932 on average experienced worse death rates in the early 1920s, while cities that spent less on average had child death rates below the mean. However, by 1926, the positions of the bottom and top quartiles had reversed.

Figures 1.4 and 1.5 perform the same exercise for cities at the top and bottom quartiles of poverty relief spending. Stratifying the cities by their level of poverty relief spending reveals an even more striking difference in trends. The number of infant deaths per thousand people fell substantially relative to the average within cities in the top quartile of poverty relief spending. However, cities that chose not

[8]Plotting the infant mortality rate for the top and bottom quartiles for the subset of cities with birth data reveals a similar pattern.

to invest in this experienced a growing gap between them and the average large city in the United States. Figure 1.5 illustrates similar trends.

The information plotted in figures 1.2 through 1.5 suggests that in the case of small children and infants, both spending on public health education and spending on poverty assistance were associated with improving health outcomes in the cities. To take this a step further, figures 1.6a and 1.6b plot the number of infant deaths per thousand people against per capita public health education spending (Figure 1.6a) and per capita poverty relief spending (Figure 1.6b). Figures 1.6a and b also include regression lines, which establish basic correlations between the death rate for infants and the amount of the different types of spending. These estimated regression lines display the raw correlations, and show what conclusions would have been drawn from the data with a method often used to evaluate the success of the policies in the 1920s and 1930s.[9]

The coefficient of -0.046 in the regression line in Figure 1.6a implies that a reduction of one infant death would have been associated with an additional 21,645 2007 dollars of per capita public health education spending. The coefficient of -0.0049 on poor relief spending in Figure 1.6b is an order of magnitude smaller. About 204,000 2007 dollars of poverty relief spending was associated with the reduction of one infant death. Figures 1.7 through 1.89 plot the health education and poverty relief spending against the death rates for children aged 1-4 (Figures 1.7a and b), and 5-9 (Figures 1.8a and b). The coefficients for the older age groups are all negative, but they are much smaller in magnitude than those for the infants and small children. From these basic correlations it appears that smaller amounts of health education spending than of poverty relief spending are associated with lower mortality rates for each of the different age groups. Additionally, infant death rates respond the most to increases in either type of spending.

For these basic correlations to estimate a causal effect, spending on public health education and poverty relief would need to have been completely uncorrelated with

[9]See, for instance, the Department of Labor Annual Reports, as well as Lathrop 1919, Abbot 1922, Tobey 1925, Vaughn 1922, and Levy 1920

any other factors that may have influenced death rates. Since this assumes elements such as per capita income or general schooling do not influence mortality and public health education jointly, this is clearly a very strong assumption. If charitable and public health spending was greater in cities with more per capita income and per capita income was correlated with lower mortality rates, then failing to include a measure of income would lead to estimating a much larger effect of the spending than was actually the case.

To control for the potential biases, I collected various measures of income and other correlates that likely influenced child mortality. The income measure is average annual earnings from the manufacturing sector. To control for differences in the distribution of income, an additional measure of the number of tax returns filed as a share of the municipal population in a year was collected. This gives the number of jointly filing couples in each city with incomes above \$5,000 (about \$60,000 in 2007 dollars), and individual filers with incomes over \$2,000. Although this is a measure of the income share held by the highest income people, it also provides indirect information about the share in the bottom tail of the distribution because I am simultaneously controlling for average manufacturing earnings. Information on demographics, municipal spending on hospitals, sanitation and other activities, and other data that could be related to both spending and mortality was also collected. These variables and their sources are explained in the Appendix A.

1.4 Econometric model and results

The panel data set is used in the following estimation equation:

$$UMR^a_{i,t} = \beta_1 PHE_{i,t-1} + \beta_2 CPR_{i,t} + \sum_{j=1}^{J} \beta_{J+3} X_{j,i,t} + \beta_{J+5} C_i + B_{J+5} Y_t + \epsilon^a_{i,t} \quad (1.1)$$

Where $UMR^a_{i,t}$ is the urban mortality rate for age group a in city i and year t. $PHE_{i,t-1}$ is the amount of spending on public health education occurring during the prior year in city i. This includes spending on the medical inspection of school children and spending distributed towrds educating persons about proper hygiene, milk preparation techniques and other things that could be done to preserve child life. I include the lagged term because it likely took some amount of time for people to implement the information they learned about pre and post-natal practices and caring for young children.[10] $CPR_{i,t}$ is is the amount of current year per capita poverty relief spending on children in city i. This variable includes spending on mothers' pensions, spending on almshouses and orphanages and other poverty relief spending directed towards children. $\sum_{j=1}^{J} \beta_{J+3} X_{j,i,t}$ is a set of J covariates include the county demographic variables percent black, percent illiterate, percent rural, and percent foreign born, as well as other city spending on sanitation, hospitals, education, health other than child health, and other charitable spending. It also contains the income and income distribution measures, as well as variables controlling for the amount of pollution within a city and the mortality rate of adults aged 20-29 to control for trends in mortality common across age groups. The errors are assumed to have mean zero, conditional on the covariates in the mortality equations and defined as the unobserved characteristics affecting mortality in city i, year t for each of the different age groups. I allow these error terms to be correlated between the different age groups, but because I estimate the mortality rate with the same covariates, this reduces to a basic OLS with covariates model (Wooldridge 2002). Finally, C_i and Y_t are vectors of city and year effects, respectively.

1.4.1 Fixed effects

To control for nationwide, annual shocks, associated with macroeconomic policy or other factors common across the sample cities in a specific year, year fixed effects are included in the analysis. The city fixed effects control for unmeasured factors that did not vary through time but did vary across cities. The most important feature that fits this definition is the quality of water treatment and sanitation infrastructure. Annual city sanitation spending is included among the set of co-

[10] Other distributed lag structures were estimated, both with respect to spending on public health education and spending on poor relief. The different lag structures estimated were t, t-1, and t-2. The estimated coefficients were stable across the different specifications.

variates, but most of this spending was devoted to street sweeping, trash collection and some maintenance. As a result, the city annual spending is not well correlated with infrastructure quality. In 66 of the 67 cities there were no major capital improvements to the water treatment and sanitation infrastructure over the period; therefore, absent depreciation, the quality of the infrastructure over the period was likely time-invariant in each city.[11] Better infrastructure would have tended to reduce death rates, implying a negative relationship between the sanitation and water treatment facilities and death rates. If the fixed effects to control for these major facilities were left out, the sign of the omitted variable bias will be determined by the relationship between sanitation and water treatment and a city's choice about public health education. If cities with better sanitation and water treatment infrastructure saw them as substitutes for public health education, they would have spent less on public health education. The combination of the negative relationship between infrastructure and death rates and the negative correlation between infrastructure and health education would impart a positive bias to the public health education coefficient. On the other hand, if cities with better infrastructure saw the public health education as a complement to the infrastructure, they might have invested in more public health education. This would then lead to a negative bias for the coefficient of public health education in the regressions without city fixed effects.

The coefficient of the poverty relief variable might also be affected by the quality of sanitation and water treatment infrastructure. If areas with better sanitation infrastructure were areas with more poverty relief spending, the combination of this positive correlation and the negative correlation between infrastructure and death rates would have led to a negative omitted variable bias for the coefficient on poverty relief.[12]

[11]Milwaukee added a water treatment plant in 1926.

[12]After controlling for income, income distribution, various types of city spending, and city and year fixed effects, I believe I have controlled for much of the potential endogeneity that might arise. I have attempted to use instrumental variable analysis in case there is some left over endogeneity. Thus far, I have not found any variables that are sufficiently correlated with health or poverty spending to get past weak instrument problems. Some of the potential instruments tried, but

Table 1.2 presents estimates from equation 1 for the age groups 0-1, 1-4 and 5-9. Comparisons of the coefficients in Column 1 in Table 1.2 with their corresponding matches in figures 1.6-1.8 suggest that the poverty relief coefficients in figures 1.6a through 8a were biased in a negative direction. Once the set of covariates and city and year fixed effects are included in the specification, the coefficients on the poverty relief spending in Column 1 are all less negative than those displayed in the figures. The only Column 1 coefficient that is statistically significant is in the infant death rate equation.

Once the covariates and fixed effects are included, and the level of sanitation and water treatment infrastructure controlled for, public health spending is associated with reductions in infant deaths, and to some degree deaths of children aged 1-4 (p-value of 0.11). The negative relationship was much stronger for public health education spending than it was for city welfare spending for children. One additional dollar per capita spent on child health education related activities in the prior year was associated with a 0.0358 point reduction in the mortality rate for infants. This implies that about an additional 28,000 2007 dollars were associated with one infant death avoided. An additional dollar spent per capita on poverty relief for children, which included funds distributed outside almshouses, mothers' pensions and other spending for the aid of children was associated with a 0.00142 point reduction in the mortality rate for infants, implying that about 700,000 2007 dollars were associated with one infant death avoided. Compared to the estimates of the statistical values of life calculated in Costa and Kahn (2004) for 1940 (about 1 to 1.5 million in 2007 dollars), allocating money to either charitable spending or public health education would yield benefits much greater than the costs.[13] Additionally, comparing these estimates to the results in Fishback, Haines, and Kantor (2007) (converted to 2007 dollars, about $2.3 million was the estimated the relief cost per infant death pre-

which have shown strength, are state-level voting patterns, the timing of a state's women's suffrage enactment, and whether or not a state chose to participate in the Sheppard-Towner Act.

[13]According to a 1912 County Health Organization pamphlet distributed in North Carolina, Irving Fisher also calculated the net worth of an American infant life, and found it to be worth about $90 in 1912, or about $2,150 when that number is converted to 2007 dollars.

vented.) suggests that the public health and poverty assistance programs prior to the 1930s saved a statistical infant life at a much lower cost than New Deal relief (although the New Deal relief was not specifically targeted at saving infant lives like the public health and poverty assistance spending in the 1920s were).[14] But while both public health and poverty relief appeared to reduce mortality at a good cost-benefit ratio, the spending on public health education resulted in the largest benefits; far larger than those conferred by poverty relief.

1.4.2 Controlling for mortality trends in cities

It is possible that inclusion of city and year fixed effects still does not completely control for all of the potentially confounding factors affecting both mortality and the spending variables of interest. For instance, by 1923, nearly all of the cities in the sample have had their sanitation and water treatment infrastructure in place city for some time. It is possible there is trended depreciation in the quality of the infrastructure that differs from city to city depending on when their infrastructure was built.

This decline in quality of infrastructure would be associated with higher death rates, a negative relationship. If cities then spent more on public health education to offset the depreciation in infrastructure, the relationship is negative and the combination of the two negative relationships would lead to a positive omitted variable bias. If cities with deteriorating quality also spent less on public health education, due, for instance, to them losing interest, the correlation between the sanitation structure quality and public health education is positive. Combining the negative relationship between infrastructure quality and death rates and the positive correlation between infrastructure quality and public health education leads to a negative omitted variable bias.

[14] I expect the reasons for this are that (1) much of this public health education was aimed at the reduction of infant and maternal mortality and likely did not affect deaths from homicides and suicides as the New Deal relief spending did, and (2) the higher infant mortality rates offered more opportunities for reducing them with lower cost solutions.

A random trend variable

To control for these differences in depreciation and their potential biases, I estimate a model that includes a random trend variable $g_i t$ in equation 1. The inclusion of this variable removes the trend in mortality in each of the different cities, thereby identifying the effect of public health education and poor relief spending off of deviations within that trend. The coefficient estimates for each of the different age groups are given in Table 1.3. Not surprisingly, the coefficient estimates on public health education spending are no longer statistically significant. Additionally, the estimated coefficients between the two models are statistically different at the 99 percent level of significance. Because I am now identifying off of deviations in the trends, it may be that there is not enough variation left to precisely estimate the coefficients for the different variables. Examining this, I find that after controlling for city and year fixed effects, the variation in $PHE_{i,t}$ is reduced by about 70 percent, while the variation in $CPA_{i,t}$ is reduced by about 30 percent. Controlling for city fixed effects and city-specific trends, the variation in $PHE_{i,t}$ is reduced by about 92 percent, while the variation in $CPA_{i,t}$ is reduced by about 68 percent.

1.5 Concluding remarks

Besides the collecting of birth and death registration certificates and treating cases of malaria and tuberculosis, public health education was the primary method of interaction between state health departments and the public in the early 20th century. During the 1920s, many different states and cities engaged in educating the public about proper ways to care for infants and how to keep children healthy. Many municipal health departments held "well-child" conferences set up infant-welfare stations to observe the health of newborns and sent out bulletins and newspaper press releases (American Public Health Service 1923). During the same period, infant and child mortality rates fell drastically. With the cities in the sample, all of the age groups studied experienced declines of at least 25 percent in their mortality rates, with infants and children aged 1-4 respectively showing even greater drops of 50 and 60 percent. In this analysis I examine the extent to which these mortality declines can be explained by expanded spending on public health education and poverty relief in the 1920s and 1930s. Using a model with city and year fixed effects, the analysis shows that both types of programs contributed to reductions in infant mortality. Using a fixed effects model, public health education was more cost effective than poverty relief. Adjusted to year 2007 dollars, approximately $28,000 spent on public health education was associated with the prevention of an infant death. Meanwhile an additional $700,000 in poverty relief spending was associated with the same effect. These cost figures are much lower than those found in modern studies of Medicare expenditures and studies of the impact of work relief during the 1930s. Once the trends are controlled for using a random trend variable, the effect of public health education on reducing child mortality rates is greatly attenuated. This potentially suggests that with public health education, it is the stock of knowledge that is important, not a sustained flow of marginal increases. On the other hand, poverty relief appears to affect child and infant mortality much differently. The effect of poverty relief on child and infant mortality remains in the model that controls for mortality trends using a city-specific trend variable. As such, in the case of poverty relief, it appears that a sustained flow of income is important in reducing child and infant mortality rates. Although prior scholars could not directly measure these changes, they hypothesized that the simple lessons taught in the public health programs were very effective. The analysis shows that, while not diminishing the importance of poverty relief efforts, the public health programs instituted prior to the New Deal have been among the most cost effective programs in American history.

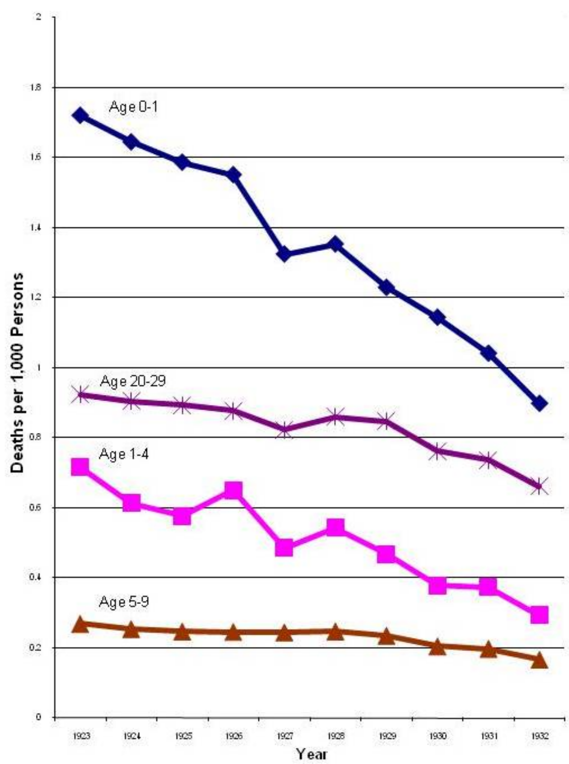


Figure 1.1: Deaths per 1,000 Persons by Age Group

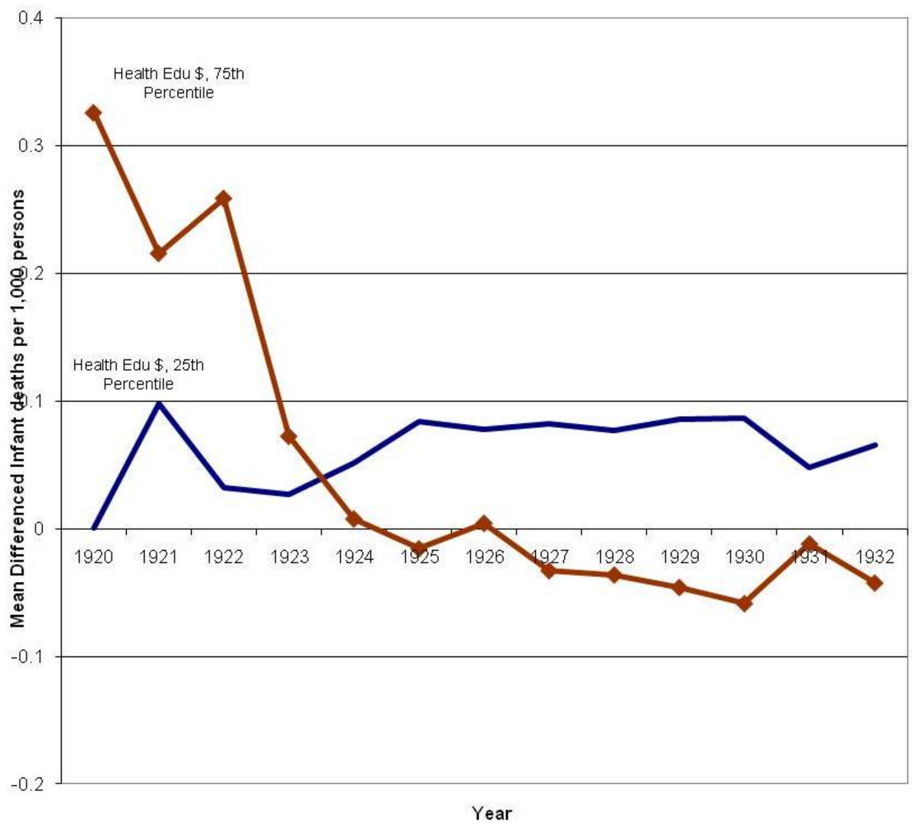


Figure 1.2: Infant Death Trends in Cities with More and Less Health Education Spending

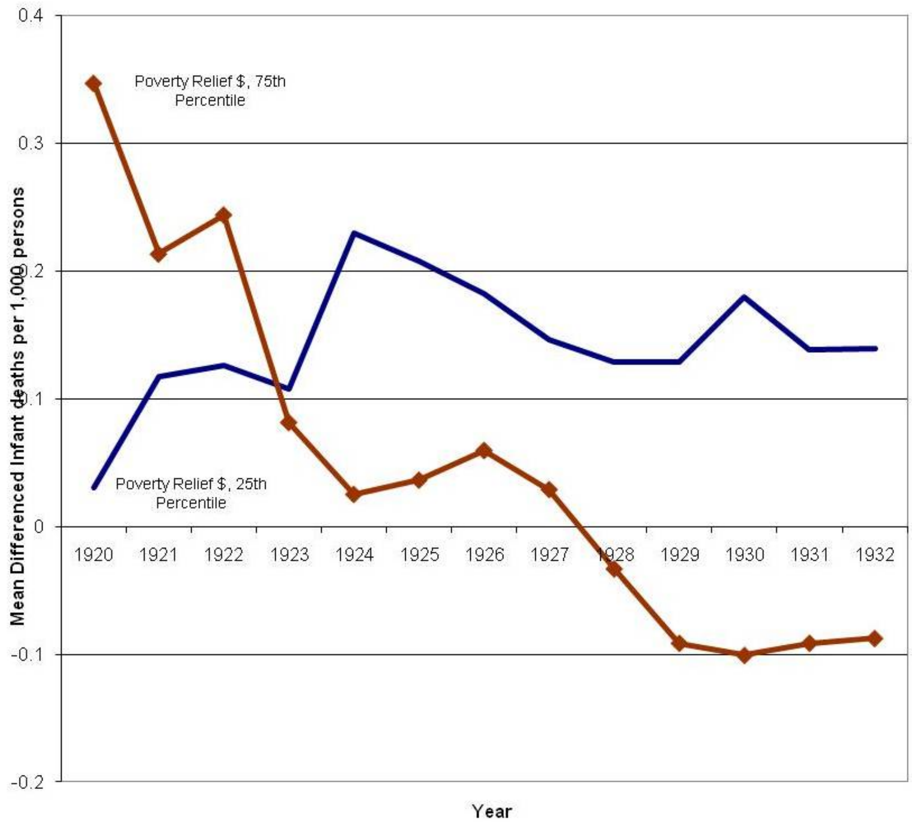


Figure 1.3: Infant Death Trends in Cities with More and Less Poverty Relief Spending

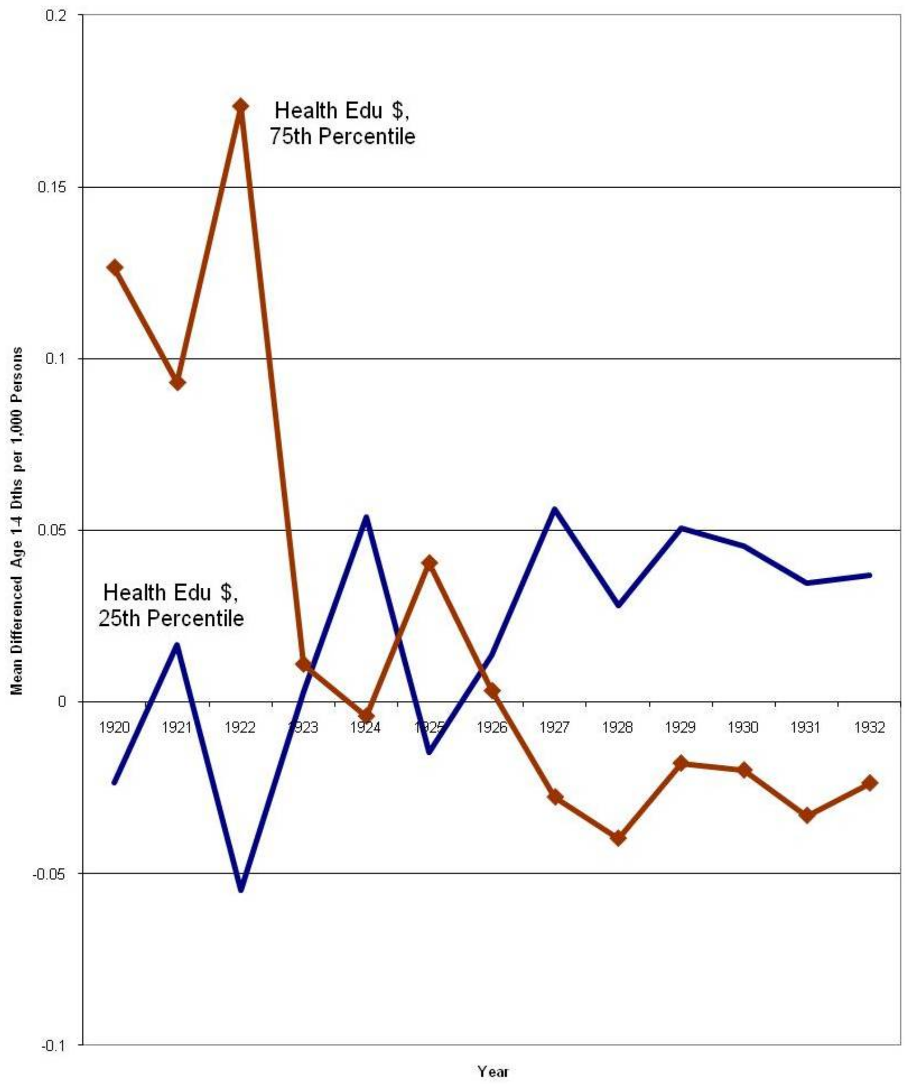


Figure 1.4: Children Age 1-4 Death Trends in Cities with More and Less Health Education Spending

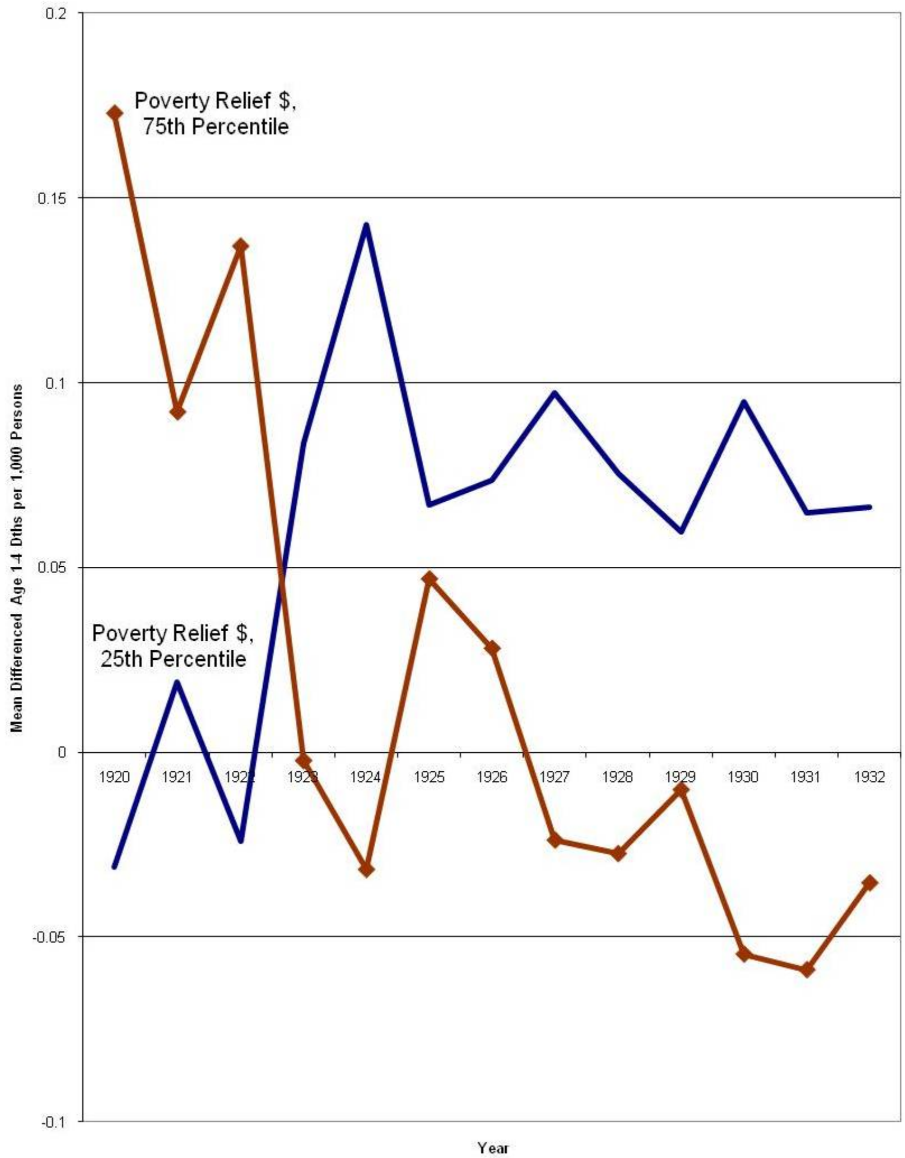
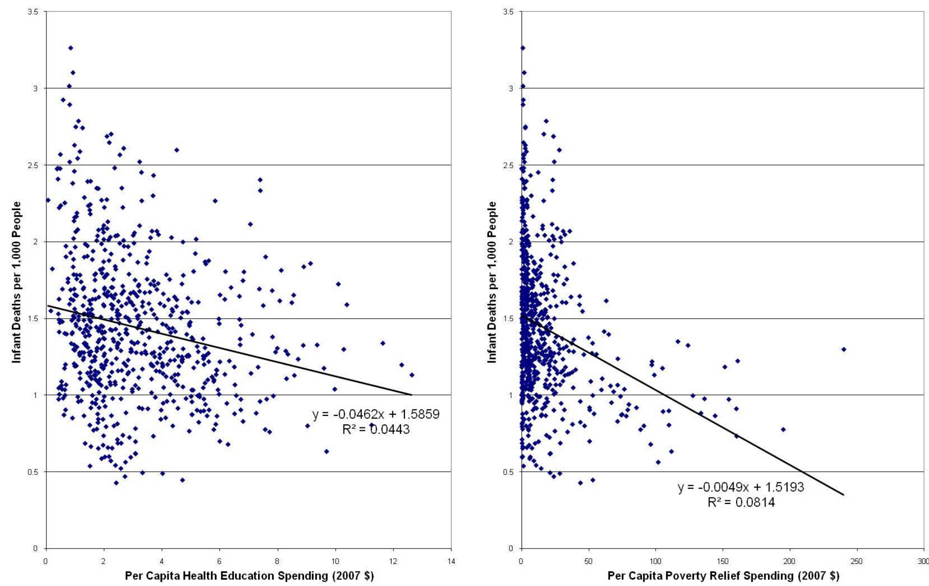
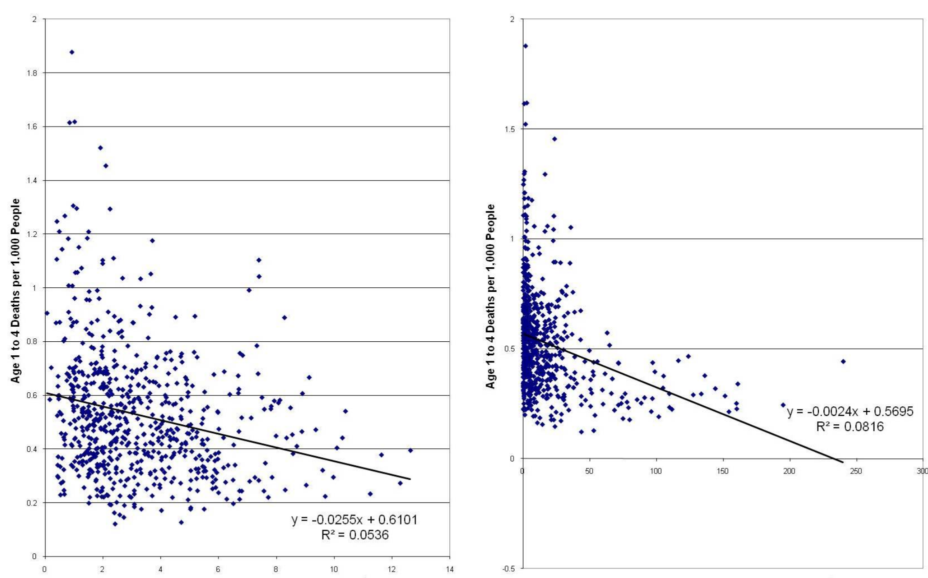


Figure 1.5: Children Age 1-4 Death Trends in Cities with More and Less Poverty Relief Spending



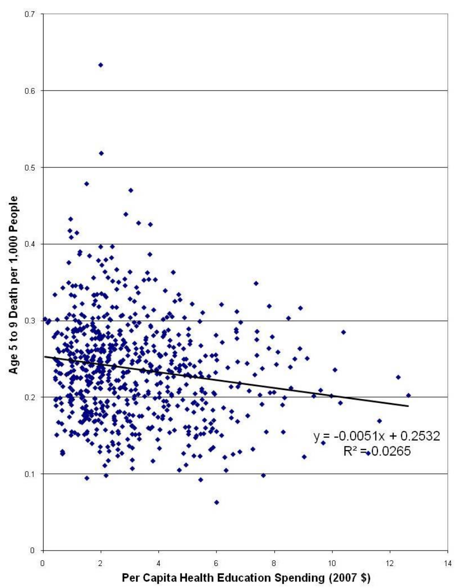
(a) Health Education (b) Poverty Relief

Figure 1.6: Death Rate for Infants Against Spending Variables of Interest (Per Capita 2007 $)

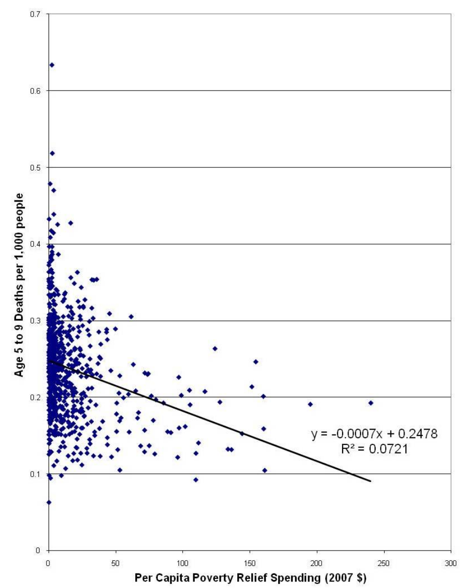


(a) Health Education (b) Poverty Relief

Figure 1.7: Death Rate for Children Age 1-4 Against Spending Variables of Interest (Per Capita 2007 $)



(a) Health Education



(b) Poverty Relief

Figure 1.8: Death Rate for Children Age 5-9 Against Spending Variables of Interest (Per Capita 2007 $)

Table 1.1: Summary Statistics

Per Capita Spending Variables	**Mean**	**Min**	**Max**	**Zeros**
Key Spending Variables				
Health education spending	3.15	0.07	12.63	0
Child poverty relief spending	16.23	0	239.83	19
Other Spending Variables				
Other health spending	10.10	1.27	39.46	0
Sanitation spending	32.54	7.75	102.44	0
Other charitable spending	5.47	0	50.29	127
Spending on hospitals	12.66	0	106.99	92
Spending on schools and libraries	190.50	68.22	383.88	0
Income and Income Distribution Correlates	Mean	Min	Max	Zeros
Avg. annual mfg wages	15,174.34	3,403.06	28,341.15	0
Number of tax returns filed	0.07	0.01	0.22	0
Demographic Measures	Mean	Min	Max	Zeros
Percent black	0.08	0.00	0.43	0
Percent foreign born	0.19	0.00	0.46	0
Percent illiterate	0.03	0.01	0.11	0
Percent rural	0.12	0	0.39	113

Table 1.2: Fixed Effects Estimates

	Age 0-1	Age 1-4	Age 5-9
City spending variables of interest			
Lagged PHE	-0.0358	-0.0082	-0.0002
	(0.0091)***	(0.005)	(0.0026)
Poverty Relief	-0.0014164	0.0002	0.0002
	(0.000367)***	(0.0002)	(0.0002)
Women's suff. ("Pre 1914" omitted)			
1915-1919	-0.3711	-0.4952	-0.0360
	(0.4995)	(0.24326)*	(0.0958)
1920	0.2384	-0.0700	0.0777
	(0.2780)	(0.148)	(0.0595)
Other spending variables			
Other health spending	-0.0003	-0.0006	-0.0005
	(0.0046)	(0.003)	(0.0008)
Sanitation spending	0.0006	0.0009	0.00002
	(0.0018)	(0.001)	(0.0004)
Other charitable spending	0.0051	-0.0005	-0.00013
	(0.0008)***	(0.001)	(0.0003)
Hospital spending	0.0008	-0.0012	-0.0009
	(0.0023)	(0.001)	(0.0005)
Education spending	0.0034	0.0007	-0.0001
	(0.0007)***	(0.0002)**	(0.0002)
City income variables			
Manufacturing wages	0.000008	0.0000002	0.000003
	(0.00001)	(0.000002)	(0.0000015)*
# of workers in heavy industry	-0.0001	0.00001	0.00001
	(0.00003)	(0.00002)	(0.000004)**
# of tax returns filed	1.7095	0.5803	0.6192
	(1.7192)	(0.4302)	(0.2226)**
Surrounding county demographics			
Percent black	-3.4510	-0.5199	0.2842
	(1.5654)*	(1.38)	(0.3834)
Percent illiterate	17.9793	9.0506	0.5026
	(3.2804)***	(3.3485)**	(1.2740)
Percent rural	-1.8852	-0.7204	-0.6484
	(1.2792)	(0.48)	(0.3045)*
Percent foreign born	-0.6760	0.1639	-0.3138
	(1.1520)	(0.75)	(0.1579)*
State weather variables			
Avg. yearly temp.	-0.0054	-0.0058	-0.0002
	(0.0093)	(0.01)	(0.0018)
Lag Mths of severe wet	0.0075	0.0013	0.0006
	(0.0052)	(0.002)	(0.0015)
Lag Mths of severe drght	0.0081	0.0033	0.0010
	(0.00275)**	(0.002)	(0.0010)
Other variables			
Mortality rate for adults aged 20-29	0.0147	0.0051	0.0058
	(0.0293)	(0.012)	(0.0045)
Constant	-0.1008	0.2780	0.1929
	(0.4340)	(0.368)	(0.1687)
Observations	603	603	603
Adjusted R-squared	0.8958	0.7922	0.589

Robust standard errors in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 1.3: City-Specific Mortality Trend Model

	Age 0-1	Age 1-4	Age 5-9
City spending variables of interest			
Lagged PHE	-0.0021	0.0016	0.0063
	(0.0048)	(0.006)	(0.0022)**
Poverty Relief	-0.001416	0.0001	0.0005
	(0.00069)*	(0.0003)	(0.00017)**
Women's suff. ("Pre 1914" omitted)			
1915-1919	-0.4811	-0.1215	0.0705
	(0.3632)	(0.136)	(0.0485)
1920	0.9995	0.5588	0.0691
	(0.2861)***	(0.22317)**	(0.0782)
Other spending variables			
Other health spending	0.0010	-0.0002	-0.0020
	(0.0042)	(0.003)	(0.00088)*
Sanitation spending	-0.0005	0.0018	0.0003
	(0.0011)	(0.0027)	(0.0006)
Other charitable spending	0.0120	-0.0003	-0.0007
	(0.0035)***	(0.002)	(0.0008)
Hospital spending	-0.0019	-0.0016	-0.0010
	(0.0021)	(0.0012)	(0.0011)
Education spending	0.0013	-0.00005	0.0002
	(0.00058)*	(0.001)	(0.0003)
City income variables			
Manufacturing wages	0.000031	0.00001	0.00001
	(0.000005)***	(0.00001)	(0.0000026)*
# workers in hvy ind.	0.000013	0.00001	-0.000002
	(0.00003)	(0.00001)	(0.00001)
# of tax returns filed	-1.0957	1.0882	0.5920
	(1.8130)	(0.5574)*	(0.2274)**
State weather variables			
Avg. yearly temperature	-0.0114	-0.0113	0.0007
	(0.0140)	(0.008)	(0.0023)
Lag mths of severe wet	0.0071	0.0002	-0.0001
	(0.0058)	(0.0027)	(0.0014)
Lag mths of severe drght	0.0061	0.0007	0.0006
	(0.0031)*	(0.002)	(0.0010)
Constant	1.8208	0.9903	0.0015
	(0.8783)*	(0.30476)**	(0.1152)
Observations	603	603	603
Adj R^2	0.9211	0.7993	0.6215

Robust standard errors in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

CHAPTER 2

THE EFFECTS OF WEATHER SHOCKS ON CROP PRICES IN UNFETTERED MARKETS: THE UNITED STATES PRIOR TO THE FARM PROGRAMS, 1895-1932

2.1 Introduction

Recently much attention has been given to studying the effects of human contributions toward an increasingly warmer, wetter and more variable climate. If we assume that the climate change scenarios are correct, it is important to determine the economic implications of not only global increases in temperature and precipitation, but also how local variation in weather may affect productive activities. Changes in temperature and precipitation, and weather disasters like droughts, floods, heat waves, and blizzards, have direct effects on crop yields and the vitality of farm animals. These weather events may affect the prices that farmers receive for their products, thus also farm incomes and land values. The effects of localized supply shocks on the prices of agricultural commodities when they leave the farm, or their farm gate prices, will differ across the types of commodities. Local weather shocks may have minimal influence on the local prices of crops sold in international markets. However, for crops primarily used and sold at the local level, such disasters may lead to significant local price responses.[1]

During the World Trade Organization negotiations, the rest of the world has been pressuring the United States and Europe to cease interfering with agricultural markets while trying to support their domestic farmers. Most modern studies of agricultural price responses to weather shocks have been focused on these heavily regulated markets, but such studies provide little information on how unfettered

[1] This could be due to tariffs or other trade barriers, or alternatively is simply a result of the inherent properties of the good (i.e. the good is heavy, bulky, or does not store well).

markets, which might arise out of the trade negotiations, will operate.

Our goal is to examine the sensitivity of agricultural prices and output to local and non-local weather fluctuations over a large span of time in the United States prior to 1932, when markets were relatively unfettered by farm programs. In this paper, we examine the United States' three great staple crops (cotton, corn and wheat), as well as hay. Cotton and wheat are crops with high value-to-weight ratios, and ones that are not heavily used in other agricultural productive activities. During the period of consideration, cotton prices, adjusted to 1982-84 dollars, averaged about 113 cents per pound. Wheat prices averaged about 15 cents per pound in 1982-84 dollars. Corn and hay, on the other hand, are both used as feed for livestock and have value-to-weight ratios less than that of cotton or wheat. For comparison, between 1895 and 1932, corn averaged about 9 cents per pound in 1982-84 dollars, and hay about 5 cents per pound in the same denomination. Both corn and hay are primarily used and sold at the local level. For example, at the beginning of our period of consideration, over 77 percent of corn was retained and consumed in the county of production, whereas only about 40 percent of wheat, and only a negligible fraction of cotton was consumed in the locality where it was grown.

We expect that when agricultural commodities have high transportation costs and are heavily used in productive activities at the local level, prices will be sensitive to changes in local weather. Conversely, for commodities sold in non-local markets, prices will be affected much less by changes in local weather but will be sensitive to geographically broad changes in weather conditions. While we do not explicitly estimate the relationship between transportation costs and price volatility, by looking at the differences between two crops with relatively high value-to-weight ratios and two crops with relatively low value-to-weight ratios, we can explore how the reduction of transactions costs through globalization might potentially mitigate the effects of localized weather shocks.

The paper proceeds as follows. In section 2 we describe how we measure weather. Prior papers generally determine weather through temperature and precipitation measured at different points of the year; we use temperature and precipitation, but also include variables derived from the Palmer Drought Index to capture the dramatic events of long, sustained periods of drought or wetness. Section 3 describes the data sources for the weather and livestock information, and describes summary statistics for the different farm and weather variables included in the analysis. In Section 4 we discuss how price effects should differ between commodities that are relatively more and less tradable and how the presence of international markets reduces the sensitivity of prices to local weather shocks. We then outline the model used to estimate the effect of weather on farmgate prices and output in section 5, and talk about the several different methods used to estimate the weather effects. A discussion of the results from these models is given in Section 6 for each of the different commodities. In Section 7 we specify and estimate a model that attempts to control for farmer expectations and try to determine the effects of unexpected shocks. Section 8 concludes.

2.2 Local weather and drought

Between 1895 and 1932 there was a great deal of variation in weather conditions, yields on different harvests, and commodity prices. In his 1926 *Business Annals*, published for the National Bureau of Economic Research, Willard Thorp assembled narrative information from commercial sources on the success or failures of the cotton, corn and wheat harvests over the previous century, as well as data on crop prices. Within our period, he reported multiple instances of record crop harvests, as well as several years of failures. For example, 1921 witnessed poor harvests of both corn and wheat, and a failure of the cotton harvest, leading to price increases even in the midst of a recession (Thorp 1926, pp. 144). As predicted from a supply-driven equilibrium model, good harvests generally led to lower prices and bad harvests to higher prices. Some crops hewed more closely to this relationship than others. Inspection of Thorp's annals reveals that when the cotton harvest was poor, short or failed, cotton prices were always listed as rising or being high. This relationship was only generally true for corn and wheat.

Many different reasons contributed to the good and bad harvests reported by Thorp in his business annals. Defective seed, pests, and disease played a small role in reducing yields. As has been well documented, the boll weevil, which entered the United States around 1892, was particularly devastating to cotton harvests in the South (Lange, Olmstead, Rhode, 2009). Table 2.1 reports USDA estimates of reduction of cotton yield by cause between 1909 and 1932. As was often the case with plant diseases and other insect and animal pests, the effects of the weevil was tied to weather conditions. Specifically, weevil damage was worse in wet, warm years.

Even though most people have focused on the destruction wrought by the boll weevil, drought and other weather fluctuations caused more crop losses than did that nasty pest (USDA Yearbook 1922, Kramer 1983). Droughts, floods, hot winds and other climatic shocks destroyed numerous harvests across the United States. Heavy rain led to multiple devastating floods on the Ohio, Missouri and Mississippi basins, ruining harvests and destroying farmland that sometimes took several years to recover. In the Mississippi Basin, there were twelve separate instances of major flooding, culminating in 1927 with the famous flood that led to the Flood Control Act of 1928 (Trotter et. al 1998).[2]

Drought was also a major problem. According to the USDA estimates, drought conditions in 1911 destroyed a quarter of the corn, wheat and hay harvests. The drought eliminated 35 pounds per acre of the cotton harvest, a loss that is 50 percent higher than the amount of cotton destroyed by the boll weevil in any of the surrounding four years (USDA Yearbook 1922). Deficient moisture conditions severely curtailed the corn harvests again in 1916 and 1918, and in 1917 frost reduced the corn and wheat yields by nearly 14 and 12 percent, respectively (USDA Yearbook 1922). The droughts in the late 1910s also contributed to an outbreak of stem rust which devastated the spring wheat crop in Northern Plains. The Great Drought of 1930, which USDA Secretary Arthur M. Hyde called "the worst drought ever recorded in this country," served as a prelude to the infamous Dust Bowl conditions

[2]These were 1903, 1907, 1908, 1912, 1913, 1916, 1920, 1922, 1923, 1927, 1929, and 1932.

of the 1930s (Hamilton 1982, p. 850).

Given the importance of weather in determining the strength of the cotton, corn and wheat harvests, it is not surprising that so many studies have been devoted to determining the relationship between temperature, precipitation and crop yields. Just within our sample years, Annie Hannan counted over 2,000 studies that examined the influence of weather on crops (Hannan 1932).

Standard practice in the agronomic literature is to measure weather through a combination of both linear and non-linear effects on crop yields. Because both very low and very high levels of precipitation and temperature adversely affect crop yield, assuming a simple linear relationship between weather and prices and subsequently profit, would be a misspecification. Generally, the non-linearity introduced is a quadratic in precipitation and temperature. We choose a similar path, but also include measures for drought and wetness conditions using the Palmer Drought Severity Index (PDSI).[3]

2.3 Data and summary statistics

To study the effects of state-level weather fluctuations on farm-gate prices, we combine two existing datasets; one containing historical crop information, the other historical weather information. State-level information on yield, harvests, and prices for commodities produced and sold across the United States was taken from the Agricultural Time Series-Cross Section Dataset (ATICS), compiled from USDA records by Thomas Cooley, Stephen DeCanio and M. Scott Matthews.[4] This dataset covers the contiguous United States from 1866 to 1969, although the limited availability of weather data constrained the analysis to the period after 1895. After 1932, the federal government intervened heavily in many agricultural markets, with payments to limit acreage under production, price supports for some agricultural commodities, and the formation of federal crop insurance in 1938. This new legislation, changes

[3]We define the term "drought" as a prolonged and abnormal moisture deficiency.

[4]Crop prices represent the farm gate price on December 1st. The data are freely available from the National Agricultural Statistics Service

in technology, increases in the use of fertilizer, and other factors affecting the relationship between weather and crop yields, caused a variety of effects that have yet to be sorted out. To isolate the activities of relatively unregulated markets, we limit the analysis to the 37 years between 1895 and 1932.

To adjust for inflation, commodity prices are adjusted to reflect 1982-1984 dollars using the CPI series developed by Lawrence Officer. Table 2.2a gives summary statistics on output and prices per pound in 1982-1984 dollars for the selected farm commodities, and Table 2.2b the summary statistics for the different weather variables used in estimation.

There was great temporal and spatial variation in both output and prices for all four of the crops; second, there is also large variation between the commodities. Within any one year cotton and wheat prices displayed much less regional variation than prices for corn and hay. This is likely a function of two key differences. Corn and hay were used as animal feed on the farm and in local markets, while cotton was used primarily as an input for manufacturing in U.S. and English cities and wheat was marketed internationally. Further, cotton was much less costly to transport than the other crops. The differences in markets lead us to believe that state level corn and hay prices will fluctuate more with state level weather shocks than will wheat and cotton prices.

Map 1 shows the distribution of production across the contiguous United States for the four crops in 1929, a year relatively free of inclement weather. Both the cotton, corn and wheat belts are clearly visible. Also evident is that corn and hay production is more widely distributed across the country, consistent with the existence of local markets for these commodities. Every state engaged in at least some production of corn and hay, whereas cotton was limited to the more southern latitudes. Wheat was concentrated in the Midwest, with no production occurring in Florida, Mississippi, Alabama, New Hampshire, Massachusetts, Rhode Island or Connecticut.

Monthly data on temperature, precipitation and the Palmer drought measures were compiled by the National Climatic Data Center. To correct for biases in raw weather data that arose from different measurement times across the stations, the average temperature and precipitation data are adjusted for time of day using the model suggested by Karl et. al (1986).

Because the agricultural data are measured annually and the planting and harvesting dates differ among states, for simplicity we convert all weather variables to yearly averages. Summary statistics for average temperature, average precipitation, months of extreme or severe drought, months of extreme or severe wetness, and the Palmer Z standard deviation are given in Table 2.2b. For the period under consideration, these series are stationary in the time series sense.[5]

The weather in most states producing cotton tended to be warmer and wetter than in states producing corn, hay and wheat. Temperatures also displayed less variability in the cotton states.[6] In all of the states, average yearly precipitation, which is an average of average monthly precipitation from January to December, ranges from about a third of an inch per month to just over 6 inches per month. Temperature is averaged similarly to precipitation, and also represents a twelve month average of the January to December monthly averages. In our sample, this ranges from about 35 Fahrenheit to about 74 Fahrenheit in the corn, hay and wheat producing states and from about 50 Fahrenheit to about 74 Fahrenheit in the cotton producing states. Months of extreme or severe drought and extreme or severe wetness are calculated using a form of the Palmer Drought Severity Index, the Palmer Hydrological Drought Index (PHDI).[7] Between 1895 and 1932, some states endured years with serious drought conditions for all twelve months, while other states enjoyed years with twelve months straight of normal moisture levels. The number of months of extreme or severe wetness ranged similarly.

[5]This is tested using the Fisher test suggested by Maddala and Wu (1999).

[6]During the period under consideration, arid western states began producing cotton using irrigation.

[7]The PHDI represents about a year's worth of abnormal moisture conditions, while the PDSI represents about nine months worth.

2.4 Price effects and transportation costs

State-level price responses from state-level supply shocks will be mitigated for tradable goods sold in an international market because individual states will play too small a role to affect prices significantly. At the extreme, if a farm commodity is perfectly tradable internationally, the observed state average farm-gate (local) price is a function only of non-local, or international supply and demand and farm-gate prices vary only with changes affecting the aggregate market.[8] Conditional on weather that affects producers as a whole, state farm-gate commodity prices would not be influenced much by state-level weather shocks. Conversely, for perfectly non-tradable commodities with prohibitively high transportation or storage costs, the observed state farm-gate price is a function only of supply and demand in the local market. If a severe drought hits, local traders do not import goods from other states or countries to mitigate the price shock. Additionally, weather affecting producers outside of the local area will not affect local prices.

Commodities generally fall in between these two extreme cases. Assuming tradability is only a function of the specific properties for a certain crop and it does not vary temporally or spatially, we follow the setup of Mundlak and Larson (1992) and write the observed logged price of commodity c in state s during year t as function of both the international and local supply and demand:

$$ln(P_{s,c,t}^{obs}) = \tau_c ln(P_{c,t}^{int}) + (1-\tau_c) ln(P_{s,c,t}^{loc}) \tag{2.1}$$

where τ_c is an index of the strength of the local market for crop c. The strength of the market itself is a function of transaction costs and local uses for the crop. Crops such as cotton and wheat, which are easy to store, easy to transport, have a relatively high value-to-weight ratio and few local uses, have a τ_c closer to one. On the other hand, crops such as corn and hay, which are used in other areas of agricultural production, have higher transport costs, and stronger local markets, have a τ_c much closer to zero.

[8]This statement assumes that transport costs between the farm-gate and markets stay constant over time.

Figures 2.1 and 2.2 show the differential effects of negative local supply shocks for goods with relatively high and low transport costs. In each of these figures, local supply is upward sloping, relatively inelastic and given by $S1$ and $S2$. Also present in each is the internationally determined prices P_{buy}^{int}, the price paid by international buyers and P_{sell}^{int}, the effective price received by commodity producers after subtracting their marginal cost of bringing the commodity to the international market.[9] The distance between P_{buy}^{int} and P_{sell}^{int} is C_i^{int}, the size of the marginal cost to bring commodity i to the international market. Within each of the figures, the bolded line represents the effective market demand curve faced by suppliers in the local area, and the bolded dashed line represents the relevant areas of the market supply curves for local suppliers.

Figure 2.1 represents the local market for commodities such as corn and hay that have higher transport costs and strong local markets. Because the cost of bringing these types of goods to international market is high compared to a good such as cotton, fluctuations in the local market play a larger role. Additionally, fluctuations in the international market will play a relatively smaller role. In Figure 2.1, producers begin in an exporting situation on supply curve $S1_L$, where the market quantity Q_b is the amount sold at the local level, and the quantity $Q_a - Q_b$ is exported to the outside market. Because the marginal transport cost C_i^{int} is large, very little of the good is exported, and a relatively small supply shock would cause local producers to sell exclusively to the local market. For example, an adverse weather shock that causes supply to fall from $S1_L$ to $S2_L$ drives the quantity produced from Q_a to Q_b and cuases the price at which the farmers sell to rise from P_a to P_b. If the supply reduction were particularly severe, it could potentially drive the market price above P_{buy}^{int} and cause local markets to import from the outside market.

The effect of the international market on the local market is also mitigated. Changes in price determined by the international market, or P_{buy}^{int}, will affect local market prices if it pulls the "band" between P_{buy}^{int} and P_{sell}^{int} in Figure 2.1 above or below the local equilibrium price. Local prices tend to be less susceptible to changes

[9]If we assume constant marginal transaction costs, this is also the average cost

in the international market for goods with higher transport costs and wider bands.

Figure 2.2 representes a local market for commodities such as cotton that have low transportation costs and are not generally used and sold at the local level. In this case, the costs associated with bringing the good to the international market, C_i^{int}, are small. We again begin in an exporting situation with a local market price at P_a, total production of Q_a, and local purchases of Q_c. In this situation, however, the production exported to outside markets $(Q_a - Q_b)$ is much larger. It takes a much larger adverse supply shock to force the local market to rely exclusively on local production. In the event of a supply shock that moves the local supply curve to $S2_L$, to price shock is much lower than in Figure 2.1, as the price rises much less from P_a to P_b. Because of the lower transport costs, a local supply shock would be more likely to lead to a situation where local consumers pay the international price paid by buyers.

Considering a simplified version for farm commodity prices where weather is the only input, then

$$ln(P_{s,c,t}) = \tau_c ln(P(Q_{c,t}^{int}(OPW_{s,t}))) + (1 - \tau_c) ln(P(Q_{s,c,t}(w_{s,t}))) \tag{2.2}$$

where Q^{int} is the international level of the commodity sold, $w_{s,t}$ is a measure of weather conditions in state s and year t, and OPW is a measure of weather conditions across all other states producing the commodity, defined in Section 5.1 below.

2.5 Empirical model

To test the significance and magnitude of the relationship between farm commodity prices and adverse weather, we use the year-to-year variations in both weather and commodity prices to specify a regression model including state and year fixed effects. Including these fixed effects will net out much of the unobserved variable bias that seems to plague the cross-sectional models present in much of the agronomic literature (Deschenes and Greenstone 2007). In this way, we can look at the entire United States, instead of, for instance, limiting our scope to non-irrigated counties or to states that were net exporters of the different crops. Additionally, we can conduct the analysis without worry that states in different climate zones may have different levels of transportation structure.

Local weather is measured using time-bias corrected temperature and precipitation, their squared terms, the number of months of extreme or severe drought, the number of months of extreme or severe wetness, and moisture variation. The last three variables are derived from the Palmer Z Index. In addition to the effect of local weather on local prices, we are also interested in the effect of weather fluctuations by other producers who are competing in the national and international market. For this reason, we construct a set of "Other Producer's Weather" (OPW) variables.

2.5.1 Other producer's weather

For goods with weak local markets and sold primarily as exports, geographically broad changes in weather conditions affecting the aggregate market play the dominant role in determining local prices. To capture the effect of weather-driven supply shocks in the outside market, we create a set of variables that measure changes in weather affecting all other producing states in the U.S.

If the United States economy was completely closed and trade occurred only between the different states, including these variables would completely capture the effect of an international market. However, while the U.S. was not a closed economy and cotton and wheat were being bought and sold in a true international market, we argue that including aggregate measures of weather affecting domestic producers will proxy well for shocks to the entire market outside the local state market. This proxy works well because the United States is a sufficiently large portion of the overall international supply for each of the different crops.

We construct these OPW variables for each of the four commodities. These weather variables are weighted by each state's relative share of national production in 1929, which is chosen as the weighting reference year due to its near absence of inclement weather. These variables control for the effect on local farm-gate prices from weather shocks affecting other producers of the commodity.

To create the weights, we first calculate the national share of production within state s for commodity c in 1929. This is represented by $\eta_{s,c}$ and defined as:

$$\eta_{s,c} = \frac{Q_{s,c}}{\sum_{j=1}^{S} Q_{j,c}} \tag{2.3}$$

where $Q_{s,c}$ is the total output of commodity c produced by state s in 1929.

We then use this to construct the weighted averages of the different weather variables (OPW). For weather variable W, OPW is defined as:

$$OPW^{W}_{s,c,t} = \frac{1}{1-\eta_{s,c}} \sum_{j \neq s}^{S} \eta_{j,c} * W_{j,t} \tag{2.4}$$

In the above equation, W could be average yearly precipitation in state j and year t, average temperature in state j and year t, or one of the other weather variables included in the analysis.

2.5.2 Reduced form models

For each time period, we estimate the following reduced form models for price and quantity:

$$ln(P_{s,c,t}) = \alpha_s + \gamma_t + W_{s,t}\beta_c + OPW_{s,c,t}\omega_c + \epsilon_{1,s,c,t} \tag{2.5}$$

$$ln(Q_{s,c,t}) = \alpha_s + \gamma_t + W_{s,t}\beta_c + OPW_{s,c,t}\omega_c + \epsilon_{2,s,c,t} \tag{2.6}$$

where $ln(P_{s,c,t})$ is the logged real price for commodity c in year t and state s, $ln(Q_{s,c,t})$ is the logged quantity,[10] α_s is a set of state fixed effects that control for unmeasured time-invariant determinants of the farm-gate price, γ_t is a set of year indicators that control for annual shocks common to all states, $W_{s,t}$ is a vector of weather variables in year t and state s that could potentially affect local prices, and $OPW_{s,c,t}$ is the vector of weather variables in year t for other producers of commodity c outside state s. The disturbance terms $\epsilon_{1,s,c,t}$ and $\epsilon_{2,s,c,t}$ are assumed to have conditional mean zero and defined as the other factors influencing farmgate prices and output besides weather.

[10] After examination of the different output distributions, we concluded that they were all closer to a log-normal distribution than a normal distribution

Although there are certainly other factors that could potentially affect farm commodity prices, after controlling for fixed effects, it is not likely that these unobserved effects will cause the local weather variables to be correlated with the error term. While the variables that proxy for weather fluctuations in other producing states are weighted by that state's share of national production, it is also unlikely that the OPW variables will be correlated with the error term. The share of production used to weight the weather in the other states is fixed in 1929, and thus cannot vary over time in response to weather shocks. Any influence of the production share in 1929 on the error term will be controlled with the state fixed effects.

The dependent variables in the analysis are the logged values of the real prices and quantities for the different commodities. Corn and hay had higher transport costs and were more commonly used in agricultural production than cotton and wheat. Therefore, we expect that the state farm-gate prices of corn and hay were more responsive to adverse local weather shocks than were the state farm-gate prices of cotton and wheat. We might expect all of these commodities to experience changes in prices due to fluctuations in the weather in the rest of the nation. Such fluctuations will influence the placement of the upper and lower prices in the price band created by transportation and transactions costs. How much the weather outside the state will matter is an empirical question.

2.6 Results

Tables 2.3a-d show regression results with the logged prices of cotton and corn as the dependent variables (a-cotton, b-corn, c-hay, and d-wheat) and Tables 2.4a-d give regression results with logged quantity as the dependent variable. All of the models present in Tables 2.3 and 2.4 include state and year fixed effects.[11] Column 1 in each table presents results from measuring weather using just temperature, precipitation and their squared terms. Including this basic model allows comparison

[11] The inclusion of the fixed effects cause the to be so close to one. Without the fixed effects the R-squared ranges from 0.15 to 0.35 across the specifications.

to the prior work that used only measures of temperature and precipitation and sets a baseline for comparison when the additional weather variables are included. Column 2 includes the variables controlling for the number of months of extreme or severe wetness and extreme or severe drought. For all of the different crops, including the extreme or severe wetness and drought measures did not affect the coefficients on average yearly temperature or its squared term. Their inclusion tended to slightly attenuate the coefficients for precipitation and its squared term because these variables are measures of the extreme parts of the distribution of drought and wetness that arises from changes in current and prior precipitation.

Columns 3-5 represent the different models that include the *OPW* variables. Column 3 includes just temperature, precipitation, and their squared terms. Column 4 adds in the number of months of extreme or severe drought and wetness, and Column 5 includes the standard deviation of the Palmer Z index to control for effects of changes in weather variability. Tables 2.4a-d show the results when the logged quantity within the state is estimated as a function of the weather variables.

The different crops exhibited different sensitivities to local and non-local weather events, although there were some commonalities across the tables. In comparisons of specifications for a crop, the coefficient estimates tended to be similar across the different model specifications. The inclusion of the variables and the Palmer drought and wetness measures had little effect on the coefficient estimates for local average temperature and average precipitation.

2.6.1 The dominant shifts in supply or demand associated with weather changes

Our analysis of local supply and demand adjustments in Figures 2.1 and 2.2 focuses on supply shifts because in most cases they are the dominant shifts associated with local weather changes. For corn and hay, where narrative evidence suggests a great deal of local consumption, the weather coefficients in the Table 2.3 price regressions and the Table 2.4 quantity regressions are consistent with a supply shift dominating any demand effects associated with changes in the weather. The coefficient of local temperature in the crop price equations had the opposite sign of the coefficient of local temperature in the crop output equations for nearly every crop. This was also true for local precipitation.

For example, the local temperature coefficients for corn output in Table 2.4b showed that a rise in local temperature raised corn output at a diminishing rate. The coefficients of local temperature in Table 2.3b in the corn price regressions had the opposite sign, so that increases in local temperature lowered farm-gate corn prices at a diminishing rate. The coefficients of local precipitation have similar opposing signs in the Table 2.3b price regressions and the Table 2.4b quantity regressions in Table 2.4b. Comparisons of the coefficients in the hay regressions in Tables 2.4c and 2.4b show the same opposing signs for the local weather coefficients in the price and quantity regressions. Increases in local temperature decreased hay output at a diminishing rate, while raising the farm-gate hay price at a diminishing rate. Meanwhile, increases in local precipitation lowered hay output at a diminishing rate and raised hay prices at a diminishing rate.

In the cotton and wheat markets, where transport costs were low and there was limited local consumption, the relationships of local weather with prices and quantities had the opposing signs associated with dominant supply shifts for precipitation, but not for temperature. In both the cotton and wheat markets, increases in local precipitation raised output at a diminishing rate and lowered price at a diminishing rate (see Tables 2.3a and 2.4a for cotton and 2.3d and 2.4d for wheat). On the other hand, increases in local temperature increased both prices and quantities in both the cotton and wheat markets. As we will see below, the effects of local weather on prices in the cotton and wheat markets were weak relative to the effects in the corn and hay markets, which is consistent with a setting where local conditions had little effect on prices.

Many of the same patterns arise when examining the impact of weather outside the state on the state's prices and quantities. For corn, the impact of other producer's weather on local prices (Table 2.3b) and the effect of weather on output (Table 2.4b) display the same pattern as the effects of local weather on corn prices and output. Increases in temperature raised corn output at a diminishing rate in general, and the temperature rise elsewhere was associated with declines in farm-gate corn prices at a diminishing rate. Increases in precipitation show the same patterns. Similarly, the hay market results for the impact of weather elsewhere on farm-gate prices in Table 2.3c and the general effect of weather on output in 2.4c display the exact same pattern as for the impact of local weather. More precipitation raised hay output at a diminishing rate in general, and more precipitation elsewhere lowered hay prices at a diminishing rate; higher temperatures lowered hay output at a diminishing rate and higher temperatures elsewhere raised hay prices at a diminishing rate.

The markets for cotton, described in Tables 2.3a and 2.4a, and for wheat, described in Tables 2.3d and 2.4d, again have mixed effects of weather elsewhere on output and farm-gate prices. In both the cotton and wheat markets, a rise in temperature lowered output at a diminishing rate while raising prices at diminishing rate. On the other hand, demand shifts seemed to have been more dominant with respect to changes in precipitation elsewhere. In the cotton market increases in precipitation raised output at a diminishing rate in general but increases in precipitation elsewhere also raised the state farm-gate price at a diminishing rate. In the wheat market increases in precipitation lowered output at a diminishing rate while increases in precipitation elsewhere also raised the state price.

2.6.2 Cotton prices and weather: tables 2.3a and 2.4a and figures 2.3a-2.3d

Cotton output was sensitive to fluctuations in temperature, extreme or severe wetness and changes in weather variability, as shown by the coefficients of the temperature and extreme or severe wetness variables in Table 2.4a. Despite this sensitivity of output, state cotton prices did not respond much to local weather changes. Increases in local weather variability tended to slightly decrease the state farm-gate price of cotton, but state prices were most sensitive to changes in precipitation and drought conditions in other producing states.

Figures 2.3a and 2.3b plot price response functions that show the percentage change in the state price associated with an increase of one degree Fahrenheit in local state temperature (2.3a) and in temperature elsewhere (2.3b). Figure 2.3c shows the percentage change in the state farm-gate price in response to an increase of one inch of rainfall in that state's precipitation. Figure 2.3d shows the percentage change in the state farm-gate price in response to an additional inch of average precipitation experienced by producers in the rest of the U.S. These estimates are derived from the coefficients in column 5 of Table 2.3a. We plot the relationships because the inclusion of both linear and squared terms for temperature in the log price equations causes the relationship between temperature and the price to change as the temperature rises. The same linear and squared terms are used in the precipitation measures.

In general, the results suggest that state cotton prices were not very sensitive to fluctuations in local weather. None of the local temperature or precipitation coefficients in Table 2.3a are statistically significant, and the price response functions are much flatter and closer to zero percent change than for any other crop. As seen in Figure 2.3a, the percentage change in cotton prices associated with an increase in local state temperature decreases slightly as temperature increases. Until about 60 degrees F, a 1 degree rise in annual average temperature is associated with very little change in price. As the temperature approaches the mid-70s, a one degree rise in local temperature is associated with roughly a one percent drop in state cotton prices. The other price response functions for cotton in figures 2.3b through 2.3d are flatter and even closer to zero than for local temperatures. The only statistically significant relationships between weather and prices is for precipitation in areas outside the state, but the plotted relationship in Figure 2.3d shows very weak responsiveness even at the upper ranges of precipitation. The cotton price responses to precipitation and temperature changes are much weaker than those for corn and hay described below.

2.6.3 Corn prices and weather - tables 2.3b and 2.4b and figures 2.4a-2.4d

Corn output at the state level responded to increases in temperature and precipitation in roughly the same way. The coefficients in Table 2.4b show that increases in average annual temperature led to increases in corn production at a diminishing rate, as did increases in precipitation. The local increases in corn output are associated with increases in temperature and precipitation, which themselves are also associated with reductions in local corn prices as shown in Table 2.3b. Similarly, increases in temperature and precipitation that likely would have increased output in other states, also contributed to lower corn prices in the state of interest. Our findings for the era before the powerful influences of the federal farm programs are similar in that regard to Mundlak and Larson's (1992) findings that international markets played an important role in determining local prices.

Comparisons of Figures 2.4a and 2.3a show that state farm-gate corn prices were far more responsive to local temperatures than were cotton prices. Once the average annual temperature exceeded 40 degrees, corn prices started rising in response to increases in temperature and the responsiveness rose significantly from there. Similarly, the state corn price response function for temperature changes occurring in the rest of the country (Figure 2.4b) had a much stronger positive slope than state cotton price response in Figure 2.3b. However, comparisons of Figures 2.3c with 2.4c and 2.3d with 2.4d show that state precipitation and precipitation elsewhere had much stronger impacts at higher levels of precipitation on corn prices than on cotton prices.

2.6.4 Hay prices and weather - tables 2.3c and 2.4c and figures 2.5a-2.5d

Hay, like corn, was sensitive to many of the different weather variables, both local and non-local. Local weather variables that had statistically significant coefficients included the temperature and precipitation variables, as well as the variables that proxy for extreme or severe drought and wetness conditions. The price response function to local temperature was negatively sloped for hay in Figure 2.5a. After about 54 degrees Fahrenheit, as state-level temperatures rose, prices fell. At levels below that, a 1 percent rise in average yearly temperature was associated with up to a 1 percent rise in the farm-gate price. In contrast, the relationship between hay prices and local precipitation in Figure 2.5c had a mild U shape. Local monthly precipitation had little impact on state hay prices until it reached the upper end of the range. When precipitation approached seven inches per month, an additional inch rise in precipitation led to more than a one percent rise in that state's hay price. This is nearly double the effect seen for corn in Figure 2.4c.

The price response function in Figure 2.5b for temperatures in the rest of the country shows that a one unit increase in temperature elsewhere contributed to about a one percent decrease in the state farm-gate price at every temperature level. However, none of the temperature coefficients were statistically significant, so there may have been no effect. The precipitation coefficients were statistically significant, and the path of the price response to precipitation outside the state in Figure 2.5d is nearly a direct contrast to the response path to local precipitation in Figure 2.5c. The response function for out-of-state precipitation is hump-shaped, and as the precipitation approaches an average of 7 inches per month elsewhere, a one unit increase leads to 2.2 percent reduction in hay prices in the state. This response is nearly twice as large in a negative direction in comparison to the positive response to increased local precipitation. In general, the state hay price responses to higher levels of precipitation either locally or elsewhere are much larger in magnitude than for any other crop.

2.6.5 Wheat prices and weather - tables 2.3d and 2.4d and figures 2.6a-2.6d

Wheat is sold in an international market, and as might be expected, the responses of a state's prices to local weather shocks were muted for both local temperature and precipitation. The response functions for both types of weather in figures 2.6a and 2.6c are much flatter and closer to the origin throughout the range than for hay in figures 2.5a and 2.5c and corn in figures 2.4a and 2.4c. They more closely resemble the responses seen for cotton, the other strongly international crop, in figures 2.3a and 2.3c. The slight sensitivity to local weather fluctuations is likely due to wheat being grown in many of the different states, even though it is primarily concentrated in the Dakotas and Kansas. If a local weather shock did not affect the local price too much, it would still make sense to purchase locally grown wheat at a slightly higher price. However, if local supply was hit hard, then it would not be too difficult to import from the outside market.

Wheat prices were very sensitive to temperatures in other parts of the country. The wheat price response function to temperatures elsewhere in Figure 2.6b looks very similar to the corn price response function in Figure 2.4b. A one degree rise in average temperature elsewhere as temperatures elsewhere were near 33 degrees, led to a reduction in state corn prices of nearly 2 percent. At higher temperatures, there was a much stronger response in the other direction. A rise in temperature elsewhere by one degree as temperatures elsewhere neared the high end around 70 degrees, led to an increase of state corn prices of nearly 5 percent.

The effects of precipitation elsewhere on wheat prices were also statistically significant, as seen in Table 2.6d. The response function to precipitation elsewhere looks similar to the one for hay, but the negative effect of more precipitation elsewhere at high levels of precipitation is about half the size of the one for hay.

2.7 Concluding remarks

The study of the impact of weather on crop prices in unfettered markets has become increasingly important for two reasons. First, one of the major worries associated with climate change relates to increased fluctuations in weather, which, in turn, will influence food supplies and food prices. Weather shocks that lead to reductions in output and rising food prices can have major negative effects on health as people shift their consumption to lower-priced foods, often with less nutritional quality.[12]

Second, in the World Trade Organization negotiations, lesser developed countries have been pressuring the developed nations to change their farm policies to stop interfering with markets and to stop propping up farms in the developed countries. In evaluating these changes, therefore, it is important to examine the way that unfettered crop markets work. Studies of the U.S. since the 1930s cannot illuminate much about the operation of unfettered markets because of the extensive

[12] See Komlos 1987; Haines, Craig, and Weiss 2003; and Steckel 1992. As another example, Galloway (1985) used annual data for London from 1670 to 1830 to show how bad weather and poor harvests conspired to raise both agricultural prices and mortality.

farm programs in place; therefore, we need to look at the preceding period.

We estimated the responsiveness of crop prices to both localized weather shocks, as well as weighted measures of the weather shocks experienced by other producers of each crop. Four crops, cotton, corn, hay and wheat, were chosen for study not only because they were fundamental to the U.S. economy, but also because their inherent characteristics differed in an important way. Both cotton and wheat have relatively high value-to-weight ratios, and between 1895 and 1932, were sold in a true international market. For these crops, localized weather shocks might have affected the size of the harvest within a state, but the effects on state commodity prices should have been limited. Over the 37-year period studied, that is what we find.

Corn and hay represent the flip side of that coin. Corn and hay have lower value-to-weight ratios than wheat and cotton, and thus higher transport costs. While cotton and wheat were not generally consumed locally, most of the value from corn and hay came from local uses and local agronomic activity. For corn and hay, localized weather shocks would have been expected to influence the prices in a state. Indeed, the results of the analysis show that hay and corn prices were substantially more responsive to local weather shocks than were cotton and wheat prices.

We identified the differential weather effects using the year-to-year variation in weather and commodity prices after controlling for time-invariant features of each state and controlling for national shocks such as warfare that would have influenced the markets. As a result, the analysis avoids much of the unobserved variable bias that seems to plague cross-sectional models. Furthermore, we focused on the period between 1895 and 1932 because there were not only substantial fluctuations in weather that influenced the yields of different commodities, but also because there was much less government interference in markets to protect farmers from falling prices. Despite the absence of price supports, state level cotton and wheat prices were not affected much by the weather shocks within the state.

The nationwide weather shocks and decline in prices that followed World War I helped usher in support for the federal farm programs of the New Deal. Several of these programs, such as the 1938 Federal Crop Insurance Act, were intended to address the income and production variability that the weather shocks induced. Others such as the Soil Conservation Act sought to remedy the environmental damage associated with the Dust Bowl. These federal farm programs persisted and expanded over the next 80 years and strongly influenced the ways that farm prices at the farm-gate responded to weather shocks within states and across the nation. Our next move is to investigate how the relationships between weather and prices changed as a result of these farm policies. Many of the policies were designed to diminish the downward volatility in farm gate prices and control the sale of the crop within the state and in outside markets. Such controls potentially reduced income volatility for farmers, although given the fluctuations in prices in response to local shocks for some crops in the unfettered markets, perhaps not as much as might have been thought. On the other hand, the programs may well have led to higher crop prices in the long run, with the consequent impact on health, within the United States. Understanding these tradeoffs will contribute to improvements in the quality of the policies chosen in the future.

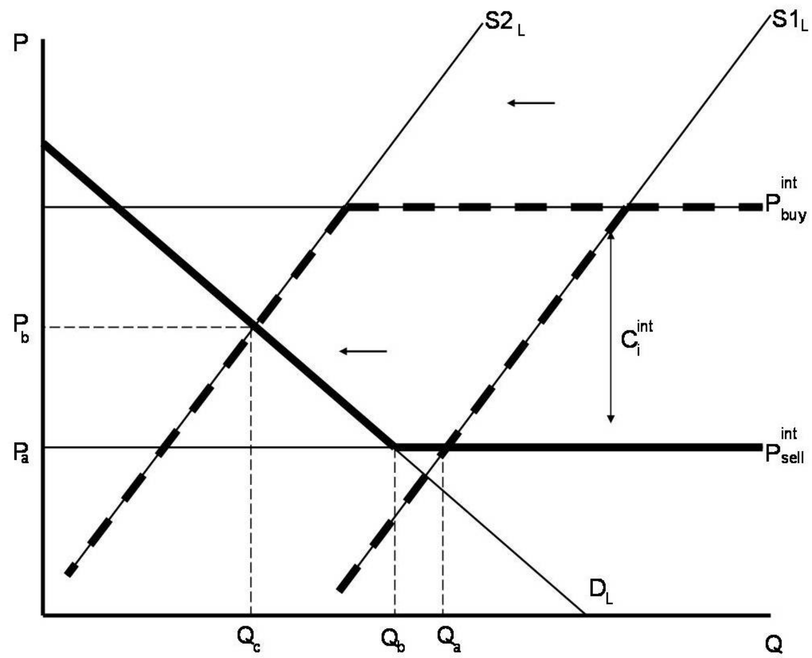


Figure 2.1: High transport costs and weak local market, adverse supply shock

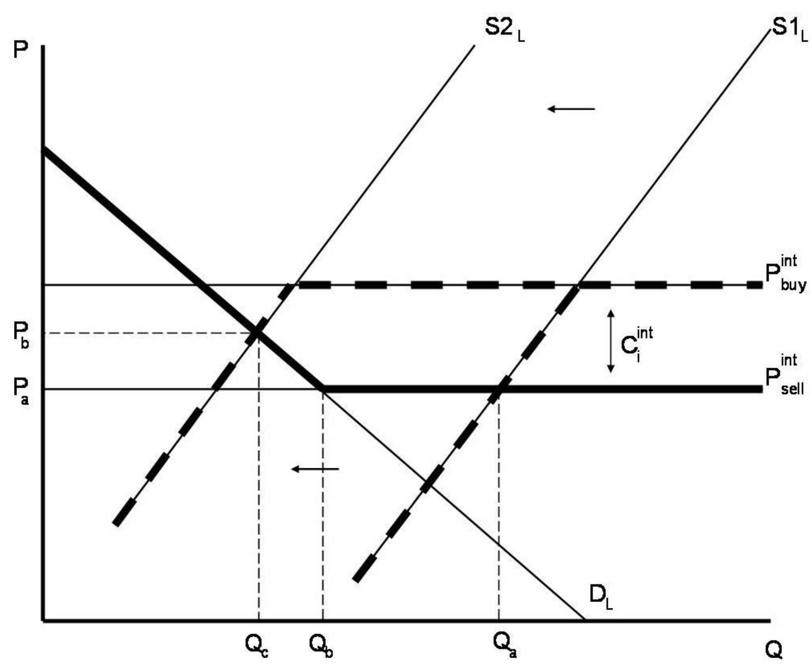
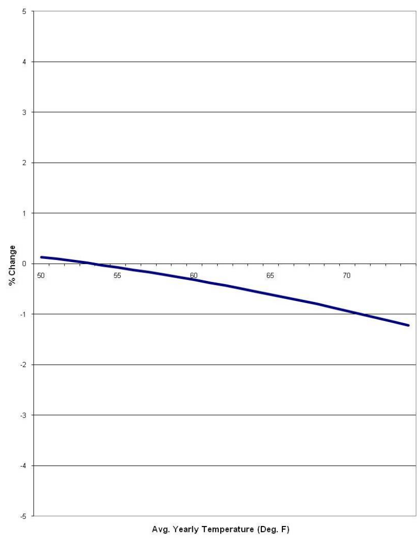
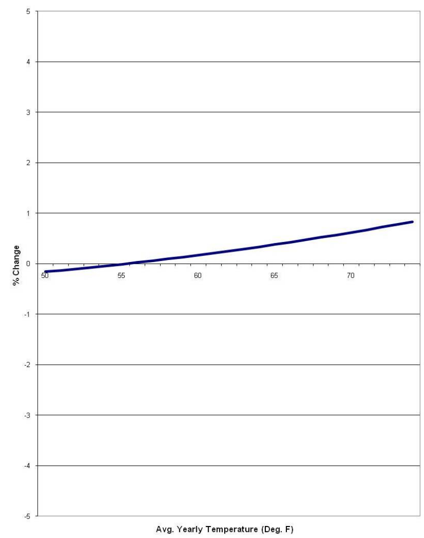


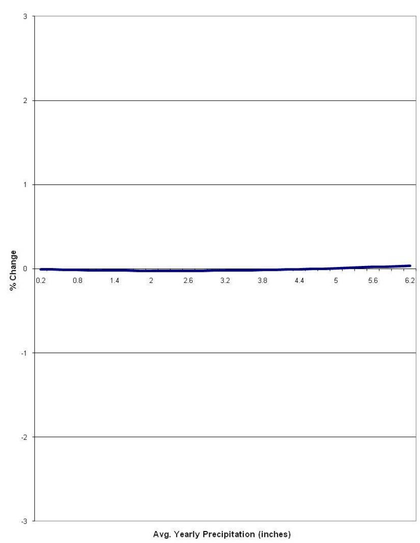
Figure 2.2: Low transport costs and strong local market, adverse supply shock



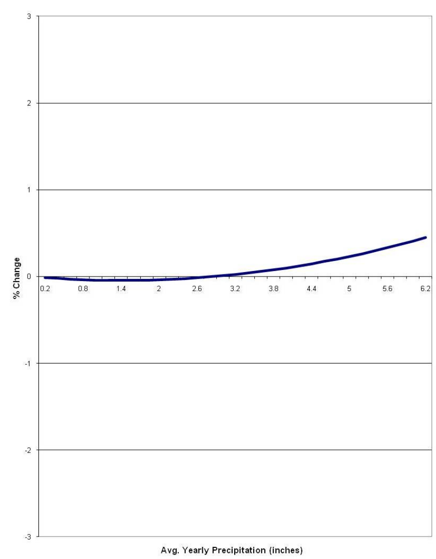
(a) $\frac{\partial \% \Delta P_{cotton}}{\partial \% \Delta Tempature_{local}}$



(b) $\frac{\partial \% \Delta P_{cotton}}{\partial \% \Delta Temperature_{OPW}}$

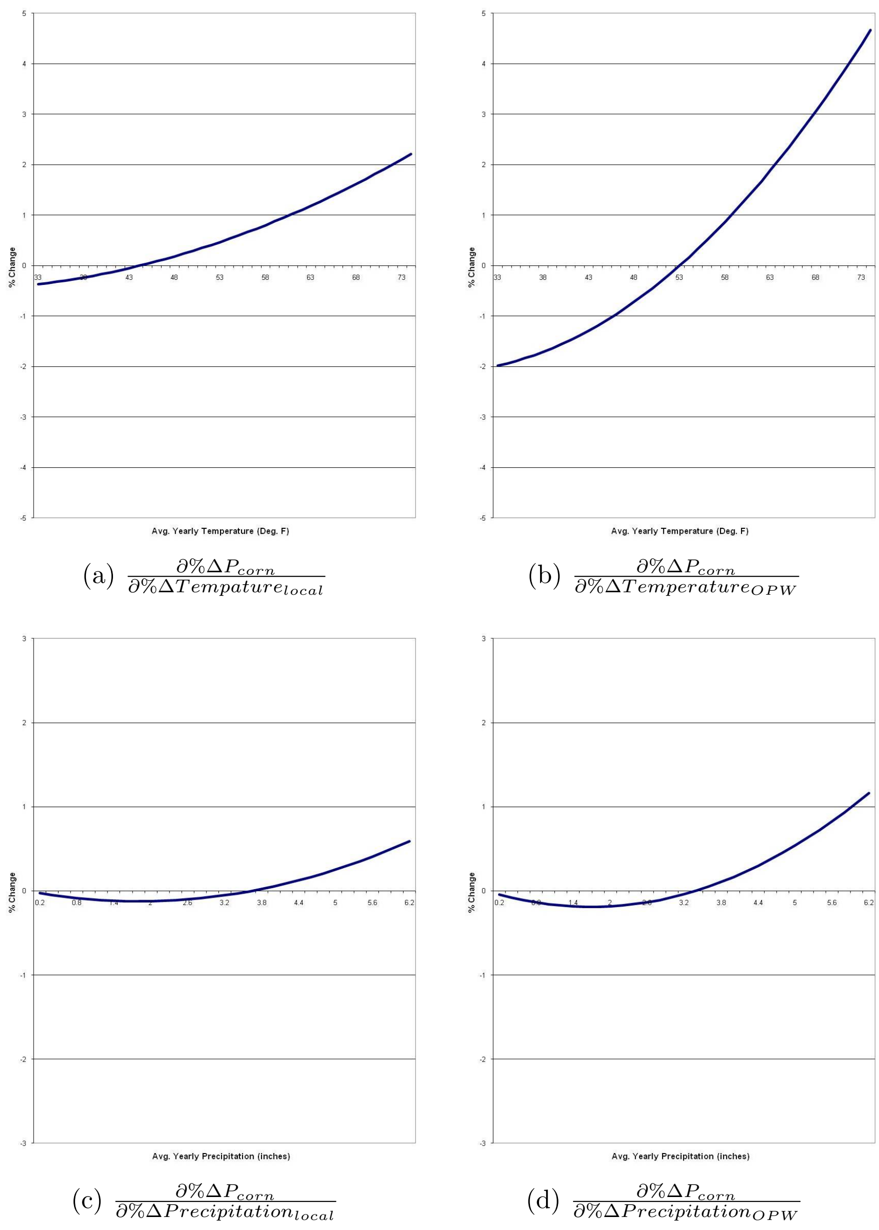


(c) $\frac{\partial \% \Delta P_{cotton}}{\partial \% \Delta Precipitation_{local}}$



(d) $\frac{\partial \% \Delta P_{cotton}}{\partial \% \Delta Precipitation_{OPW}}$

Figure 2.3: Weather effects on cotton prices



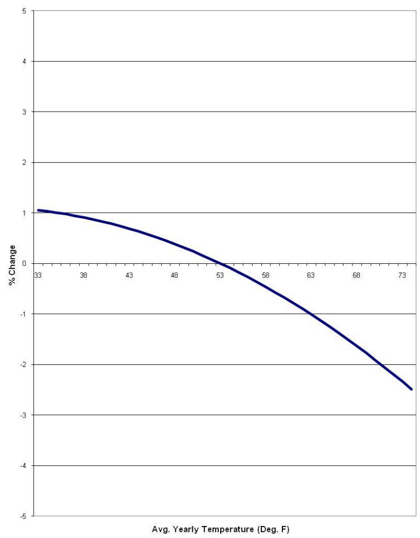
(a) $\frac{\partial \% \Delta P_{corn}}{\partial \% \Delta Tempature_{local}}$

(b) $\frac{\partial \% \Delta P_{corn}}{\partial \% \Delta Temperature_{OPW}}$

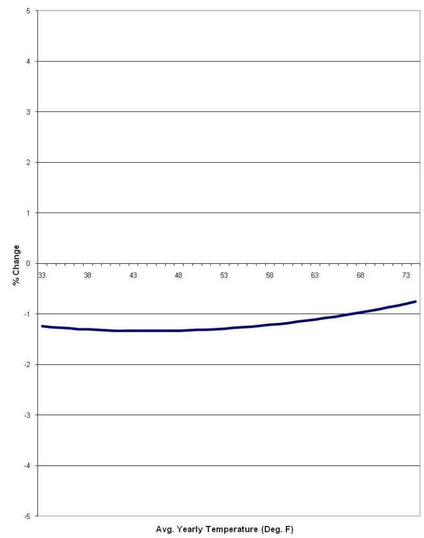
(c) $\frac{\partial \% \Delta P_{corn}}{\partial \% \Delta Precipitation_{local}}$

(d) $\frac{\partial \% \Delta P_{corn}}{\partial \% \Delta Precipitation_{OPW}}$

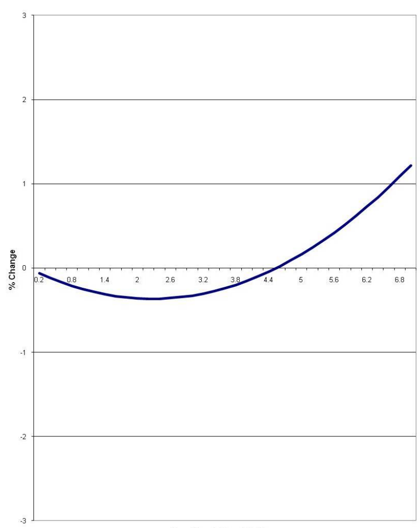
Figure 2.4: Weather effects on corn prices



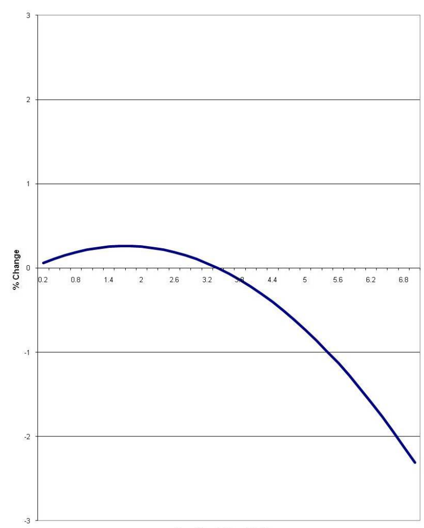
(a) $\frac{\partial \% \Delta P_{hay}}{\partial \% \Delta Tempature_{local}}$



(b) $\frac{\partial \% \Delta P_{hay}}{\partial \% \Delta Temperature_{OPW}}$

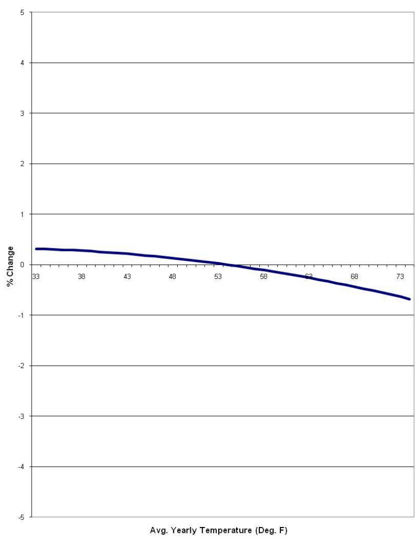


(c) $\frac{\partial \% \Delta P_{hay}}{\partial \% \Delta Precipitation_{local}}$

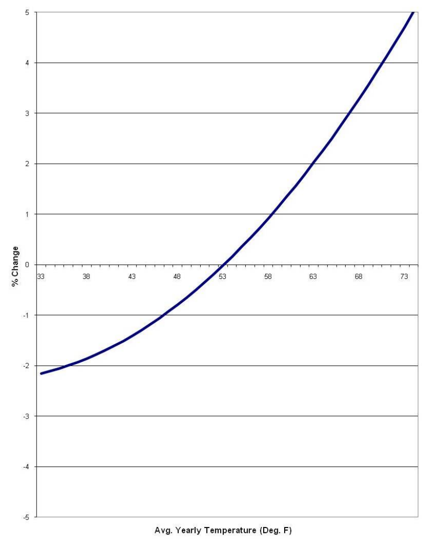


(d) $\frac{\partial \% \Delta P_{hay}}{\partial \% \Delta Precipitation_{OPW}}$

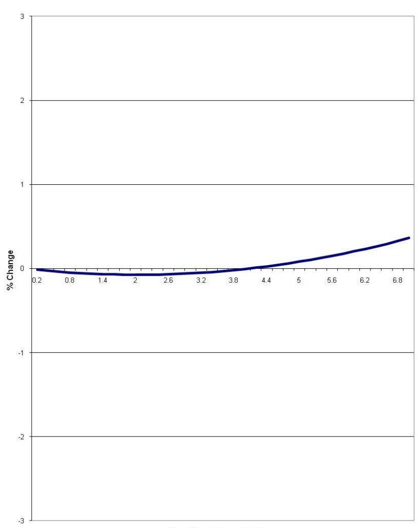
Figure 2.5: Weather effects on hay prices



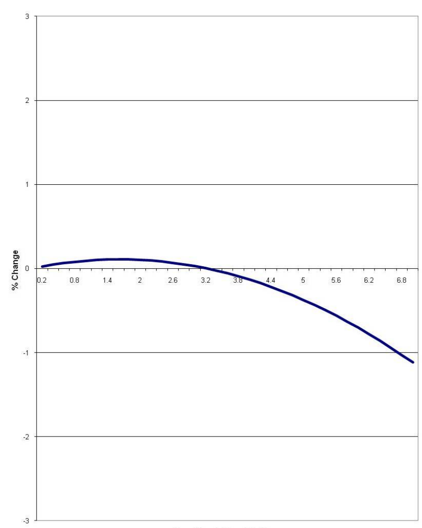
(a) $\frac{\partial \% \Delta P_{wheat}}{\partial \% \Delta Tempature_{local}}$



(b) $\frac{\partial \% \Delta P_{wheat}}{\partial \% \Delta Temperature_{OPW}}$



(c) $\frac{\partial \% \Delta P_{wheat}}{\partial \% \Delta Precipitation_{local}}$



(d) $\frac{\partial \% \Delta P_{wheat}}{\partial \% \Delta Precipitation_{OPW}}$

Figure 2.6: Weather effects on wheat prices

Table 2.1: Percent Reduction in Crop Yields by Cause: 1909-1932

Year	Deficient Moisture	Excessive Moisture	Other Climatic	Plant Disease	Boll Weevil	Other Insects
1909	14.9	6	7.7	4.2	6.1	1.8
1910	12.2	5.1	5.3	4.3	5.1	2.4
1911	9.8	2.6	3.0	0.5	1.3	6.6
1912	8.1	7.6	5.0	4.3	3.5	3
1913	15.2	2	5.9	0.5	7.5	1.4
1914	7.9	2.9	3.0	0.2	6.1	3.7
1915	6.8	5.7	6.9	1.9	10.2	2
1916	9.2	9.1	7.0	0.9	14.2	1.6
1917	15.1	1.7	8.7	1.3	8.6	3.7
1918	23.8	0.9	4.5	2	5.4	2.6
1919	2.7	15.3	3.2	1.3	13	5.8
1920	2.2	8.8	2.1	1.1	19.7	4.3
1921	8.6	4.3	3.1	1	31.2	4.2
1922	10.2	4.9	2.4	0.8	23.3	3.4
1923	7.2	8	2.8	0.7	19.2	7.4
1924	14	4.9	2.4	0.8	8.1	3.9
1925	24.6	1.4	3.0	2.1	4.1	2.2
1926	5.3	3.2	2.9	1.5	7.1	8.9
1927	6.4	4.9	2.8	1.5	18.5	4.4
1928	4.4	7.3	4.9	1.9	14.1	3.4
1929	10.8	7.2	6.0	2.3	13.3	2.5
1930	27.7	2.8	6.3	1.7	5	1.9
1931	8.3	2.6	3.5	2	8.3	1.8
1932	8	3.9	6.1	3.2	15.2	3.1
1909-1919 Avg	11.43	5.35	5.48	1.95	7.36	3.15
1920-1932 Avg	10.59	4.94	3.72	1.58	14.39	3.95

Source: 1925 USDA Yearbook of Agriculture, and the May 1928, July 1930, June 1931 and June 1934 editions of the USDA Crops and Markets publication

Table 2.2a: Statistics for Farm Commodities, 1895-1932

Crop prices (cents per lb, 1982-84 dollars)	**Mean**	**Std. Dev.**	σ/μ	**Min**	**Max**
Cotton	113.50	44.90	0.396	31.71	356.71
Corn	9.08	3.49	0.385	2.08	26.30
Hay	5.31	2.25	0.423	0.68	13.25
Wheat	15.16	5.15	0.339	3.61	39.00
Crop Output (for producing states)	**Mean**	**Std. Dev.**	σ/μ	**Min**	**Max**
Cotton (bushels)	396,235.10	453,382.70	1.144	144	2,697,848
Corn (bushels)	55,160.22	86,691.08	1.572	21	509,507
Hay (tons)	1,763.12	1,631.243	0.925	40	7,303.5
Wheat (bushels)	17,750.53	25,269.70	1.424	13	251,885

Table 2.2b: Summary Statistics for Climate Variables

Climate Variables	**Mean**	**Std. Dev.**	**Min**	**Max**
Cotton states				
Average yearly temperature	61.198	5.004	50.564	73.324
Average yearly precipitation	3.644	1.054	0.648	6.193
Months of extreme or severe drought	0.750	1.640	0	9.857143
Months of extreme or severe wet	0.926	1.914	0	11.750
Moisture index std. dev.	1.680	0.407	0.754	4.021
Corn states				
Average yearly temperature	51.813	8.029	35.495	73.324
Average yearly precipitation	2.927	1.131	0.366	6.193
Months of extreme or severe drought	1.196	2.335	0	12
Months of extreme or severe wet	0.996	2.022	0	12
Moisture index std. dev.	1.672	0.426	0.643	4.021
Hay States				
Average yearly temperature	51.907	8.073	35.495	73.324
Average yearly precipitation	2.886	1.149	0.366	5.930
Months of extreme or severe drought	1.154	2.300	0	12
Months of extreme or severe wet	1.059	2.109	0	12
Moisture index std. dev.	1.659	0.409	0.643	4.021
Wheat States				
Average yearly temperature	51.474	7.374	35.495	69.046
Average yearly precipitation	2.778	1.145	0.366	5.889
Months of extreme or severe drought	1.299	2.430	0	12
Months of extreme or severe wet	1.001	2.036	0	12
Moisture index std. dev.	1.671	0.422	0.643	4.021

Table 2.3a: Dependent variable: ln(cotton price, cents per lb in 1982-84 dollars)

Local Weather	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Avg.Temperature	0.0185	0.0183	0.0491	0.0493	0.0427
	(0.0510)	(0.0508)	(0.0640)	(0.0645)	(0.0633)
Avg. Temperature Sq.	-0.0002	-0.0002	-0.0004	-0.0004	-0.0004
	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0005)
Avg. Precipitation	-0.0142	-0.0246	-0.0319	-0.0399	-0.0197
	(0.0229)	(0.0240)	(0.0235)	(0.0253)	(0.0277)
Avg. Precipitation Sq	0.0016	0.0020	0.0037	0.0039	0.0021
	(0.0029)	(0.0028)	(0.0029)	(0.0030)	(0.0032)
Months of XS wetness		0.0035		0.0038	0.0047
		(0.0032)		(0.0033)	(0.0033)
Months of XS drought		-0.0015		-0.0006	-0.0001
		(0.0024)		(0.0023)	(0.0022)
Palmer Z Index Std. Dev.					-0.0182
					(0.0109)*
Other Producer's Weather					
Avg.Temperature			-0.0186	-0.0321	-0.0331
			(0.0214)	(0.0211)	(0.0213)
Avg. Temperature Sq.			0.0002	0.0003	0.0003
			(0.0002)	(0.0002)	(0.0002)
Avg. Precipitation			-0.0854	-0.0665	-0.0614
			(0.0277)***	(0.0270)**	(0.0296)**
Avg. Precipitation Sq			0.0130	0.0116	0.0108
			(0.0040)***	(0.0039)***	(0.0042)**
Months of XS wetness				-0.0011	-0.0007
				(0.0021)	(0.0022)
Months of XS drought				0.0040	0.0043
				(0.0019)**	(0.0019)**
Palmer Z Index Std. Dev.					-0.0038
					(0.0085)
Constant	4.1852	4.1704	5.3998	6.9183	7.1733
	(1.5478)***	(1.5423)***	(2.6576)**	(2.6464)***	(2.6305)***
Observations	559	559	559	559	559
Adjusted R-squared	0.97	0.97	0.97	0.97	0.97

Robust standard errors in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 2.3b: Dependent variable: ln(corn price, cents per lb in 1982-84 dollars)

Local Weather	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Avg.Temperature	-0.0411	-0.0421	-0.0435	-0.0437	-0.0441
	(0.0178)**	(0.0182)**	(0.0180)**	(0.0182)**	(0.0182)**
Avg. Temperature Sq.	0.0005	0.0005	0.0005	0.0005	0.0005
	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***
Avg. Precipitation	-0.1306	-0.1166	-0.1277	-0.1166	-0.1349
	(0.0311)***	(0.0336)***	(0.0312)***	(0.0334)***	(0.0351)***
Avg. Precipitation Sq	0.0180	0.0168	0.0179	0.0170	0.0186
	(0.0041)***	(0.0043)***	(0.0041)***	(0.0042)***	(0.0043)***
Months of XS wetness		-0.0019		-0.0017	-0.0022
		(0.0021)		(0.0020)	(0.0020)
Months of XS drought		0.0015		0.0018	0.0012
		(0.0018)		(0.0018)	(0.0018)
Palmer Z Index Std. Dev.					0.0164
					(1.82)*
Other Producer's Weather					
Avg.Temperature			-0.1357	-0.1545	-0.1589
			(0.0376)***	(0.0346)***	(0.0341)***
Avg. Temperature Sq.			0.0013	0.0015	0.0015
			(0.0004)***	(0.0003)***	(0.0003)***
Avg. Precipitation			-0.1154	-0.2021	-0.2240
			(0.0609)*	(0.0697)***	(0.0662)***
Avg. Precipitation Sq			0.0186	0.0303	0.0332
			(0.0091)**	(0.0101)***	(0.0095)***
Months of XS wetness				0.0177	0.0159
				(0.0050)***	(0.0054)***
Months of XS drought				0.0101	0.0088
				(0.0036)***	(0.0037)**
Palmer Z Index Std. Dev.					0.0164
					(0.0090)*
Constant	7.0815	7.1046	9.6347	10.0489	10.1610
	(0.4743)***	(0.4804)***	(0.8512)***	(0.8175)***	(0.8062)***
Observations	1796	1796	1796	1796	1796
Adjusted R-squared	0.9	0.9	0.91	0.91	0.91

Robust standard errors in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 2.3c: Dependent variable: ln(hay price, cents per lb in 1982-84 dollars)

Local Weather	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Avg.Temperature	0.0585	0.0584	0.0850	0.0870	0.0849
	(0.0254)**	(0.0255)**	(0.0298)***	(0.0297)***	(0.0302)***
Avg. Temperature Sq.	-0.0005	-0.0005	-0.0007	-0.0008	-0.0008
	(0.0002)**	(0.0002)**	(0.0003)**	(0.0003)***	(0.0002)***
Avg. Precipitation	-0.3334	-0.2749	-0.3340	-0.2734	-0.3209
	(0.0471)***	(0.0464)***	(0.0473)***	(0.0468)***	(0.0506)***
Avg. Precipitation Sq	0.0363	0.0320	0.0355	0.0311	0.0354
	(0.0064)***	(0.0062)***	(0.0065)***	(0.0064)***	(0.0068)***
Months of XS wetness		-0.0055		-0.0054	-0.0063
		(0.0033)*		(0.0034)	(0.0034)*
Months of XS drought		0.0129		0.0131	0.0118
		(0.0027)***		(0.0027)***	(0.0027)***
Palmer Z Index Std. Dev.					0.0424
					(0.0136)***
Other Producer's Weather					
Avg.Temperature			-0.0360	-0.0487	-0.0598
			(0.0852)	(0.0817)	(0.0820)
Avg. Temperature Sq.			0.0002	0.0002	0.0003
			(0.0008)	(0.0008)	(0.0008)
Avg. Precipitation			0.2584	0.3022	0.3101
			(0.1257)**	(0.1341)**	(0.1503)**
Avg. Precipitation Sq			-0.0420	-0.0444	-0.0457
			(0.0189)**	(0.0196)**	(0.0217)**
Months of XS wetness				-0.0088	-0.0080
				(0.0092)	(0.0098)
Months of XS drought				0.0090	0.0091
				(0.0082)	(0.0085)
Palmer Z Index Std. Dev.					0.0036
					(0.0415)
Constant	8.0209	7.8679	8.1451	7.4538	7.6300
	(0.6695)***	(0.6795)***	(0.8887)***	(0.8671)***	(0.8729)***
Observations	1152	1152	1152	1152	1152
Adjusted R-squared	0.87	0.87	0.87	0.88	0.88

Robust standard errors in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 2.3d: Dependent variable: ln(wheat price, cents per lb in 1982-84 dollars)

Local Weather	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Avg.Temperature	0.0158	0.0185	0.0216	0.0232	0.0248
	(0.0144)	(0.0147)	(0.0138)	(0.0143)	(0.0142)*
Avg. Temperature Sq.	-0.0001	-0.0002	-0.0002	-0.0002	-0.0002
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Avg. Precipitation	-0.1012	-0.0852	-0.0870	-0.0704	-0.0718
	(0.0249)***	(0.0253)***	(0.0248)***	(0.0249)***	(0.0253)***
Avg. Precipitation Sq	0.0124	0.0109	0.0103	0.0088	0.0088
	(0.0034)***	(0.0034)***	(0.0034)***	(0.0033)***	(0.0033)***
Months of XS wetness		0.0001		-0.0009	-0.0012
		(0.0015)		(0.0015)	(0.0015)
Months of XS drought		0.0041		0.0030	0.0028
		(0.0016)**		(0.0015)**	(0.0015)*
Palmer Z Index Std. Dev.					0.0028
					(0.0066)
Other Producer's Weather					
Avg.Temperature			-0.1745	-0.1646	-0.1721
			(0.0245)***	(0.0252)***	(0.0255)***
Avg. Temperature Sq.			0.0016	0.0016	0.0016
			(0.0002)***	(0.0002)***	(0.0002)***
Avg. Precipitation			0.1655	0.1761	0.1364
			(0.0403)***	(0.0470)***	(0.0497)***
Avg. Precipitation Sq			-0.0242	-0.0262	-0.0211
			(0.0058)***	(0.0063)***	(0.0068)***
Months of XS wetness				-0.0039	-0.0066
				(0.0034)	(0.0037)*
Months of XS drought				-0.0041	-0.0062
				(0.0025)	(0.0028)**
Palmer Z Index Std. Dev.					0.0448
					(4.12)***
Constant	6.6023	6.5575	6.3888	6.3510	6.2901
	(0.3803)***	(0.3848)***	(0.3637)***	(0.3729)***	(0.3733)***
Observations	1597	1597	1597	1597	1597
Adjusted R-squared	0.94	0.94	0.95	0.95	0.95

Robust standard errors in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 2.4a: Dependent variable: ln(cotton output)

Local Weather	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Avg.Temperature	1.2643	1.2567	1.4058	1.3664	1.3143
	(0.2333)***	(0.2302)***	(0.2657)***	(0.2591)***	(0.2568)***
Avg. Temperature Sq.	-0.0106	-0.0106	-0.0118	-0.0115	-0.0110
	(0.0019)***	(0.0019)***	(0.0022)***	(0.0021)***	(0.002)***
Avg. Precipitation	-0.1255	-0.0891	-0.1119	-0.0898	0.1281
	(0.2301)	(0.2205)	(0.2287)	(0.2157)	(0.2220)
Avg. Precipitation Sq	-0.0077	-0.0067	-0.0100	-0.0081	-0.0271
	(0.0267)	(0.0256)	(0.0265)	(0.0250)	(0.0253)
Months of XS wetness		-0.0450		-0.0499	-0.0417
		(0.017)***		(0.017)***	(0.0165)**
Months of XS drought		-0.0218		-0.0255	-0.0216
		(0.0113)*		(0.0122)**	(0.0117)*
Palmer Z Index Std. Dev.					-0.1780
					(0.062)***
Other Producer's Weather					
Avg.Temperature			0.0247	0.0982	0.0658
			(0.0850)	(0.0907)	(0.0883)
Avg. Temperature Sq.			-0.0002	-0.0009	-0.0005
			(0.0008)	(0.0009)	(0.0009)
Avg. Precipitation			0.2369	0.3501	0.3005
			(0.1357)*	(0.1557)**	(0.1563)*
Avg. Precipitation Sq			-0.0307	-0.0486	-0.0428
			(0.0192)	(0.0209)**	(0.0207)**
Months of XS wetness				-0.0255	-0.0281
				(0.0109)**	(0.0109)**
Months of XS drought				-0.0241	-0.0254
				(0.0069)***	(0.0072)***
Palmer Z Index Std. Dev.					0.0744
					(0.0468)
Constant	-24.4308	-24.2303	-34.0117	-41.4904	-36.9132
	(7.389)***	(7.2743)***	(11.3647)***	(11.3199)***	(10.902)***
Observations	559	559	559	559	559
Adjusted R-squared	0.95	0.95	0.95	0.95	0.95

Robust standard errors in parentheses

**significant at 10%; **significant at 5%; ***significant at 1%*

Table 2.4b: Dependent variable: ln(corn output)

Local Weather	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Avg.Temperature	0.2142	0.2137	0.2203	0.2193	0.2233
	(0.0701)***	(0.0689)***	(0.0714)***	(0.0703)***	(0.0699)***
Avg. Temperature Sq.	-0.0027	-0.0027	-0.0028	-0.0027	-0.0028
	(0.0006)***	(0.0006)***	(0.0006)***	(0.0006)***	(0.0006)***
Avg. Precipitation	0.2934	0.2881	0.2877	0.2818	0.3501
	(0.106)***	(0.1106)***	(0.1073)***	(0.1121)**	(0.1173)***
Avg. Precipitation Sq	-0.0396	-0.0392	-0.0390	-0.0385	-0.0442
	(0.0134)***	(0.0135)***	(0.0136)***	(0.0138)***	(0.0141)***
Months of XS wetness		0.0001		-0.0004	0.0014
		(0.0089)		(0.0089)	(0.0091)
Months of XS drought		-0.0012		-0.0017	-0.0001
		(0.0059)		(0.0059)	(0.0059)
Palmer Z Index Std. Dev.					-0.0641
					(0.0349)*
Other Producer's Weather					
Avg.Temperature			0.2039	0.2013	0.1895
			(0.0786)***	(0.0808)**	(0.081)**
Avg. Temperature Sq.			-0.0020	-0.0020	-0.0019
			(0.0008)***	(0.0008)**	(0.0008)**
Avg. Precipitation			0.2369	0.3501	0.3005
			(0.1357)*	(0.1557)**	(0.1563)*
Avg. Precipitation Sq			-0.0301	-0.0278	-0.0186
			(0.0205)	(0.0228)	(0.0229)
Months of XS wetness				0.0041	0.0003
				(0.0105)	(0.0108)
Months of XS drought				0.0020	-0.0006
				(0.0073)	(0.0075)
Palmer Z Index Std. Dev.					0.0631
					(0.0486)
Constant	6.8718	6.8787	2.9457	3.0185	2.9937
	(1.9907)***	(1.9706)***	(2.5138)	(2.5523)	(2.5529)
Observations	1806	1806	1806	1806	1806
Adjusted R-squared	0.96	0.96	0.96	0.96	0.96

Robust standard errors in parentheses

**significant at 10%; **significant at 5%; ***significant at 1%*

Table 2.4c: Dependent variable: ln(total hay output)

Local Weather	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Avg.Temperature	-0.0712	-0.0827	-0.0638	-0.0772	-0.0734
	(0.0325)**	(0.0321)**	(0.0355)*	(0.0352)**	(0.0353)**
Avg. Temperature Sq.	0.0005	0.0006	0.0004	0.0006	0.0005
	(0.0003)	(0.0003)*	(0.0004)	(0.0004)	(0.0004)
Avg. Precipitation	0.2478	0.2295	0.2476	0.2278	0.3126
	(0.0557)***	(0.05538)***	(0.0554)***	(0.0552)***	(0.058)***
Avg. Precipitation Sq	-0.0205	-0.0187	-0.0199	-0.0180	-0.0257
	(0.0076)***	(0.0074)**	(0.0076)***	(0.0074)**	(0.0074)***
Months of XS wetness		-0.0064		-0.0063	-0.0045
		(0.0035)*		(0.0035)*	(0.0034)
Months of XS drought		-0.0111		-0.0109	-0.0085
		(0.0038)***		(0.0038)***	(0.0037)**
Palmer Z Index Std. Dev.					-0.0751
					(0.0181)***
Other Producer's Weather					
Avg.Temperature			-0.1230	-0.1397	-0.1161
			(0.0965)	(0.1018)	(0.1000)
Avg. Temperature Sq.			0.0012	0.0014	0.0011
			(0.0009)	(0.0010)	(0.0010)
Avg. Precipitation			-0.1582	-0.1984	-0.1585
			(0.1600)	(0.1733)	(0.1953)
Avg. Precipitation Sq			0.0248	0.0291	0.0243
			(0.0236)	(0.0250)	(0.0279)
Months of XS wetness				0.0138	0.0158
				(0.0108)	(0.0115)
Months of XS drought				0.0015	0.0040
				(0.0088)	(0.0095)
Palmer Z Index Std. Dev.					-0.0458
					(0.0497)
Constant	7.9382	8.2315	8.3900	8.7924	8.4668
	(0.8018)***	(0.7996)***	(1.0474)***	(1.0656)***	(1.0702)***
Observations	1152	1152	1152	1152	1152
Adjusted R-squared	0.98	0.98	0.98	0.98	0.98

Robust standard errors in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 2.4d: Dependent variable: ln(total wheat output)

Local Weather	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Avg.Temperature	0.0233	0.0400	-0.0049	0.0158	0.0058
	(0.0818)	(0.0816)	(0.0819)	(0.0820)	(0.0819)
Avg. Temperature Sq.	-0.0011	-0.0012	-0.0008	-0.0010	-0.0009
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Avg. Precipitation	0.3528	0.2771	0.3667	0.2958	0.3186
	(0.144)**	(0.1575)*	(0.1451)**	(0.1546)*	(0.1565)**
Avg. Precipitation Sq	-0.0545	-0.0485	-0.0561	-0.0504	-0.0520
	(0.0216)**	(0.02248)**	(0.0216)***	(0.0219)**	(0.0218)**
Months of XS wetness		0.0172		0.0189	0.0211
		(0.0087)**		(0.0087)**	(0.00901)**
Months of XS drought		-0.0001		0.0044	0.0063
		(0.0084)		(0.0082)	(0.0082)
Palmer Z Index Std. Dev.					-0.0308
					(0.0427)
Other Producer's Weather					
Avg.Temperature			-0.2739	-0.3457	-0.2953
			(0.1389)**	(0.13595)**	(0.1401)**
Avg. Temperature Sq.			0.0029	0.0038	0.0033
			(0.0013)**	(0.0013)***	(0.0013)**
Avg. Precipitation			-0.1582	-0.1984	-0.1585
			(0.1600)	(0.1733)	(0.1953)
Avg. Precipitation Sq			0.0732	0.1521	0.1175
			(0.0339)**	(0.0381)***	(0.0417)***
Months of XS wetness				0.1023	0.1206
				(0.0188)***	(0.0195)***
Months of XS drought				0.0412	0.0552
				(0.0128)***	(0.0144)***
Palmer Z Index Std. Dev.					-0.3052
					(0.0749)***
Constant	7.1252	6.8021	7.7913	7.5424	7.9058
	(2.19021)***	(2.1839)***	(2.1939)***	(2.1957)***	(2.2029)***
Observations	1595	1595	1595	1595	1595
Adjusted R-squared	0.93	0.93	0.93	0.94	0.94

Robust standard errors in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

CHAPTER 3

THE LABOR MARKET RETURNS TO ATTENDING A FEMALE DOMINATED SCHOOL

3.1 Introduction

Recent years have seen both heightened interest in single-sex schools and a growing concentration of females in college. In 2005, the U.S. Department of Education published new regulations allowing single-sex classrooms and schools within publicly funded school districts. That same year, over 1,000 postsecondary institutions had female enrollments of greater than 95 percent (Snyder, 2005). Over the last few decades, the gender gap that was such a concern in the 1950s and 1960s has not only closed, but has reversed (Goldin, Katz, Kuziemko 2006). As the proportion of females across colleges approaches, and is expected to pass 60 percent, it is important to attempt to identify the effects that an increasingly female college composition may have on its students.

The peer effects due from attending a women's college, as well as those due to the increasing proportion of females in college, are both functions of high concentrations of females within universities. Extant literature primarily examines the effects of this on educational outcomes, both at the high school and college levels. This paper extends that approach by analyzing what happens when students leave the classroom and enter the labor market. Specifically, I estimate the effect of attending a "female dominated" school on personal wages, both at entrance and throughout tenure of the labor market.

On the face of it, it is not immediately obvious whether attending a female dominated school will lead to better labor market outcomes for the women that attend them. If there are strong, positive peer effects for women when they are surrounded with more women in their college classes and campus life, attending a female dominated school may lead to higher future wages for its female students. These higher wages could be the result of women being more likely to major in traditionally male-dominated subjects such as math and science that lead to higher wages, or they could simply be a function of increased confidence, self-worth, assertiveness or other inherent characteristics that may lead to performing better in the labor market. On the other hand, women who graduate from a female dominated school do not experience the gender diversity present in a coeducational university and may not acquire the male interaction skills their coed counterparts are more likely to learn. Since part of succeeding in a workplace environment involves successfully dealing with people of different races, cultures and sexes, they may be less likely to be promoted or more likely to be fired.

This effect is present in regards to racial diversity and seems to be the most important. Daniel, Black and Smith (2001) look at the effects of higher levels of racial diversity in universities, and find positive labor market effects for the students after their graduation. However, the authors find that once the proportion of minorities in a college surpasses around 18 percent, the effects begin to decline. If the effects from a high concentration of females (less gender diversity) in colleges leads to the same effects as a high white or a high minority concentration in colleges does, then women who graduate from a female dominated school will likely fare worse in the labor market than women who attended a coed school.

However, researchers who look at the relationship between high concentrations of females in schools and female academic outcomes tell a much more positive story. In general, they find at worst no effect, and in some cases, very positive effects. In her paper *Peer Effects in the Classroom: Learning from Gender and Race Variation* (2000), Caroline Hoxby focuses on high school students, and using sources of idiosyncratic variation between adjacent cohorts, finds that higher levels of females in classrooms is associated with higher test scores for women, as well as men. Smith (1990) and Smith, Wolf, and Morrison (1995) examine women's colleges, and respectively find that students at those colleges have better academic experiences and are more concerned with learning and civic involvement. Astin (1977) finds students at women's colleges are more likely to obtain leadership positions, become involved in student government, develop high aspirations and graduate. Whitt (1994) also finds a positive effect on leadership development at women's colleges. In terms of education outcomes, Kim and Alvarez (1995) find that the peer effects from being surrounded by women who see themselves as intellectually able is a significant predictor of academic success. In contrast, Miller-Bernal (1989) finds women at coeducational institutions are no more likely to become involved in campus activities. Stoecker and Pascarella (1991) and Riordan (1994) also found no difference in educational attainment for women in female dominated environments.

So, while a positive link can be seen between female dominated schools and academic performance, it is not yet exactly clear whether attending a female dominated school leads to improved labor market outcomes for its students. This paper will try to answer that question, and will examine the labor market effect of attending a female dominated school within the context of treatment effects. Few papers have yet examined the correlation between having more females in a school and the future labor market outcomes for their students. Those that do have relied on traditional least squares models to draw their conclusions. However, with the evidence of nonlinearities in returns to schooling (Tobias 2003), linearity is likely a too restrictive assumption to make. Further, using propensity score methods highlights potential common support problems. This paper will use both traditional least squares methods, as well as matching and propensity score weighting techniques to estimate whether women who graduated from a college with a high concentration of females had better labor market outcomes than their coed counterparts, both at initial entry into the labor market, as well as at the five and ten year intervals to see whether these effects (or lack thereof) persist into the future.

3.2 Methodology

This paper views the effect of attending a female dominated school as a binary treatment effect, where

$T_i = 1$ if student i attends a female dominated institution

$T_i = 0$ if student i attends a coeducational institution

Then, student i's wage Y_i would be $Y_i|T_i = 1$ if they attended a female dominated school, and $Y_i|T_i = 0$ if they did not. In potential outcomes notation these are $Y_i(1)$ and $Y_i(0)$.

The purpose of this paper is to estimate the treatment effect on labor market earnings. Due to the relative sizes of the treated and untreated subsets within the dataset that I use, in the propensity score models I only estimate this effect for the treated subset. In the program evaluation literature, this is called the average treatment effect on the treated (ATT). Defining the ATT as a function of the potential outcomes, it is given as:

$$ATT \equiv E[Y_i(1) - Y_i(0)|T_i = 1] \tag{3.1}$$

Because $Y_i(0)|T_i = 1$ is not observed, to estimate the ATT the following is required:

$$T \perp (Y(0), Y(1))|p(x) \tag{3.2}$$

$$0 \leq p(x) < 1,\ p(x) = Pr(T = 1|X = x) \tag{3.3}$$

The first condition relates to how students select into female dominated versus coed schools, and requires they select on an observed set of covariates, X. The second requires that *ex ante*, all of the students in the sample have some probability of attending each of the different types of schools.

Certainly these are more restrictive assumptions than would be desirable. However, they are no less, and in many cases are more flexible than those assumptions present in the previously described literature. Given that students do not select into female dominated schools on characteristics that are both unobservable and correlated with their future wages, the ATT can be expressed as:

$$ATT \equiv E[Y_i(1) - Y_i(0)|T_i = 1, p(x)] \tag{3.4}$$

I estimate the ATT using weekly wages for 1994, 1997 and 2003 to examine if the effect diminishes, grows or is stable through time. Although I expect the effects to differ for each of the three periods, it is unclear exactly how. If attending a predominantly female school affects wages by sending a signal to employers about the quality, motivation or ability to get along with other people, then the effect should show up in the 1994 wages but disappear over time. On the other hand, if attending a female dominated school endows its students with assets that affect the likelihood of promotion, such as increased confidence or assertiveness, then the effects may not manifest themselves until later in the career life of students.

3.2.1 Heterogeneous effects

It is possible that the treatment effect from attending a school with a high concentration of females varies across subgroups. Traits such as assertiveness may be valued differently by employers of workers at the upper tail of the income distribution than employers of workers at the lower tail. Employees with higher pay may see greater returns to leadership abilities than those with lower pay (conversely, an assertive woman may be valued less in high paying occupations if her high salary leads to insecurity in her superiors). If, as the single-sex school lobbyists suggest, attending a college with a high concentration of women endows or encourages traits such as confidence and assertiveness, then women who enter higher paying jobs may see a different labor market return from attending these schools than women who enter lower paying jobs.

Also possible is that the marginal return on the human capital received from attending a female dominated school is lower for students attending highly selective universities than for students attending less selective schools. If students attending more selective schools are already endowed with relatively more human capital, then students at less selective institutions may see a greater return to attending a female dominated college.

If these or other heterogeneous effects are present, averaging across the treated population will mute the estimated effects. Any one of the above scenarios (and perhaps others I have not considered) may occur, perhaps even simultaneously. On the other hand, heterogeneous effects may not be present at all. However, this is an empirical question, so in the matching models I examine it for two sets of subgroups. The sample is first split by students above and below median income, then split by students who attended more or less selective colleges. The treatment effects are estimated for each of these different subgroups (low income, high income, those who attended less selective colleges, and those who attended more selective colleges) to determine if heterogeneous effects exist.

3.3 Data

To estimate the effect of attending a school with a high concentration of females on a woman's wages, I use the 2003 follow up of the Baccalaureate and Beyond (B&B:93/2003) survey. Administered by the National Center for Education statistics, B&B:93/2003 was given to over 11,000 individuals who had received a Bachelor's degree in 1992 or 1993. To match individuals to their institutions, I rely on the restricted version of the dataset.

Data on enrollments and institutional control (public versus private) come from the Integrated Postsecondary Education Data System (IPEDS). IPEDS is also administered by the NCES, and contains yearly information on postsecondary institutions in the United States between 1984 and the present.

3.3.1 College selectivity

The 1992 edition of Barron's Profiles of American Colleges provided information on college selectivity. As this was published in 1992, this edition reflects data from 1991-1992 freshman classes. The rankings are based on a 6 point scale, ranging from "Noncompetitive" to "Most Competitive." The factors used in determining the category for each college were the median entrance exam (SAT and ACT) scores for the entering freshman class, the percentages of freshmen scoring 500 and above on the math sections and 600 and above on the verbal section of the SAT score, the percentages of freshmen scoring 21 and above and 27 and above on the Enhanced ACT, the percentage of entering freshmen who ranked in the upper fifth of their high school class, the percentage of entering freshmen who ranked in the upper two fifths of their high school class, whether or not minimum class rank and grade point average were required of applicants, and the acceptance rate. A total of 54 students attended "special" schools that do not fit well within the Barron's selectivity rankings. These included schools such as music conservatories, art institutes and medical colleges, and cater to a specialized student on a narrow path. Since these schools are sufficiently different from research universities or liberal arts schools, they warrant their own category.

The distribution of schools by selectivity ranking for both the set of schools in the Baccalaureate and Beyond sample and the set of schools in Barron's universe is given in Figure 3.1. The distributions within both the sample and the Barron's universe are very similar, with slightly more upper tier schools sampled in B&B:93/2003. Although these reflect the number of different schools of each type, not the distribution of students at each type of school, it is still encouraging to see that the two distributions match up fairly closely.

Other studies that have used the Barron's college rankings typically collapsed the seven categories into four, then interacted those with whether the university was public or private. I follow this approach to end up with eight college types: highly selective public, highly selective private, middle selective public, middle selective private, less selective public, less selective private, special public, and special private.

3.3.2 Classifying schools as female dominated

I classify a school as female dominated if it had at least 70 percent female enrollment while the student attended. I use the enrollment data available through the NCES' Integrated Postsecondary Education Data System (IPEDS), and construct the treatment indicator by dividing the average number of FTE females enrolled by the average total FTE enrollment while the student attended the sample school.

Within my sample, 43 different schools are classified as female dominated. These range across all of the different Barron's selectivity ratings. Figures 3.2a and 3.2b show the selectivity distributions for both the female dominated schools and the coeducational schools in the dataset. The figures represent the distribution of students in the sample enrolled in the different Barron's types, for students in coed schools, as well as those in female dominated schools. The percentage of students enrolled in "Less Competitive," "Competitive," and "Very Competitive" schools is very similar between the two types of colleges. However, there is a relatively higher percentage of female dominated school students enrolled in the "Special" category. This is due primarily to higher enrollment in medical colleges. There is also a higher percentage of coed students enrolled in the top two selectivity categories, "Highly Competitive" and "Most Competitive." It will be important to control for school selectivity in estimating the effect of attending a female dominated school.

Figure 3.3 shows the distribution of colleges present in my sample by the proportion of women who were enrolled. Although the data are fairly normally distributed around a mean of just over 54 percent female, there is a mass of colleges at nearly 100 percent female, as well as a mass between about 73 and 76 percent. Although the 70 percent cutoff is admittedly somewhat arbitrary, there appears to be a natural break in the data around this level.

A cutoff of 70 percent is also relevant when thinking about the implications of the growing number of women in American colleges. Many universities around the country are experiencing female enrollment rates above 60 percent, so examining the labor market effects on women at an enrollment rate of 70 percent may, if the effects are constant across enrollment rates, predict the effects in the future if the gender gap continues to grow.

3.3.3 Paring the data

Combining the individual and institution level data and dropping non-respondents and observations with missing school data, I start with 8,765 observations. 1,186 observations are eliminated because they enrolled in the sample school before 1988. This limits the dataset to those who graduated within 6 years of initial enrollment, eliminating students who originally enrolled in the sample school decades prior to graduation and minimizing long-term attendance gaps for the sample students. Taking out the males and those observations who attended schools with no full time students leaves a dataset containing 4,412 observations. 4,087 of these observations received their Bachelor's degree from a coeducational institution, while 325 received it from a female dominated school.

Table 3.1 gives summary statistics for a selection of variables from this pared down dataset. In general, schools classified as female dominated tended to be private and of middle selectivity while coeducational institutions tended to be middle selectivity and public. Across the different Barron's rankings, the female dominated and coed schools were for the most part similarly distributed. There were differences in the proportion of students enrolled in the "Special" schools, which include art, music and fashion institutes, as well as medical colleges. These schools are nearly all predominantly female. Both sets of students are predominantly white, although those that attended female dominated schools were slightly more racially diverse. They also tended to have parents with less education, less income and were less likely to be married. In terms of their own family, these students were more likely to have a husband with a job, and on average their husbands earned higher incomes. They also were more likely to have dependents. Students at these schools also generally scored worse on their combined SAT/ACT score, and were less likely to take the tests.

As I estimate the effect of attending a female dominated school on personal weekly wages in 1994, 1997 and 2003, I am limited to persons who participated in the labor force for those years. For 1994, I lack income information on 1,068 students (996 coed students, 72 female dominated students). Enrollment in a graduate program explains this for 345 women (328 from coed universities, 17 from female dominated universities), but for the rest, I am unsure of the reason. For 1997, income information is missing for 855 students (795 coed students, 60 female dominated students). 212 of these were enrolled in either a graduate program or in another undergraduate program (195 from coed universities, 17 from female dominated universities). For 2003, income information is missing for 437 students (405 from coed universities, 32 from female dominated universities). 28 of these were enrolled in graduate or undergraduate school (25 from coed universities, 3 from female dominated universities). This selection into the labor market will bias the estimates if students from female dominated colleges select differently than students at coed schools. Looking at the above proportions of unemployed students at each of the two types of schools, this does not seem to be the case, however this could still be a potential issue.

3.4 Wage model and results

In this section I estimate the wage effect of attending a predominantly female school for women. There are many different methods available to estimate treatment effects. These range from a simple regression function with covariates (Daniel, Black and Smith 2001), to mean differencing the regression function (Rubin 1977), to propensity score matching (Rosebaum and Rubin 1983; Rosebaum and Rubin 1985), to propensity score weighting (Hirano, Imbens, Ridder 2000), to a combination of weighting and mean differencing (Hirano and Imbens 2002). To examine the effect of attending a female dominated school on labor market earnings, I employ versions of each of the above approaches. I separate the estimation models into two sets. The first set consists of the more traditional least squares models that do not use the propensity score as part of the estimation. The second set consists of the models that include the estimated propensity score. Within this set, the estimated propensity score enters the different models in different ways. It first enters simply as a regressor, then as a way to "match" treated and untreated observations, and finally is used to construct weights.

In each of the two model sets, I begin by estimating a naive model which includes only the dependent variable, treatment indicator, and in the case of the second set, the propensity score. Examining and comparing these naive models gives both an idea of the correlation between attending a female dominated school and wages for each of the three cross-sections, as well as a demonstration of how much the estimated coefficient on the treatment indicator is attenuated by inclusion of the estimated propensity score. Other models estimated as part of the first set include a basic least squares regression, as well as an instrumental variables approach and a Heckman two-step selection model. The basic least squares models estimate average treatment effect, or ATE, not the ATT. Once the propensity score is estimated and the common support of the treated and untreated groups is examined, it becomes clear that estimating the ATE is likely not appropriate within the Baccalaureate and Beyond sample. As such, the models in the second set give estimates only of the ATT. In addition to the naive model described above, I estimate nearest neighbor matching models as well as the regression adjustment and propensity score weighting method discussed in Hirano and Imbens (2002).

3.4.1 Least squares methods

Naive OLS

As an initial step, I plot income data against treatment and show the raw correlations for each of the different years in Figures 3.4, 3.5, and 3.6. From these figures it appears the correlation between income and attendance at a female dominated school is strongest in the initial 1994 survey. In 1994, the estimated treatment effect is about .21, and diminishes as the sample of students spends more time in the labor force. The estimated coefficients for 1997 and 2003 wages are about .16 and .04. It also appears that the wage dispersion is increasing slightly between 1994 and 2003.

Unless students select into female dominated and coed schools randomly, selection bias will plague the above estimates of the treatment effect. Below I follow the estimation strategy of Daniel, Black and Smith (2001) and include a set of covariates along with the treatment.

OLS with covariates

Here I estimate the labor market effects of attending a female dominated school using traditional least squares methods. I begin with a very basic OLS framework and regress the outcome of interest (1994, 1997 or 2003 logged wages) on the indicator for treatment and a set of covariates. With this model I can see if these effects persist once variation in demographics and school selectivity is controlled for. Specifically, the estimation equation is:

$$ln(Y_i) = \alpha_0 + \tau T_i + \alpha_1' X_i + \epsilon_i \tag{3.5}$$

where Y_i is either 1994, 1997 or 2003 wages, τ is the average treatment effect (or ATE) and X_i is the set of covariates. Table 3.2 includes the a summary of estimated treatment effects from the above model, as well as for the estimates from the naive model, instrumental variables model and Heckman selection model that will follow. Tables 3.3, 3.4, and 3.5 include the full set of coefficient estimates for all of the included covariates in each of the different cross-sections.

Once demographics, parental, family, and decision variables are controlled for, the estimated average treatment effect was attenuated by about 40 percent in the model for 1994 wages, about 15 percent for 1997 wages, and about 90 percent in the model for 2003 wages. Additionally, the difference between the treatment effect estimated from 1994 wages and that estimated from 1997 wages disappears.

For each of the periods, the regression estimates suggest a wage penalty to growing up in a region outside of the Mid-East,[1] although for 1997 these differences are not statistically significant. Women with less educated fathers also tended to earn less across the three periods. This relationship did not hold for mother's education. Students with wealthier parents tended to earn more, however this was only statistically significant for the 2003 cross-section. The type of high school attended by the student was statistically significant in both the 1997 and 2003 cross-sections, but was not statistically significant in 1994. In both 1997 and 2003, attending a

[1] This region includes the states Delaware, Maryland, New Jersey, New York, Pennsylvania and Washington D.C.

Catholic high school was positively associated with labor market earnings, relative both to attending a public high school, as well as a private high school of another denomination. In terms of family demographics, if a woman had children she cared for, or had an employed husband, she tended to earn less in the 1997 and 2003 labor markets. However, given that her husband was employed, the woman's income was positively correlated with her husband's. It is possible this positive coefficient estimate is a result of selection into the labor force. Women married to men with higher income may be less likely to work, and only the particularly career-minded individuals are observed. This bias due to selection into the labor force could affect the ATE estimates if women from female dominated schools select into the labor force differently than women that attended coed schools. For this reason, I estimate a Heckman two-step selection model.

Heckman selection model

One of the issues with estimating any sort of wage equation is that individuals, and in particular women, select into the labor force non-randomly. The number of children, marital status and size of a partner's salary all factor into the decision of whether or not to enter the labor force. In the context of this paper, selection into the labor force will be a problem if women who attended a female dominated school select differently from women who attended a coed institution.

In estimating the selection model, I included the spousal, parent income and dependents variables in the selection equation as well as the wage equation. Variables for current enrollment and whether or not a person had a disability were entered in the selection equation, but excluded from the wage equation. Table 3.2 contains a summary of the ATE estimates from two-step Heckman selection models for logged wages in 1994, 1997 and 2003. Tables 3.6, 3.7, and 3.8 contain the full set of coefficient estimates for both the wage and selection equations.

For each cross-section I reject the null that the selection equation and wage equation were independent, and it appears selection is an issue in all three of the cross-sections (χ^2_1 equaled about 414.52 in 1994, about 283.39 for 1997 and about

241.81 for 2003). However, the coefficient estimates of the treatment variable did not change much after controlling for selection into the labor force. The 1994 estimate increased slightly from .134 to .153, and the estimate from 1997 decreased slightly from .139 to .11. The coefficient estimate for 2003 was not statistically significant, although did increase from about .002 to .034.

In each of the three cross-sections, family demographics were most important in the student's decision to participate in the labor force, specifically with respect to the number of dependents the woman had. In the first two cross-sections, husband income was slightly positively correlated with female labor force participation, while husband employment was negatively correlated. This suggests that perhaps participation in the labor force had more to do with aspects of the marriage relationship than with the spousal income. Other significant variables included whether or not the woman was currently enrolled in school and whether or not she had a disability. Both of these variables had negative coefficients in the selection equation.

Instrumental variables

Endogeneity is a constant concern in much of the literature estimating the returns to education. College admissions officers will choose to accept students they think will be better suited to their type of school, and in making their decision to attend, students likely consider the expected return on their education. In light of this, and the possibility that the above set of covariates do not completely control for student selection into female dominated schools, I explore an instrumental variables approach.

Attendance at a female dominated school is instrumented using the proportion of women's colleges present in the state in which the student's parents reside. I expect that students raised in states with a higher proportion of women's colleges will have better information regarding their existence, as well as their costs and benefits relative to coeducational institutions. Given this, a higher proportion of women's colleges may help predict whether or not a woman decides to attend a female dominated school.

To create the instrument the number of schools designated as women's colleges in 1986 was counted for each state, then divided by the number of four year institutions present within that state in 1986.[2] Using this, Equation 7 is estimated using an IV framework. The treatment effect results are summarized for each of the years in Table 3.2. The coefficient estimates for the additional regressors are given in Table 3.9 for 1994, Table 3.10 for 1997 and Table 3.11 for 2003. First stage results are included in tables A1, A2 and A3. The F statistics for the excluded instrument in the different cross-sections are about 15 for 1994 and 2003, and about 18 for 1997. This suggests the instrument has decent explanatory power, but is not particularly strong.

While instrumenting for attending a female dominated school did not substantially affect the coefficient estimates for the other covariates, the 1994 and 1997 treatment effect estimates were affected. The estimated treatment effect from the 1994 cross-section increased to about 0.32, but the precision decreased dramatically and the coefficient was no longer statistically significant. For 1997, the estimated ATE became negative, although was also not statistically significant. The precision also decreased in the 2003 cross-section, although the estimated ATE did not differ by much from the other estimation methods.

The difference in the coefficient estimates using IV and those from the basic OLS model is potentially due to the poor small sample properties of IV. The small number of treated observations and the fact that the instrument is not particularly strong is likely causing the lack of precision of the coefficient estimates, and is potentially also causing a small amount of bias. In the next section, I move away from least squares estimators and instead use the proportion of women's colleges present in the state the student's parents were residents of as a covariate in the estimation of the propensity score.

[2] Each of these women's colleges have enrollments close to 100 percent

3.4.2 Propensity score/Matching methods

This section includes those models which make use of the estimated propensity score. As described above, the propensity score is the estimated probability a student enrolls in a female dominated school.

The propensity score is estimated with a logit model that includes the variable controlling for the proportion of women's colleges present in the state the student's parents were residents of, as well as the different sets of categorical variables present in the section 4.2.1 wage model. The results from this logit model are given in Table 3.12. A total of 54 observations who attended coed universities had missing values, so the number of observations available when implementing the matching and regression adjustment estimators is reduced.

Figure 3.7 shows the distribution of the propensity scores for both treatment and control distributions. The two distributions are fairly similar, and range from just above zero to about 0.4. This figure highlights the potential common support problems that are more hidden when using traditional least squares methods. Within this sample, there appear to be some observations in the treated group without close matches. As such, I exclude these from the treatment effect estimations by limiting the sample to observations with propensity score values below 0.35.[3] Additionally, this figure illustrates that estimating the ATE is probably not appropriate within this sample, something that would not have been known were the estimation to stop with the least squares methods. Estimating the ATE requires assuming both that (1) $\widehat{Y_i(0)|T_i} = 1$ is a good proxy for $Y_i(0)|T_i = 1$, as well as (2) $\widehat{Y_i(1)|T_i} = 0$ is a good proxy for $Y_i(1)|T_i = 0$. Since there are a large number of untreated observations across the treated support to match from, (1) is likely a plausible assumption. However, the small number of treated observations at the lower end of the untreated support suggests that the (2) may be a bad assumption. Since estimating the ATT only requires (1), only the average treated on the treated effect will be estimated in the models using the propensity score.

[3] Varying this cutoff slightly did not affect the results much

The propensity score as a regressor

Before the matching models are estimated, it is informative to first include the propensity score as a regressor in the naive model in section 4.1.1. Although this requires estimating the ATE, it will show how much variation in the estimates is soaked up by $p(x)$. Including the propensity score helps control for that part of the estimated ATE due to certain types of students having a higher probability of enrollment in a specific type of school. The results for each of the three cross-sections are given in Table 3.13. Including the "Probability of attending a female dominated college" as a regressor in the naive model attenuated the treatment effect by about 16 percent in 1994 and 1997 and by about 40 percent in 2003.

Propensity score matching

In this section I estimate the ATT using the matching estimators common in the program evaluation literature (Heckman, Ichimura and Todd 1997; Rosenbaum and Rubin 1983; Black and Smith 2003). I favor the matching models over least squares methods because they are more transparent with regards to the common support of the treatment and control groups and do not require linearly conditioning on the observables.

To eliminate the dimension problems associated with matching on a large matrix of covariates, I use the propensity score to create matches between the treated and untreated observations. For each of the cross-sections, each student who attended a female dominated college is matched with replacement to the five students at coed universities who had the closest propensity scores.[4] The difference in the logged weekly wages between the treated student and the average from the group of five closes matches was calculated for each of the treated observations. The ATT and was then calculated from this set of averages. Table 3.14 presents the estimates from this method for each of the different cross sections. Both bootstrapped and subsampled standard errors are presented in Table 3.14, however statistical inference

[4]Calculations using kernel bandwidths yielded very similar estimates

is made using only the subsampled standard errors.

Based on these estimates, attending a female dominated school was associated with higher weekly wages for 1994. Aside from women at the bottom half of the income distribution, the estimates were positive and statistically significant for each of the subgroups, as well as for the full sample. From the estimates using the full sample, it appears that women who graduated from these schools saw returns of about 16 percent. Although slightly attenuated, this effect persisted into 1997, where the returns were estimated to be about 13 percent. It does not appear this effect persisted into 2003. The 2003 estimate of the ATT is much smaller and not statistically different from zero. From the propensity score matching estimates, women who attended a predominantly female school saw returns in the labor market early in their working life, but these diminished and eventually died off after about a decade.

Heterogeneous subgroups

As expected, the effects tended to differ for the different subgroups. Based on the different subsets in 1994, the top half of the current job earnings distribution seems to have seen a return from attending a female dominated school, while the bottom half did not. This difference was not present in the 1997 and 2003 cross-sections.

There is also a difference between the ATT estimates between students who attended less selective colleges and those who attended more selective colleges, although in the 1994 cross section the difference was small. Until 2003, students at public and private universities classified as “Non-competitive,” “Less Competitive” or “Competitive” on average saw a slightly greater return to attending a female dominated school than students who attended the “More Competitive,” “Highly Competitive” and “Most Competitive” universities. For the 1994 cross section, this difference was statistically significant at the 5 percent level. The difference was not statistically significant for the 1997 cross section. One potential explanation is that students at less selective schools have a greater marginal return on the additional human capital conferred by attending a female dominated school.

In general, women who received their B.A. from a school with a proportion of females greater than 70 percent saw returns upon immediately entering the labor market, but those returns diminished over time. It also appears that different subgroups of women were affected in different ways. For 1994 and 1997, women at more selective universities saw labor market returns, but these were slightly smaller than those seen by women at less selective colleges. These differences also existed for women at different parts of the income distribution in the 1994 cross section. Women who entered higher paying jobs immediately after college tended to see more of a return than women who entered lower paying jobs. This suggests that attending a women's college may be a positive signal to higher tier employers.

Propensity score weighting and regression adjustment

I move now to estimating the same set of ATTs using the propensity score weighting and regression adjustment method. Hirano and Imbens (2002) show that a combination of these typically yields more stable values for the estimates of the treatment effect. The ATT is estimated using the equation:

$$ln(Y_i) = \alpha_0 + \tau T_i + \alpha_1^{'} X_i + \alpha_2^{'}(X_i - \bar{X}_1)T_i + \epsilon_i \tag{3.6}$$

where τ is the estimated treatment effect, X_i is a matrix of covariates, $\bar{X}_1$ is a vector of conditional means from the treated observations and ϵ_i is the error term assumed to have conditional mean zero and defined as the unobserved variables affecting a persons wage (such as motivation or networking skills). To make the treatment and control groups comparable, the following weights are used with the above equation:

$\hat{\omega}(T, x) = T + (1 - T)\frac{\hat{p}(x)}{1-\hat{p}(x)}$

Within X is included the full set of covariates seen in the previous sections. Table 3.15 summarizes the ATT estimates from the weighting and regression adjustment estimators. The t-statistics are included within the table for an idea of the preciseness of the different estimates, although these are based on standard errors that are not yet valid.

While the point estimates of the ATT do not exactly match the point estimates from the propensity score matching estimators, they are fairly close. However, the estimated ATT for the 1997 cross-section was slightly greater than that estimated from the 1994 cross-section. Tables 3.16, 3.17, and 3.18 display the full regression results, less the estimated coefficients from the conditional mean-differences. Because these conditional mean-differences are included in the estimation, the coefficients are not interpretable like those in the previous sections were. However, they are presented to show which were statistically significant. Although there were some instances in which the coefficient estimates were no longer statistically significant, in general, if the coefficient had a statistically significant coefficient in the basic OLS model, it continued to be statistically significant in the regression adjustment and propensity score weighting model.

3.5 Concluding remarks

While a significant literature exists in estimating the scholastic effect of attending a school with a large proportion of females, few studies have so far attempted to directly estimate how this translates to the labor market. Those that have use basic regression techniques and were unable to find any effect from attending a female dominated school. Using a variety of techniques, I find positive and statistically significant labor market effects. Additionally, my results suggest that these effects diminish over time, and that different subgroups of students may be affected differently. Specifically, using the propensity score matching methods I find that first, students at less selective colleges tended to realize a slightly greater return from attending a female dominated school and second, that when the students first entered the labor market in 1994, those in the upper 50 percent of the income distribution benefitted more than students in the lower 50 percent of the income distribution.

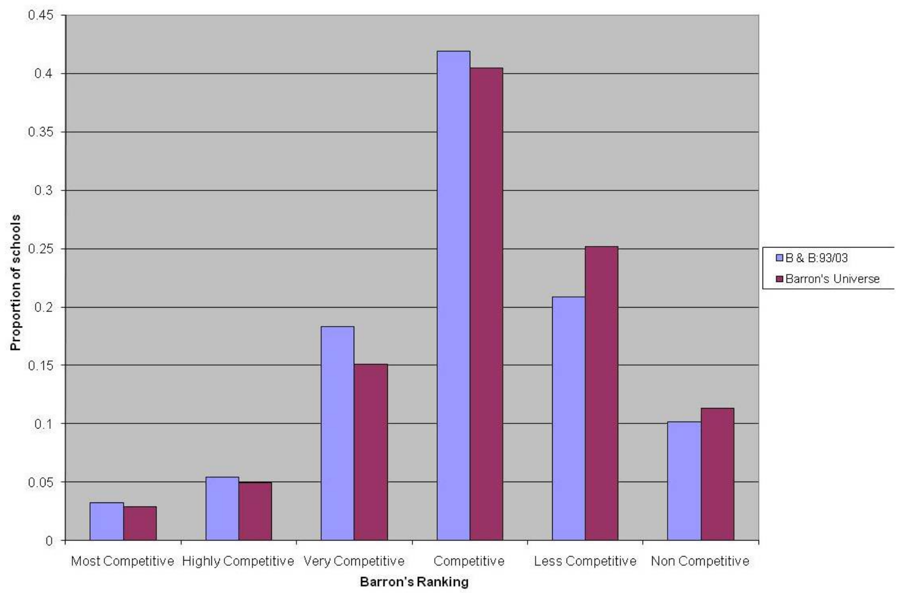
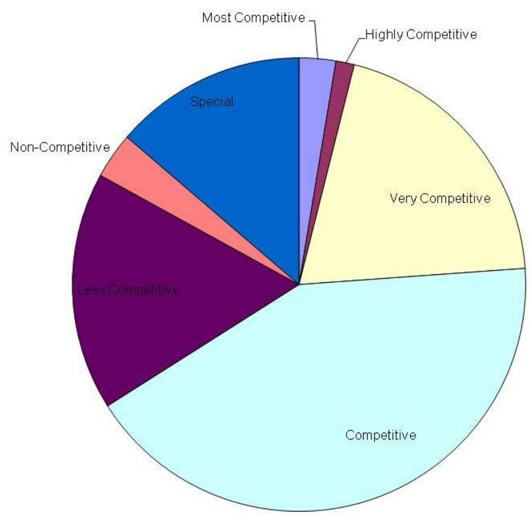
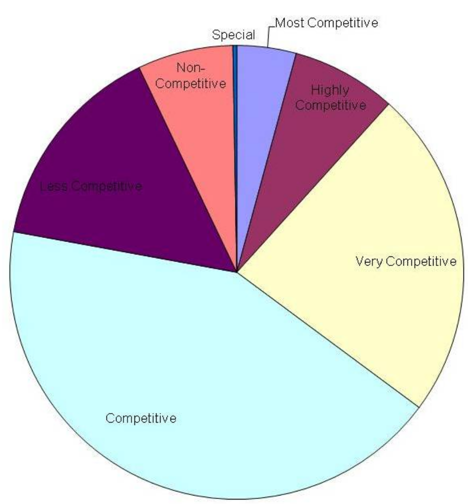


Figure 3.1: Proportion of Schools by Selectivity



(a) Female Dominated Schools



(b) Coed Schools

Figure 3.2: Selectivity distributions for coed and female dominated schools

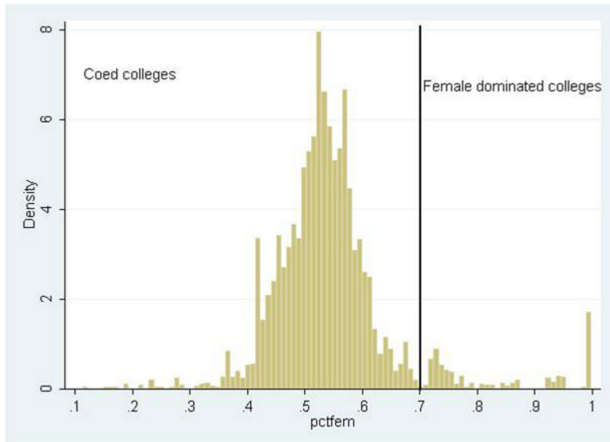


Figure 3.3: Distribution of Schools by Female Enrollment

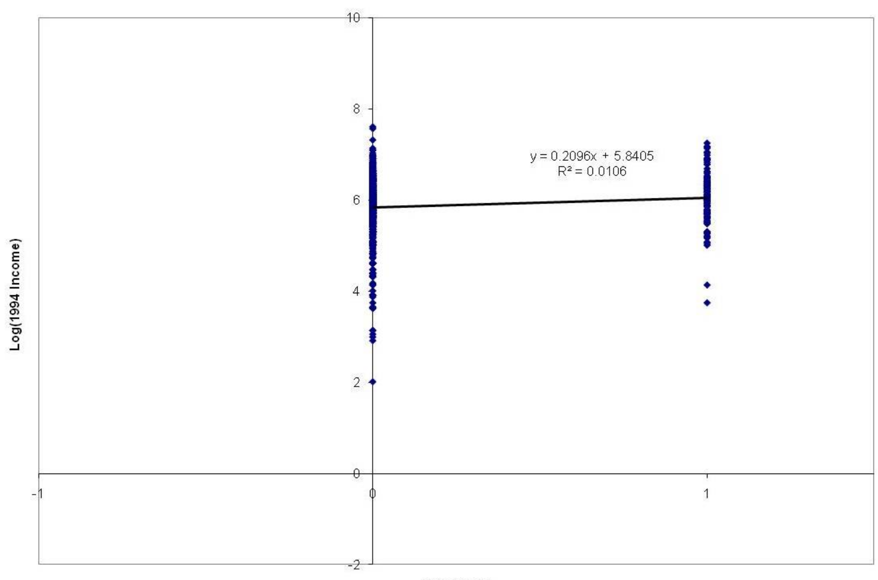


Figure 3.4: 1994 logged weekly wages against treatment

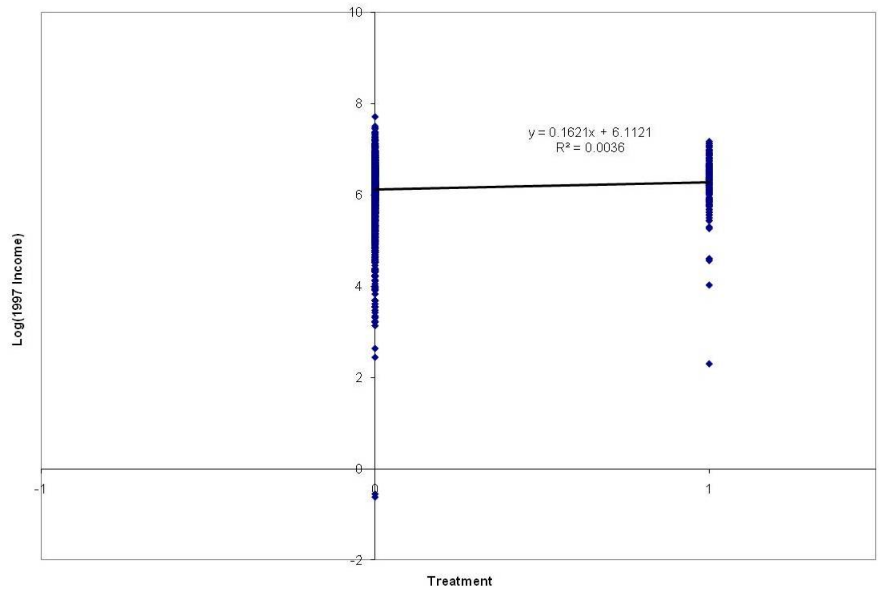


Figure 3.5: 1997 logged weekly wages against treatment

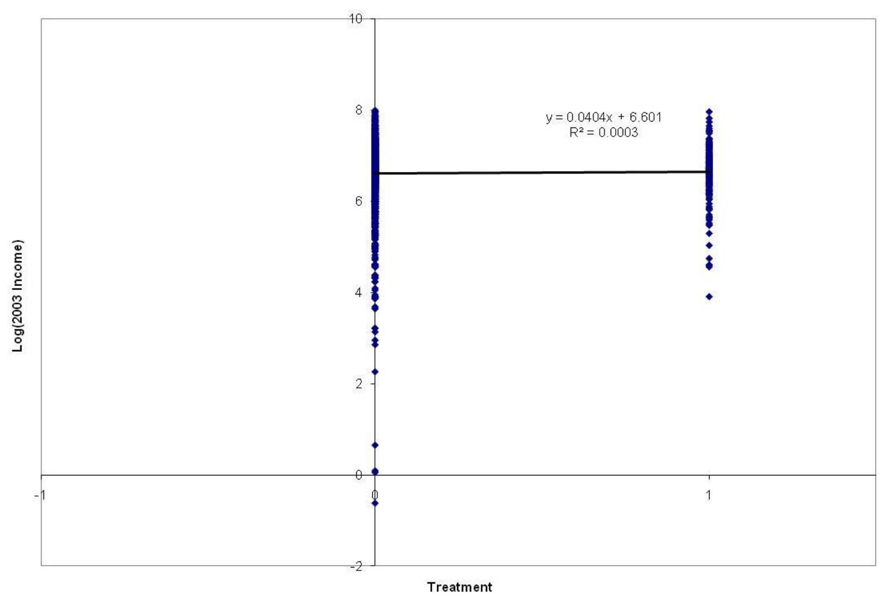


Figure 3.6: 2003 logged weekly wages against treatment

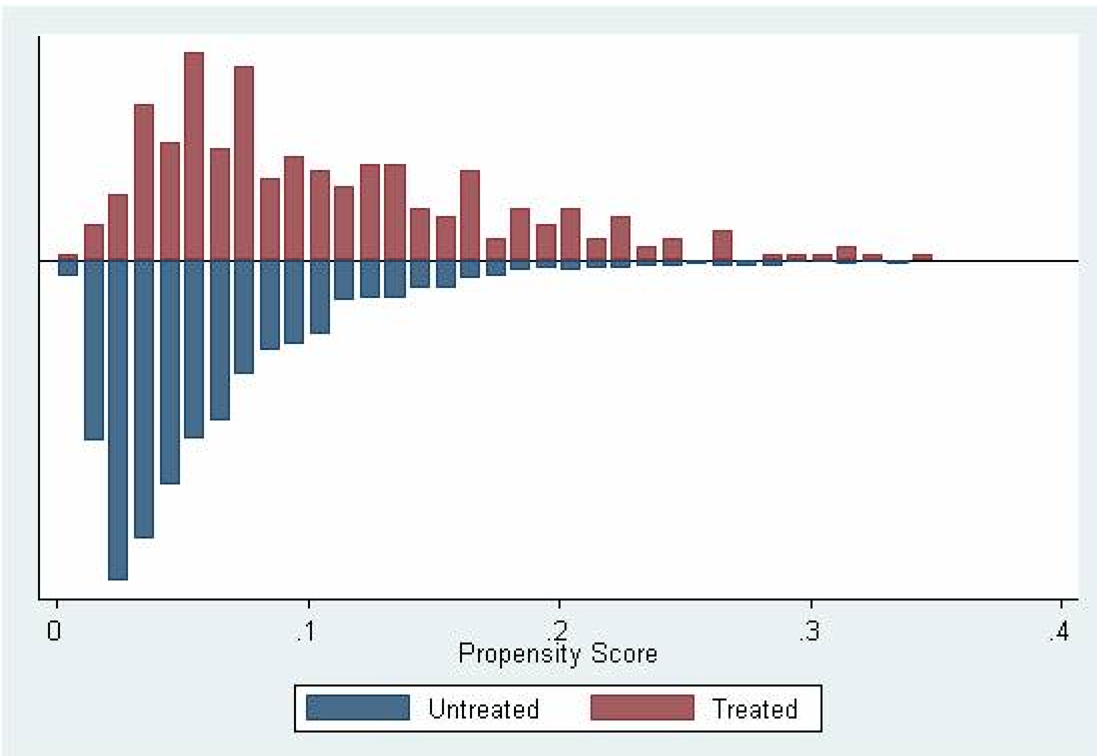


Figure 3.7: Propensity Score Graph

Table 3.1: Summary Statistics

	School Type		
	Coeducational	**Female dominated**	**t stat**
College Variables			
Low quality, public	14.36%	2.63%	-4.44
Low quality, private	7.09%	16.54%	2.47
Middle quality, public	44.92%	10.53%	-9.57
Middle quality, private	20.82%	51.88%	6.96
High quality, public	5.61%	1.50%	-1.80
High quality, private	5.83%	2.26%	-1.43
Special, public	0.00%	7.89%	2.48
Special, private	0.22%	4.51%	1.52
HBCU	3.62%	2.26%	-0.55
Ethnicity			
American Indian	0.71%	0.75%	0.02
Asian	2.75%	3.38%	0.23
Black	6.13%	6.77%	0.20
Hispanic	4.31%	4.14%	-0.06
White, non hispanic	85.47%	84.21%	-0.33
Other	0.19%	0.38%	0.12
Demographics			
Father's education	2.39	2.27	-1.53
Mother's education	2.08	2.00	-1.11
Parents married	75.39%	69.55%	-1.36
Parent inc: 0-30,000	26.17%	25.56%	-0.14
Parent inc: 30,000-60,000	46.52%	48.12%	0.36
Parent inc: 60,000+	27.31%	26.32%	-0.24
Parents helped pay for school	69.80%	58.65%	-2.51
Age in 1994	23.90	25.99	11.81
Family variables			
Spouse employed in 94	11.31%	14.66%	0.89
Spouse's 1994 income	6,095.64	10,577.07	288.04
Spouse employed FT in 97	47.83%	47.69%	-0.03
Spouse's 1996 income	17,126.98	18,941.44	161.82
Spouse employed FT in 03	67.16%	62.26%	-1.11
No dependents	89.02%	85.34%	-0.98
1 dependent	4.74%	6.39%	0.53
2 dependents	3.79%	5.64%	0.61
3 dependents	1.66%	1.13%	-0.26
4 dependents	0.49%	1.13%	0.31
Ability			
Did not take SAT/ACT	15.40%	26.32%	2.60
SAT/ACT q1	22.38%	24.81%	0.58
SAT/ACT q2	24.42%	19.55%	-1.21
SAT/ACT q3	20.36%	16.17%	-1.08
SAT/ACT q4	17.44%	13.16%	-1.16

Table 3.2: Summary of Least Squares Estimates

Treatment	**1994 Wages**			**1997 Wages**			**2003 Wages**		
	OLS	Hman	IV	OLS	Hman	IV	OLS	Hman	IV
Estimated ATE	0.134	0.153	0.317227	0.139	0.11	-0.089	0.002	0.034107	0.059
	(3.28)***	(4.07)***	(0.66)	(3.33)***	(3.02)***	(0.13)	(0.05)	(0.79)	(0.11)
Observations	3023	3869	3023	3167	3630	3167	3498	3715	3498
R squared	0.1		0.0869	0.07		0.0625	0.11		0.116

Table 3.3: 1994 Logged Wages OLS with Covariates

		Coefficient	t stat
ATE		0.13	(3.28)***
College variables			
Attended HBCU		-0.101	(2.00)**
College selectivity ("Middle quality public" excl.)			
	Missing	0.039	(0.68)
	Low quality, public	-0.032	(1.23)
	Low quality, private	0.008	(0.27)
	Middle quality, private	-0.007	(0.29)
	High quality, public	0.102	(2.53)**
	High quality, private	0.055	(1.22)
	Special, public	0.373	(3.93)***
	Special, private	-0.145	(1.28)
Parent variables			
Parent region ("Mid-east" excl.)			
	New England	-0.011	(0.26)
	Great Lakes	-0.069	(2.20)**
	Plains	-0.09	(2.67)***
	Southeast	-0.065	(2.41)**
	Southwest	-0.027	(0.74)
	Rocky Mountains	-0.057	(1.03)
	Far West	-0.011	(0.34)
Parent age (40-45 excl.)			
	Missing	0.034	(0.48)
	35-40	0.03	(0.36)
	45-50	0.022	(0.79)
	50+	0.011	(0.39)
Mother education (BA excl.)			
	Missing	-0.115	(0.88)
	High School	0.006	(0.21)
	Some college	0.019	(0.69)
	MA	-0.012	(0.33)
	PhD/Other professional	-0.056	(0.78)
Father education (BA excl.)			
	Missing	-0.087	(0.79)
	High School	-0.074	(2.48)**
	Some college	4.7E-06	(0.00)
	MA	-0.016	(0.54)
	PhD/Other professional	-0.06	(1.51)
Mom had more edu. than dad ("No" excl.)			
	Missing	0.02	(0.66)
	Yes	-0.002	(0.09)
Parent marital status ("Yes" excl.)			
	Missing	0.046	(0.40)
	Married	0.019	(0.71)
Parents helped with college ("Yes" excl.)			
	No	0.021	(0.86)
	Missing	0.197	(2.12)**
Parent income (20-60k excl.)			
	0-20k	-0.055	(2.49)**
	60k+	0.026	(1.15)
Family variables			
Spouse employed		0.011	(0.34)
Spouse's income		1.57E-06	(2.05)**
Spouse's income sq.		1.81E-12	(2.16)**
Number of dependents		-0.041	(2.61)***
Demographics			
Earned another degree/license prior to BA		0.046	(2.23)**
Ethnicity ("Hispanic" excl.)			
	Missing	-0.071	(0.47)
	Native American	0.051	(0.42)
	Asian	-0.126	(1.59)
	Black	-0.059	(1.07)
	White	-0.127	(2.98)***
SAT/ACT merged quartile ("Quartile 3" excl.)			
	Missing	0.006	(0.16)
	Quartile 1	-0.017	(0.60)
	Quartile 2	0.016	(0.57)
	Quartile 4	0.037	(1.19)
High school type ("Private, Catholic" excl.)			
	Missing	-0.213	(1.57)
	Public	0.018	(0.46)
	Private, non-Catholic	0.04	(0.81)
	Private, non-religious	-0.029	(0.48)
Age in 1992		0.072	(4.53)***
Age in 1992 sq.		-0.001	(3.53)***
Undergrad decision making variables			
Crime rate important ("Yes" excl)			
	No	0.003	(0.13)
	Missing	-0.152	(1.20)
Graduation rate important ("Yes" excl.)		0.005	(0.21)
Placement rate important ("Yes" excl.)		-0.067	(3.26)***
Constant		4.755	(19.74)***
Observations		3007	
R-squared		0.1	

Robust t statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 3.4: 1997 Logged Wages: OLS with Covariates

		Coefficient	**t stat**
ATE		0.14	(3.33)***
College variables			
Attended HBCU		0.022	(0.42)
College selectivity ("Middle quality public" excl.)			
	Missing	-0.046	(0.69)
	Low quality, public	-0.011	(0.33)
	Low quality, private	-0.027	(0.72)
	Middle quality, private	-0.012	(0.34)
	High quality, public	0.055	(1.12)
	High quality, private	0.09	(1.42)
	Special, public	0.056	(0.35)
	Special, private	-0.24	(1.48)
Parent variables			
Parent region ("Mid-east" excl.)			
	New England	-0.013	(0.24)
	Great Lakes	0.006	(0.14)
	Plains	-0.04	(0.86)
	Southeast	-0.038	(0.87)
	Southwest	0.03	(0.65)
	Rocky Mountains	-0.052	(0.75)
	Far West	0.081	(1.77)*
Parent age (40-45 excl.)			
	Missing	0.187	(2.74)***
	35-40	0.015	(0.18)
	45-50	0.023	(0.62)
	50+	0.06	(1.69)*
Mother education (BA excl.)			
	Missing	0.107	(0.55)
	High School	0.078	(1.79)*
	Some college	0.071	(1.61)
	MA	-0.043	(0.74)
	PhD/Other professional	0.043	(0.56)
Father education (BA excl.)			
	Missing	-0.283	(1.85)*
	High School	-0.073	(1.88)*
	Some college	-0.092	(2.31)**
	MA	-0.047	(1.13)
	PhD/Other professional	-0.066	(1.39)
Mom had more edu. than dad ("No" excl.)			
	Missing	0.081	(2.18)**
	Yes	-0.029	(0.70)
Parent marital status ("Yes" excl.)			
	Missing	-0.381	(1.54)
	Married	0.056	(1.48)
Parents helped with college ("Yes" excl.)			
	No	0.002	(0.07)
	Missing	-0.051	(0.59)
Parent income (20-60k excl.)			
	0-20k	-0.054	(1.95)*
	60k+	0.019	(0.61)
Family variables			
Spouse employment ("Unemployed" excl.)			
	Missing	-0.012	(0.25)
	Employed full-time	-0.1	(1.98)**
Spouse's income		4.5E-06	(3.61)***
Spouse's income sq.		2.0E-11	(9.40)***
Number of dependents		-0.078	(4.35)***
Demographics			
Earned another degree/license prior to BA		0.008	(0.31)
Ethnicity ("Hispanic" excl.)			
	Missing	-0.206	(0.86)
	Native American	0.113	(0.76)
	Asian	0.144	(1.61)
	Black	0.121	(1.84)*
	White	-0.031	(0.53)
	Other	0.122	(0.37)
SAT/ACT merged quartile ("Quartile 3" excl.)			
	Missing	-0.057	(1.25)
	Quartile 1	-0.066	(1.79)*
	Quartile 2	-0.015	(0.45)
	Quartile 4	-0.008	(0.22)
High school type ("Private, Catholic" excl.)			
	Missing	-0.224	(1.72)*
	Public	-0.123	(3.07)***
	Private, non-Catholic	-0.058	(1.15)
	Private, non-religious	-0.198	(2.35)**
Age in 1992		0.057	(2.90)***
Age in 1992 sq.		-0.001	(2.74)***
Job variables			
Tenure		0.006	(5.39)***
Tenure sq.		-1.6E-05	(2.48)**
Undergrad decision making variables			
Crime rate important ("Yes" excl)			
	No	0.001	(0.04)
	Missing	-0.064	(0.54)
Graduation rate important ("Yes" excl.)		-0.007	(0.20)
Placement rate important ("Yes" excl.)		-0.057	(2.14)**
Constant		5.286	(18.06)***
Observations		3150	
R-squared		0.07	

Robust t statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 3.5: 2003 Logged Wages: OLS with Covariates

		Coefficient	t stat
ATE		0.002	(0.05)
College variables			
Attended HBCU		-0.009	(0.12)
College selectivity ("Middle quality public" excl.)			
	Missing	0.018	(0.18)
	Low quality, public	-0.112	(3.33)***
	Low quality, private	-0.047	(0.86)
	Middle quality, private	0.014	(0.50)
	High quality, public	0.097	(2.26)**
	High quality, private	0.082	(1.61)
	Special, public	0.227	(2.37)**
	Special, private	-0.119	(1.10)
Parent variables			
Parent region ("Mid-east" excl.)			
	New England	0.015	(0.37)
	Great Lakes	-0.051	(1.49)
	Plains	-0.133	(3.21)***
	Southeast	-0.093	(3.22)***
	Southwest	-0.101	(2.35)**
	Rocky Mountains	-0.097	(1.56)
	Far West	-0.056	(1.24)
Parent age (40-45 excl.)			
	Missing	-0.011	(0.15)
	35-40	0.004	(0.04)
	45-50	-0.018	(0.56)
	50+	-0.016	(0.49)
Mother education (BA excl.)			
	Missing	0.147	(0.98)
	High School	0.06	(1.74)*
	Some college	0.018	(0.54)
	MA	0.029	(0.78)
	PhD/Other professional	-0.025	(0.38)
Father education (BA excl.)			
	Missing	-0.158	(1.16)
	High School	-0.041	(1.22)
	Some college	-0.028	(0.84)
	MA	-0.031	(0.88)
	PhD/Other professional	-0.049	(1.09)
Mom had more edu. than dad ("No" excl.)			
	Missing	0.031	(0.97)
	Yes	-0.009	(0.30)
Parent marital status ("Yes" excl.)			
	Missing	-0.172	(0.58)
	Married	-0.032	(1.14)
Parents helped with college ("Yes" excl.)			
	No	-0.046	(1.47)
	Missing	-0.012	(0.13)
Parent income (20-60k excl.)			
	0-20k	-0.034	(1.35)
	60k+	0.074	(2.78)***
Family variables			
Spouse employment ("Unemployed" excl.)			
	Missing	-0.179	(4.34)***
	Employed full time	-0.149	(3.89)***
	Employed part-time	-0.02	(0.34)
Dependents age 0-17		-0.073	(6.07)***
Demographics			
Earned another degree/license prior to BA		0.026	(1.11)
Ethnicity ("Hispanic" excl.)			
	Missing	-0.039	(0.21)
	Native American	0.01	(0.07)
	Asian	0.025	(0.33)
	Black	-0.069	(1.04)
	White	-0.164	(3.07)***
	Other	-0.165	(0.62)
SAT/ACT merged quartile ("Quartile 3" excl.)			
	Missing	-0.048	(1.27)
	Quartile 1	-0.076	(2.41)**
	Quartile 2	-0.042	(1.38)
	Quartile 4	0.011	(0.30)
High school type ("Private, Catholic" excl.)			
	Missing	-0.295	(1.93)*
	Public	-0.111	(2.42)**
	Private, non-Catholic	-0.067	(1.20)
	Private, non-religious	-0.04	(0.61)
Age in 2003		0.042	(2.09)**
Age in 2003 sq.		-0.00049	(2.18)**
Job variables			
Tenure		0.071	(7.15)***
Tenure sq.		-0.004	(4.20)***
Undergrad decision making variables			
Crime rate important ("Yes" excl)			
	No	0.017	(0.50)
	Missing	-0.006	(0.04)
Graduation rate important ("Yes" excl.)		-0.058	(1.97)**
Placement rate important ("Yes" excl.)		0.009	(0.35)
Constant		6.179	(14.73)***
Observations		3455	
R-squared		0.12	

Robust t statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 3.6: 1994 Logged Wages: Heckman two-step

		Wage eq.		*Selection eq*	
		Coefficient	**t stat**	**Coefficient**	**t stat**
ATE		0.1534	(4.07)***		
College variables					
Attended HBCU		-0.0867	(1.90)*		
College selectivity ("Middle quality public" excl.)					
	Missing	-0.0237	(0.47)		
	Low quality, public	-0.0331	(1.36)		
	Low quality, private	-0.0212	(0.71)		
	Middle quality, private	0.0071	(0.33)		
	High quality, public	0.1008	(2.82)***		
	High quality, private	0.0580	(1.32)		
	Special, public	0.3193	(2.80)***		
	Special, private	-0.1956	(1.91)*		
Parent variables					
Parent region ("Mid-east" excl.)					
	New England	-0.0093	(0.25)		
	Great Lakes	-0.0736	(2.63)***		
	Plains	-0.1120	(3.63)***		
	Southeast	-0.0639	(2.46)**		
	Southwest	-0.0222	(0.68)		
	Rocky Mountains	-0.0542	(1.13)		
	Far West	0.0044	(0.14)		
Parent age (40-45 excl.)					
	Missing	0.0197	(0.34)		
	35-40	0.0291	(0.37)		
	45-50	0.0131	(0.51)		
	50+	0.0013	(0.05)		
Mother education (BA excl.)					
	Missing	-0.1896	(1.23)		
	High School	0.0132	(0.48)		
	Some college	0.0086	(0.36)		
	MA	-0.0065	(0.21)		
	PhD/Other professional	-0.0234	(0.42)		
Father education (BA excl.)					
	Missing	-0.0265	(0.19)		
	High School	-0.0729	(2.58)***		
	Some college	-0.0054	(0.21)		
	MA	-0.0186	(0.69)		
	PhD/Other professional	-0.0612	(1.83)*		
Mom had more edu. than dad ("No" excl.)					
	Missing	0.0149	(0.53)		
	Yes	0.0080	(0.32)		
Parent marital status ("Yes" excl.)					
	No	-0.0060	(0.05)		
	Missing	0.0087	(0.35)		
Parents helped with college ("Yes" excl.)					
	No	0.0283	(1.26)		
	Missing	0.1567	(1.73)*		
Parent income (20-60k excl.)					
	0-20k	-0.0522	(2.15)**	-0.0199	(0.37)
	60k+	0.0364	(1.48)	-0.0611	(1.19)
Family variables					
Dependents aged 0-5				-0.1111	(1.62)
Dependents aged 6-12				-0.0965	(1.29)
Spouse employed				-0.1317	(1.90)*
Spouse's income				6.94E-06	(3.21)***
Spouse's income sq.				-6.91E-12	(3.34)***
Demographics					
Currently enrolled in school				-0.6281	(12.99)***
Disability		-0.0037	(0.07)		
Earned another degree/license prior to BA		0.0535	(2.86)***		
Ethnicity ("Hispanic" excl.)					
	Missing	0.0160	(0.10)	-0.4210	(1.31)
	Native American	0.0010	(0.01)	0.2850	(0.80)
	Asian	-0.0438	(0.51)	-0.2395	(1.44)
	Black	-0.0688	(1.12)	-0.0332	(0.24)
	White	-0.1410	(2.92)***	0.0261	(0.23)
SAT/ACT merged quartile ("Quartile 3" excl.)					
	Missing	-0.0261	(0.77)		
	Quartile 1	-0.0493	(1.92)*		
	Quartile 2	-0.0128	(0.53)		
	Quartile 4	0.0177	(0.63)		
High school type ("Private, Catholic" excl.)					
	Missing	-0.1495	(1.22)		
	Public	0.0314	(0.89)		
	Private, non-Catholic	0.0475	(1.07)		
	Private, non-religious	-0.0423	(0.81)		
Age in 1992		0.0765	(5.51)***		
Age in 1992 sq.		-0.0009	(4.32)***		
Undergrad decision making variables					
Crime rate important ("Yes" excl)					
	No	0.0139	(0.60)		
	Missing	-0.0948	(0.83)		
Graduation rate important ("Yes" excl.)		0.0123	(0.52)		
Placement rate important ("Yes" excl.)		-0.0652	(3.54)***		
Inverse Mills Ratio		-1.4716	(20.41)***		
Constant		4.8883	(23.19)***	0.8709	(7.76)***
Observations		3861		3861	

Robust z statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 3.7: 1997 Logged Wages: Heckman two-step

		Wage eq.		*Selection eq*	
		Coefficient	**t stat**	**Coefficient**	**t stat**
ATE		0.1110	(3.03)***		
College variables					
Attended HBCU		-0.0434	(0.86)		
College selectivity ("Middle quality public" excl.)					
	Missing	-0.0746	(1.16)		
	Low quality, public	-0.0035	(0.12)		
	Low quality, private	-0.0155	(0.42)		
	Middle quality, private	0.0024	(0.08)		
	High quality, public	0.0233	(0.56)		
	High quality, private	0.0607	(1.16)		
	Special, public	0.0049	(0.04)		
	Special, private	-0.3121	(2.38)**		
Parent variables					
Parent region ("Mid-east" excl.)					
	New England	-0.0822	(1.99)**		
	Great Lakes	-0.0454	(1.34)		
	Plains	-0.1069	(2.86)***		
	Southeast	-0.0738	(2.16)**		
	Southwest	-0.0409	(1.01)		
	Rocky Mountains	-0.0810	(1.15)		
	Far West	0.0356	(0.93)		
Parent age (40-45 excl.)					
	Missing	0.0900	(1.60)		
	35-40	-0.0224	(0.32)		
	45-50	0.0044	(0.13)		
	50+	0.0424	(1.35)		
Mother education (BA excl.)					
	Missing	0.0111	(0.08)		
	High School	0.0301	(0.94)		
	Some college	0.0224	(0.73)		
	MA	-0.0275	(0.60)		
	PhD/Other professional	0.0360	(0.60)		
Father education (BA excl.)					
	Missing	-0.2567	(2.50)**		
	High School	-0.0546	(1.73)*		
	Some college	-0.0850	(2.77)***		
	MA	-0.0330	(0.92)		
	PhD/Other professional	-0.1198	(3.03)***		
Mom had more edu. than dad ("No" excl.)					
	Missing	0.0609	(1.96)**		
	Yes	-0.0432	(1.32)		
Parent marital status ("Yes" excl.)					
	No	-0.3348	(2.35)**		
	Missing	0.0441	(1.50)		
Parents helped with college ("Yes" excl.)					
	No	0.0288	(1.08)		
	Missing	-0.0977	(1.33)		
Parent income (20-60k excl.)					
	0-20k	-0.0267	(0.93)	-0.0867	(1.64)
	60k+	0.0733	(2.32)**	-0.1355	(2.40)**
Family variables					
Dependents aged 0-6				-0.3506	(8.84)***
Dependents aged 7-12				0.0012	(0.01)
Spouse employment ("Unemployed" excl.)					
	Missing			-0.0957	(1.03)
	Employed full time			-0.1366	(1.49)
	Employed part time			4.28E-06	(3.64)***
Spouse's income				-2.57E-11	(8.69)***
Spouse's income sq.				-5.16E-01	(9.78)***
Demographics					
Currently enrolled in school				-0.521	(9.76)***
Disability		-0.2315	(3.22)***		
Earned another degree/license prior to BA		0.0108	(0.48)		
Ethnicity ("Hispanic" excl.)					
	Missing	-0.0719	(0.29)	-0.0442	(0.13)
	Native American	0.1923	(1.19)	0.1260	(0.37)
	Asian	0.2676	(2.88)***	-0.4164	(2.80)***
	Black	0.0622	(0.92)	-0.0072	(0.06)
	White	-0.0713	(1.22)	0.0465	(0.50)
SAT/ACT merged quartile ("Quartile 3" excl.)					
	Missing	-0.0315	(0.84)		
	Quartile 1	-0.0625	(2.05)**		
	Quartile 2	0.0016	(0.06)		
	Quartile 4	0.0168	(0.53)		
High school type ("Private, Catholic" excl.)					
	Missing	-0.2372	(2.05)**		
	Public	-0.0726	(2.06)**		
	Private, non-Catholic	-0.0317	(0.68)		
	Private, non-religious	-0.1050	(1.48)		
Age in 1992		0.0465	(3.08)***		
Age in 1992 sq.		-0.0007	(2.98)***		
Job variables					
Tenure		0.0036	(4.24)***		
Tenure sq.		-8.67E-06	(1.68)*		
Undergrad decision making variables					
Crime rate important ("Yes" excl)					
	No	0.0114	(0.39)		
	Missing	0.0393	(0.34)		
Graduation rate important ("Yes" excl.)		-0.0057	(0.20)		
Placement rate important ("Yes" excl.)		-0.0609	(2.52)**		
Inverse Mills Ratio		-1.9755	(16.99)***		
Constant		5.6525	(23.82)***	1.2463	(8.90)***
Observations		3623		3623	

Robust z statistics in parentheses

**significant at 10%; **significant at 5%; ***significant at 1%*

Table 3.8: 2003 Logged Wages: Heckman two-step

	Wage eq.		*Selection eq*	
	Coefficient	**t stat**	**Coefficient**	**t stat**
ATE	0.034	(0.79)		
College variables				
Attended HBCU	0.008	(0.11)		
College selectivity ("Middle quality public" excl.)				
Missing	0.032	(0.34)		
Low quality, public	-0.112	(3.54)***		
Low quality, private	-0.026	(0.53)		
Middle quality, private	0.015	(0.57)		
High quality, public	0.076	(1.90)*		
High quality, private	0.105	(2.26)**		
Special, public	0.165	(1.72)*		
Special, private	-0.190	(1.71)*		
Parent variables				
Parent region ("Mid-east" excl.)				
New England	0.015	-0.38		
Great Lakes	-0.040	-1.22		
Plains	-0.135	(3.47)***		
Southeast	-0.096	(3.33)***		
Southwest	-0.097	(2.38)**		
Rocky Mountains	-0.073	-1.16		
Far West	-0.025	-0.61		
Parent age (40-45 excl.)				
Missing	-0.004	(0.05)		
35-40	0.006	(0.07)		
45-50	-0.010	(0.30)		
50+	0.000	(0.01)		
Mother education (BA excl.)				
Missing	0.095	(0.54)		
High School	0.074	(2.24)**		
Some college	0.026	(0.84)		
MA	0.020	(0.57)		
PhD/Other professional	-0.004	(0.07)		
Father education (BA excl.)				
Missing	-0.131	(1.03)		
High School	-0.060	(1.90)*		
Some college	-0.042	(1.33)		
MA	-0.041	(1.17)		
PhD/Other professional	-0.060	(1.49)		
Mom had more edu. than dad ("No" excl.)				
Missing	0.023	(0.72)		
Yes	-0.018	(0.61)		
Parent marital status ("Yes" excl.)				
No	-0.228	(0.87)		
Missing	-0.041	(1.55)		
Parents helped with college ("Yes" excl.)				
No	-0.050	(1.73)*		
Missing	-0.013	(0.15)		
Parent income (20-60k excl.)				
0-20k	-0.030	(1.11)	-0.023	(0.35)
60k+	0.073	(2.62)***	0.052	(0.80)
Family variables				
Dependents aged 0-4			-0.091	(2.51)**
Dependents aged 5-17			-0.160	(3.98)***
Spouse employment ("Unemployed" excl.)				
Missing			-0.240	(1.94)*
Employed full time			-0.152	(1.30)
Employed part time			-0.021	(0.12)
Spouse's 1997 income			1.12E-06	(0.80)
Spouse's 1997 income sq.			-2.36E-12	(0.50)
Demographics				
Currently enrolled in school			-0.042	(0.44)
Disability	-0.173	(2.71)***		
Earned another degree/license prior to BA	0.027	(1.18)		
Ethnicity ("Hispanic" excl.)				
Missing	0.007	(0.04)	-0.128	(0.29)
Native American	0.066	(0.38)	0.051	(0.15)
Asian	0.078	(0.92)	-0.169	(0.89)
Black	-0.073	(1.03)	-0.013	(0.08)
White	-0.137	(2.43)**	-0.045	(0.34)
SAT/ACT merged quartile ("Quartile 3" excl.)				
Missing	-0.054	(1.42)		
Quartile 1	-0.099	(3.20)***		
Quartile 2	-0.054	(1.83)*		
Quartile 4	0.025	(0.75)		
High school type ("Private, Catholic" excl.)				
Missing	-0.250	(2.00)**		
Public	-0.123	(2.74)***		
Private, non-Catholic	-0.077	(1.37)		
Private, non-religious	-0.056	(0.87)		
Age	0.025	(1.32)		
Age sq.	-2.83E-04	(1.37)		
Job variables				
Tenure	0.069	(7.59)***		
Tenure sq.	-0.004	(4.81)***		
Undergrad decision making variables				
Crime rate important ("Yes" excl.)				
No	0.026	(0.83)		
Missing	-0.030	(0.24)		
Graduation rate important ("Yes" excl.)	-0.051	(1.75)*		
Placement rate important ("Yes" excl.)	0.011	(0.42)		
Inverse Mills Ratio	-1.522	(15.52)***		
Constant	6.399	(16.20)***	1.543	(8.87)***
Observations	3717		3717	

Robust z statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 3.9: 1994 Logged Wages: IV-Proportion of Women's Colleges in State

		Coefficient	t stat
ATE		0.317	(0.66)
College variables			
Attended HBCU		-0.090	(1.50)
College selectivity ("Middle quality public" excl.)			
	Missing	0.019	(0.24)
	Low quality, public	-0.032	(1.22)
	Low quality, private	-0.017	(0.23)
	Middle quality, private	-0.029	(0.47)
	High quality, public	0.100	(2.53)**
	High quality, private	0.056	(1.25)
	Special, public	0.197	(0.42)
	Special, private	-0.261	(0.79)
Parent variables			
Parent region ("Mid-east" excl.)			
	New England	-0.013	(0.30)
	Great Lakes	-0.060	(1.58)
	Plains	-0.079	(1.78)*
	Southeast	-0.061	(2.13)**
	Southwest	-0.025	(0.68)
	Rocky Mountains	-0.048	(0.82)
	Far West	0.000	0.00
Parent age (40-45 excl.)			
	Missing	0.032	(0.45)
	35-40	0.025	(0.31)
	45-50	0.022	(0.79)
	50+	0.010	(0.35)
Mother education (BA excl.)			
	Missing	-0.117	(0.88)
	High School	0.008	(0.26)
	Some college	0.018	(0.69)
	MA	-0.014	(0.38)
	PhD/Other professional	-0.064	(0.85)
Father education (BA excl.)			
	Missing	-0.085	(0.76)
	High School	-0.075	(2.51)**
	Some college	-0.001	(0.02)
	MA	-0.015	(0.51)
	PhD/Other professional	-0.056	(1.34)
Mom had more edu. than dad ("No" excl.)			
	Missing	0.020	(0.67)
	Yes	-0.003	(0.11)
Parent marital status ("Yes" excl.)			
	No	0.064	(0.53)
	Married	0.020	(0.76)
Parents helped with college ("Yes" excl.)			
	No	0.022	(0.87)
	Missing	0.199	(2.21)**
Parent income (20-60k excl.)			
	0-20k	-0.055	(2.50)**
	60k+	0.026	(1.17)
Family variables			
Spouse employed		0.012	(0.36)
Spouse's income		1.55E-06	(2.03)**
Spouse's income sq.		-1.86E-12	(2.14)**
Number of dependents		-0.037	(1.98)**
Demographics			
Earned another degree/license prior to BA		0.046	(2.23)**
Ethnicity ("Hispanic" excl.)			
	Missing	-0.058	(0.38)
	Native American	0.049	(0.40)
	Asian	-0.132	(1.64)
	Black	-0.065	(1.13)
	White	-0.126	(2.97)***
SAT/ACT merged quartile ("Quartile 3" excl.)			
	Missing	0.006	(0.17)
	Quartile 1	-0.021	(0.69)
	Quartile 2	0.014	(0.51)
	Quartile 4	0.036	(1.15)
High school type ("Private, Catholic" excl.)			
	Missing	-0.207	(1.53)
	Public	0.022	(0.54)
	Private, non-Catholic	0.048	(0.91)
	Private, non-religious	-0.023	(0.37)
Age in 1992		0.070	(4.04)***
Age in 1992 sq.		-0.001	(3.33)***
Undergrad decision making variables			
Crime rate important ("Yes" excl.)			
	No	0.002	(0.06)
	Missing	-0.161	(1.32)
Graduation rate important ("Yes" excl.)		0.009	(0.32)
Placement rate important ("Yes" excl.)		-0.065	(3.10)***
Constant		4.783	(18.74)***
Observations		3007.000	

Robust z statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 3.10: 1997 Logged Wages: IV-Proportion of Women's Colleges in State

		Coefficient	t stat
ATE		-0.089	(0.13)
College variables			
Attended HBCU		3.00E-03	(0.04)
College selectivity ("Middle quality public" excl.)			
	Missing	-0.020	(0.19)
	Low quality, public	-0.011	(0.33)
	Low quality, private	5.00E-03	(0.04)
	Middle quality, private	0.018	(0.17)
	High quality, public	0.058	(1.14)
	High quality, private	0.089	(1.43)
	Special, public	0.278	(0.42)
	Special, private	-0.091	(0.19)
Parent variables			
Parent region ("Mid-east" excl.)			
	New England	-0.019	(0.34)
	Great Lakes	-0.008	(0.15)
	Plains	-0.062	(0.85)
	Southeast	-0.046	(1.06)
	Southwest	0.022	(0.45)
	Rocky Mountains	-0.068	(0.87)
	Far West	0.063	(0.97)
Parent age (40-45 excl.)			
	Missing	0.203	(2.41)**
	35-40	0.029	(0.31)
	45-50	0.024	(0.65)
	50+	0.062	(1.71)*
Mother education (BA excl.)			
	Missing	0.110	(0.58)
	High School	0.076	(1.76)*
	Some college	0.071	(1.62)
	MA	-0.042	(0.71)
	PhD/Other professional	0.049	(0.61)
Father education (BA excl.)			
	Missing	-0.287	(1.91)*
	High School	-0.072	(1.84)*
	Some college	-0.091	(2.34)**
	MA	-0.048	(1.15)
	PhD/Other professional	-0.067	(1.41)
Mom had more edu. than dad ("No" excl.)			
	Missing	0.081	(2.21)**
	Yes	-0.025	(0.58)
Parent marital status ("Yes" excl.)			
	No	-0.404	(1.59)
	Missing	0.057	(1.52)
Parents helped with college ("Yes" excl.)			
	No	0.004	(0.11)
	Missing	-0.048	(0.55)
Parent income (20-60k excl.)			
	0-20k	-0.056	(2.04)**
	60k+	0.019	(0.60)
Family variables			
Spouse employment ("Unemployed" excl.)			
	Missing	-0.014	(0.27)
	Employed full-time	-0.101	(2.01)**
Spouse's income		4.47E-06	(3.41)***
Spouse's income sq.		-2.03E-11	(9.44)***
Number of dependents		-0.080	(4.37)***
Demographics			
Earned another degree/license prior to BA		0.012	(0.42)
Ethnicity ("Hispanic" excl.)			
	Missing	-0.229	(0.93)
	Native American	0.116	(0.79)
	Asian	0.150	(1.66)*
	Black	0.121	(1.87)*
	White	-0.031	(0.54)
	Other	0.113	(0.34)
SAT/ACT merged quartile ("Quartile 3" excl.)			
	Missing	-0.051	(1.15)
	Quartile 1	-0.058	(1.43)
	Quartile 2	-0.014	(0.40)
	Quartile 4	-0.006	(0.17)
High school type ("Private, Catholic" excl.)			
	Missing	-0.228	(1.56)
	Public	-0.130	(2.43)**
	Private, non-Catholic	-0.077	(1.06)
	Private, non-religious	-0.211	(2.12)**
Age in 1992		0.059	(2.82)***
Age in 1992 sq.		-0.001	(2.73)***
Job variables			
Tenure		0.006	(5.41)***
Tenure sq.		-1.56E-05	(2.33)**
Undergrad decision making variables			
Crime rate important ("Yes" excl)			
	No	0.003	(0.08)
	Missing	-0.053	(0.43)
Graduation rate important ("Yes" excl.)		-0.010	(0.29)
Placement rate important ("Yes" excl.)		-0.059	(2.27)**
Constant		5.273	(17.79)***
Observations		3150	

Robust z statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 3.11: 2003 Logged Wages: IV-Proportion of Women's Colleges in State

		Coefficient	t stat
ATE		0.059	(0.11)
College variables			
Attended HBCU		-0.005	(0.05)
College selectivity ("Middle quality public" excl.)			
	Missing	0.011	(0.09)
	Low quality, public	-0.112	(3.36)***
	Low quality, private	-0.056	(0.54)
	Middle quality, private	0.006	(0.08)
	High quality, public	0.097	(2.26)**
	High quality, private	0.082	(1.62)
	Special, public	0.171	(0.31)
	Special, private	-0.152	(0.45)
Parent variables			
Parent region ("Mid-east" excl.)			
	New England	0.016	(0.39)
	Great Lakes	-0.048	(1.11)
	Plains	-0.128	(2.27)**
	Southeast	-0.092	(2.85)***
	Southwest	-0.1	(2.29)**
	Rocky Mountains	-0.095	(1.44)
	Far West	-0.052	(0.89)
Parent age (40-45 excl.)			
	Missing	-0.013	(0.17)
	35-40	0.003	(0.03)
	45-50	-0.018	(0.56)
	50+	-0.016	(0.50)
Mother education (BA excl.)			
	Missing	0.145	(0.98)
	High School	0.061	(1.77)*
	Some college	0.018	(0.54)
	MA	0.028	(0.73)
	PhD/Other professional	-0.027	(0.39)
Father education (BA excl.)			
	Missing	-0.157	(1.15)
	High School	-0.041	(1.22)
	Some college	-0.028	(0.85)
	MA	-0.031	(0.89)
	PhD/Other professional	-0.049	(1.10)
Mom had more edu. than dad ("No" excl.)			
	Missing	0.023	(0.72)
	Yes	0.006	(0.19)
Parent marital status ("Yes" excl.)			
	No	-0.17	(0.58)
	Missing	-0.032	(1.09)
Parents helped with college ("Yes" excl.)			
	No	-0.047	(1.51)
	Missing	-0.016	(0.15)
Parent income (20-60k excl.)			
	0-20k	-0.033	(1.36)
	60k+	0.074	(2.79)***
Family variables			
Spouse employment ("Unemployed" excl.)			
	Missing	-0.18	(4.34)***
	Employed full time	-0.149	(3.93)***
	Employed part-time	-0.018	(0.30)
Dependents age 0-17		-0.073	(5.99)***
Demographics			
Earned another degree/license prior to BA		0.025	(1.07)
Ethnicity ("Hispanic" excl.)			
	Missing	-0.039	(0.21)
	Native American	0.01	(0.07)
	Asian	0.021	(0.23)
	Black	-0.07	(1.06)
	White	-0.165	(3.08)***
	Other	-0.171	(0.63)
SAT/ACT merged quartile ("Quartile 3" excl.)			
	Missing	-0.049	(1.27)
	Quartile 1	-0.078	(2.28)**
	Quartile 2	-0.042	(1.40)
	Quartile 4	0.011	(0.30)
High school type ("Private, Catholic" excl.)			
	Missing	-0.294	(1.95)*
	Public	-0.11	(2.27)**
	Private, non-Catholic	-0.065	(1.07)
	Private, non-religious	-0.038	(0.55)
Age		0.042	(2.10)**
Age sq.		-0.00049	(2.16)**
Job variables			
	Tenure	0.071	(7.22)***
	Tenure sq.	-0.004	(4.20)***
Undergrad decision making variables			
Crime rate important ("Yes" excl.)			
	No	0.017	(0.51)
	Missing	-0.003	(0.02)
Graduation rate important ("Yes" excl.)		-0.057	(1.80)*
Placement rate important ("Yes" excl.)		0.01	(0.37)
Constant		6.169	(14.67)***
Observations		3455	

Robust z statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 3.12: Propensity Score Logit Model

		Coefficient	t stat
Pct. women's colleges in parent's state		11.323	(5.01)***
Parent variables			
Parent region ("Mid-east" excl.)			
	New England	-0.605	(2.16)**
	Great Lakes	-0.303	(1.24)
	Plains	-1.071	(3.10)***
	Southeast	-0.232	(1.24)
	Southwest	0.105	(0.37)
	Rocky Mountains	-0.407	(0.79)
	Far West	-0.43	(1.40)
Parent age (40-45 excl.)			
	Missing	0.485	(1.32)
	45-50	-0.2	(0.81)
	50+	-0.028	(0.17)
Mother education (BA excl.)			
	Missing	0.645	(0.72)
	High School	0.264	(1.20)
	Some college	0.123	(0.57)
	MA	0.229	(0.86)
	PhD/Other professional	0.297	(0.70)
Father education (BA excl.)			
	Missing	-0.477	(0.56)
	High School	-0.111	(0.52)
	Some college	-0.132	(0.61)
	MA	0.057	(0.24)
	PhD/Other professional	-0.064	(0.21)
Mom had more edu. than dad ("No" excl.)			
	Missing	0.1	(0.48)
	Yes	0.26	(1.35)
Parent marital status ("Yes" excl.)			
	Missing	-0.14	(0.91)
	Married	-0.533	(0.58)
Parents helped with college ("Yes" excl.)			
	No	0.191	(1.06)
	Missing	0.605	(1.25)
Parent income (20-60k excl.)			
	0-20k	-0.118	(0.73)
	60k+	0.02	(0.12)
Demographics			
Earned another degree/license prior to BA		0.447	(3.10)***
Speaks a language other than english		0.131	(0.93)
Ethnicity ("Hispanic" excl.)			
	Missing	-0.471	(0.42)
	Native American	-0.073	(0.09)
	Asian	0.557	(1.17)
	Black	-0.205	(0.49)
	White	-0.053	(0.16)
	Other	0.713	(0.58)
SAT/ACT merged quartile ("Quartile 3" excl.)			
	Missing	0.635	(2.51)**
	Quartile 1	0.356	(1.68)*
	Quartile 2	0.031	(0.14)
	Quartile 4	-0.016	(0.07)
High school type ("Private, Catholic" excl.)			
	Missing	-0.155	(0.23)
	Public	-0.756	(3.12)***
	Private, non-Catholic	-0.539	(1.66)*
	Private, non-religious	-0.82	(1.84)*
Age in 1992		-0.022	(0.25)
Age in 1992 sq.		0.001	(0.58)
Undergrad decision making variables			
Crime rate important ("Yes" excl)			
	No	-0.101	(0.54)
	Missing	-0.807	(1.05)
Graduation rate important ("Yes" excl.)		-0.41	(2.26)**
Placement rate important ("Yes" excl.)		-0.462	(2.95)***
Constant		-1.935	(1.38)
Observations		3949	

Absolute value of z statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 3.13: Naive Least Squares with Propensity Score

	94 ln(wages)	**97 ln(wages)**	**03 ln(wages)**
ATE	0.183	0.134	0.026
	(4.41)***	(3.26)***	(0.65)
Pr(Treatment)	1.003	0.51	0.361
	(5.38)***	(2.05)**	(1.89)*
Constant	5.776	6.081	6.59
	(382.84)***	(309.26)***	(372.93)***
Observations	3016	3171	3558
R-squared	0.02	0.01	0.01

Robust t statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table 3.14: Nearest Neighbor Matching Estimates: logged weekly wages

	Full sample	**Bottom 50%**	**Top 50%**	**Less selective colleges**	**More selective colleges**
1994					
ATT	**0.1679**	-0.0045	**0.1318**	**0.1621**	**0.1285**
Std error	0.0094	0.0204	0.0112	0.0159	0.0375
BS std. error	0.0472	0.0669	0.0344	0.0585	0.0951
Treated obs	207	77	130	133	48
Untreated obs	2,800	1,466	1,334	1,802	993
Total obs	3,007	1,543	1,464	1,935	1,041
1997					
ATT	**0.1346**	0.0674	**0.0504**	**0.1888**	**0.1833**
Std error	0.0157	0.0331	0.0102	0.0173	0.0467
BS std. error	0.0528	0.1085	0.0260	0.0627	0.1777
Treated obs	217	68	149	144	42
Untreated obs	2,949	1,259	1,690	1,909	1,000
Total obs	3,166	1,327	1,839	2,053	1,042
2003					
ATT	0.0340	0.0273	-0.0112	**0.0643**	**-0.2164**
Std error	0.0218	0.0235	0.0099	0.0162	0.0286
BS std. error	0.0451	0.0704	0.0301	0.0593	0.0976
Treated obs	240	96	144	152	52
Untreated obs	3,286	1,570	1,716	2,106	1,139
Total obs	3,526	1,666	1,860	2,258	1,191

Std errors are subsampled using 5,000 repetitions
BS (bootstrapped) std. errors used 1,000 repetitions
Bolded numbers indicate those statistically significant at the 5% level

Table 3.15: Regression Adjustment and Propensity Score Weighting Summary

1994	
ATT estimate	0.15
	(6.45)***
Observations	3007
R squared	0.31
1997	
ATT estimate	0.182
	(6.94)***
Observations	3153
R squared	0.18
2003	
ATT estimate	0.004
	(0.16)
Observations	3463
R squared	0.17

Table 3.16: 1994 Logged Wages: Regression Adjustment & Weighting

	Coefficient	t stat
ATT	0.15	(6.45)***
College variables		
Attended HBCU	-0.089	(1.20)
College selectivity ("Middle quality public" excl.)		
Missing	0.049	(0.50)
Low quality, public	-0.003	(0.08)
Low quality, private	0.07	(1.52)
Middle quality, private	-0.004	(0.13)
High quality, public	0.128	(2.24)**
High quality, private	0.092	(1.54)
Special, public	0.036	(0.59)
Special, private	0.101	(0.37)
Parent variables		
Parent region ("Mid-east" excl.)		
New England	-0.006	(0.14)
Great Lakes	-0.07	(1.74)*
Plains	-0.074	(1.21)
Southeast	-0.057	(1.69)*
Southwest	-0.015	(0.33)
Rocky Mountains	-0.098	(1.04)
Far West	-0.024	(0.46)
Parent age (40-45 excl.)		
Missing	0.097	(1.20)
35-40	0.058	(0.48)
45-50	0.004	(0.10)
50+	0.017	(0.39)
Mother education (BA excl.)		
Missing	-0.129	(0.59)
High School	0.067	(1.72)*
Some college	0.102	(2.59)***
MA	0.052	(1.07)
PhD/Other professional	-0.008	(0.11)
Father education (BA excl.)		
Missing	-0.04	(0.20)
High School	-0.041	(1.12)
Some college	-0.033	(0.87)
MA	-0.033	(0.78)
PhD/Other professional	-0.109	(1.99)**
Parent marital status ("Yes" excl.)		
Missing	-0.045	(0.23)
Married	-0.01	(0.36)
Parents helped with college ("Yes" excl.)		
No	0.018	(0.57)
Missing	0.237	(2.30)**
Parent income (20-60k excl.)		
0-20k	-0.082	(2.74)***
60k+	0.013	(0.41)
Family variables		
Spouse employed	-0.023	(0.52)
Spouse's income	1.40E-06	(1.46)
Spouse's income sq.	1.14E-12	(1.08)
Number of dependents	-0.045	(2.50)**
Demographics		
Earned another degree/license prior to BA	0.025	(0.98)
Ethnicity ("Hispanic" excl.)		
Missing	-0.075	(0.40)
Native American	-0.001	(0.01)
Asian	-0.066	(0.74)
Black	-0.122	(1.55)
White	-0.161	(2.69)***
SAT/ACT merged quartile ("Quartile 3" excl.)		
Missing	0.016	(0.34)
Quartile 1	-0.013	(0.34)
Quartile 2	0.012	(0.30)
Quartile 4	0.023	(0.51)
High school type ("Private, Catholic" excl.)		
Missing	-0.079	(0.65)
Public	-0.024	(0.54)
Private, non-Catholic	-0.016	(0.27)
Private, non-religious	-0.083	(1.00)
Age in 1992	0.072	(3.77)***
Age in 1992 sq.	-0.001	(3.06)***
Undergrad decision making variables		
Crime rate important ("Yes" excl.)		
No	0.014	(0.41)
Missing	-0.388	(2.68)***
Graduation rate important ("Yes" excl.)	0.001	(0.04)
Placement rate important ("Yes" excl.)	-0.077	(2.72)***
Constant	4.826	(16.11)***
Observations	3007	
R-squared	0.31	

Absolute value of t statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*
Estimated model also includes the mean-differenced versions of all the variables shown

Table 3.17: 1997 Logged Wages: Regression Adjustment & Weighting

		Coefficient	t stat
ATT		0.182	(6.94)***
College variables			
	Attended HBCU	0.011	(0.12)
College selectivity ("Middle quality public" excl.)			
	Missing	0.012	(0.09)
	Low quality, public	-0.0001	(0.00)
	Low quality, private	0.044	(0.74)
	Middle quality, private	0.028	(0.70)
	High quality, public	0.064	(0.88)
	High quality, private	0.117	(1.55)
	Special, public	-0.382	(4.83)***
	Special, private	-0.685	(6.38)***
Parent variables			
Parent region ("Mid-east" excl.)			
	New England	-0.06	(1.11)
	Great Lakes	0.02	(0.40)
	Plains	-0.097	(1.29)
	Southeast	-0.085	(2.03)**
	Southwest	-0.064	(1.11)
	Rocky Mountains	-0.075	(0.66)
	Far West	0.033	(0.52)
Parent age (40-45 excl.)			
	Missing	0.225	(2.15)**
	35-40	0.152	(0.95)
	45-50	0.061	(1.06)
	50+	0.133	(2.37)**
Mother education (BA excl.)			
	Missing	0.304	(1.26)
	High School	0.112	(2.18)**
	Some college	0.088	(1.78)*
	MA	0.048	(0.78)
	PhD/Other professional	0.009	(0.10)
Father education (BA excl.)			
	Missing	-0.287	(1.30)
	High School	-0.057	(1.20)
	Some college	-0.079	(1.63)
	MA	-0.06	(1.12)
	PhD/Other professional	-0.015	(0.22)
Mom had more edu. than dad ("No" excl.)			
	Missing	0.029	(0.60)
	Yes	-0.015	(0.34)
Parent marital status ("Yes" excl.)			
	Missing	-0.378	(1.54)
	Married	0.001	(0.03)
Parents helped with college ("Yes" excl.)			
	No	0.015	(0.37)
	Missing	-0.063	(0.52)
Parent income (20-60k excl.)			
	0-20k	-0.076	(2.10)**
	60k+	-0.025	(0.62)
Family variables			
Spouse employment ("unemployed" excl.)			
	Missing	-0.129	(1.88)*
	Employed full time	-0.103	(1.47)
Spouse's income		9.02E-07	(0.73)
Spouse's income sq.		-2.24E-11	(3.79)***
Number of dependents		-0.045	(1.98)**
Demographics			
Earned another degree/license prior to BA		0.038	(1.17)
Ethnicity ("Hispanic" excl.)			
	Missing	0.06	(0.26)
	Native American	0.123	(0.64)
	Asian	0.198	(1.72)*
	Black	0.064	(0.66)
	White	-0.091	(1.22)
	Other	-0.02	(0.04)
SAT/ACT merged quartile ("Quartile 3" excl.)			
	Missing	-0.115	(2.02)**
	Quartile 1	-0.057	(1.17)
	Quartile 2	-0.015	(0.30)
	Quartile 4	-0.045	(0.79)
High school type ("Private, Catholic" excl.)			
	Missing	-0.099	(0.62)
	Public	-0.118	(2.03)**
	Private, non-Catholic	-0.072	(0.94)
	Private, non-religious	-0.158	(1.53)
Age in 1992		0.053	(2.18)**
Age in 1992 sq.		-0.001	(2.07)**
Undergrad decision making variables			
Crime rate important ("Yes" excl.)			
	No	0.068	(1.59)
	Missing	-0.171	(0.96)
Graduation rate important ("Yes" excl.)		-0.105	(2.57)**
Placement rate important ("Yes" excl.)		-0.004	(0.13)
Constant		5.574	(14.56)***
Observations		3153	
R-squared		0.18	

Absolute value of t statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*
Estimated model also includes the mean-differenced versions of all the variables shown

Table 3.18: 2003 Logged Wages: Regression Adjustment & Weighting

		Coefficient	t stat
ATT		0.004	(0.16)
College variables			
	Attended HBCU	0.023	(0.28)
College selectivity ("Middle quality public" excl.)			
	Missing	0.051	(0.41)
	Low quality, public	-0.102	(2.51)**
	Low quality, private	-0.02	(0.37)
	Middle quality, private	0.034	(0.93)
	High quality, public	0.112	(1.72)*
	High quality, private	0.084	(1.27)
	Special, public	-0.022	(0.10)
Parent variables			
Parent region ("Mid-east" excl.)			
	New England	0.035	(0.72)
	Great Lakes	-0.032	(0.71)
	Plains	-0.083	(1.20)
	Southeast	-0.096	(2.59)***
	Southwest	-0.086	(1.68)*
	Rocky Mountains	-0.057	(0.53)
	Far West	0.002	(0.04)
Parent age (40-45 excl.)			
	Missing	0.119	(1.33)
	35-40	0.068	(0.46)
	45-50	0.032	(0.61)
	50+	0.028	(0.55)
Mother education (BA excl.)			
	Missing	0.318	(1.59)
	High School	0.127	(2.79)***
	Some college	0.049	(1.11)
	MA	0.12	(2.20)**
	PhD/Other professional	0.009	(0.10)
Father education (BA excl.)			
	Missing	-0.237	(1.25)
	High School	0.003	(0.08)
	Some college	0.049	(1.12)
	MA	-0.007	(0.15)
	PhD/Other professional	0.008	(0.14)
Mom had more edu. than dad ("No" excl.)			
	Missing	-0.041	(0.96)
	Yes	-0.026	(0.65)
Parent marital status ("Yes" excl.)			
	Missing	-0.213	(1.05)
	Married	-0.03	(0.84)
Parents helped with college ("Yes" excl.)			
	No	-0.042	(1.17)
	Missing	0.023	(0.22)
Parent income (20-60k excl.)			
	0-20k	-0.075	(2.27)**
	60k+	0.068	(1.91)*
Family variables			
Spouse employment ("Unemployed" excl.)			
	Missing	-0.164	(2.83)***
	Employed full time	-0.165	(3.04)***
Dependents age 0-17		-0.087	(6.24)***
Demographics			
Earned another degree/license prior to BA		0.021	(0.72)
Ethnicity ("Hispanic" excl.)			
	Missing	0.157	(0.72)
	Native American	-0.274	(1.57)
	Asian	0.026	(0.27)
	Black	-0.102	(1.18)
	White	-0.214	(3.33)***
	Other	0.15	(0.38)
SAT/ACT merged quartile ("Quartile 3" excl.)			
	Missing	-0.104	(2.05)**
	Quartile 1	-0.107	(2.42)**
	Quartile 2	-0.083	(1.85)*
	Quartile 4	-0.039	(0.77)
High school type ("Private, Catholic" excl.)			
	Missing	-0.248	(1.84)*
	Public	-0.138	(2.73)***
	Private, non-Catholic	-0.063	(0.94)
	Private, non-religious	-0.127	(1.35)
Age in 2003		0.076	(2.78)***
Age in 2003 sq.		-0.001	(2.81)***
Undergrad decision making variables			
Crime rate important ("Yes" excl.)			
	No	-0.04	(1.04)
	Missing	-0.094	(0.59)
Graduation rate important ("Yes" excl.)		-0.037	(1.02)
Placement rate important ("Yes" excl.)		0.005	(0.17)
Constant		5.674	(9.98)***
Observations		3463	0.17
R-squared		0.17	

Absolute value of t statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*
Estimated model also includes the mean-differenced versions of all the variables shown

APPENDIX A

CHAPTER 1 DATA NOTES

A.1 Notes on the Dataset Constructed for Chapter 1

Demographic data by county, published by the Bureau of the Census in 1920 and 1930, is used to control for the number of black, illiterate and foreign born in and near a city. The lower panel of Table 1 provides summary statistics for these variables. Controlling for demographic trends will be important, since the foreign born and black populations generally had much higher mortality rates than the native white population, and were targeted by some of the social programs aimed to reduce child mortality (Lindenmeyer 1995). There was wide variation in the demographics between counties, with the populations over 40 percent black in some counties and populations over 45 percent foreign born in other counties in some years. The percentage of rural people in a county also varied widely, with counties such as Orleans or Suffolk (New Orleans and Boston, respectively) being measured as entirely urban, but others being a quarter or more rural. The number of illiterate people as a percentage of the county population varied less between cities and years, but enough so that it will important to control for it in the empirical specification.

To control for average income and income distributions, which the Children's Bureau initially believed were so crucial in determining child mortality rates, two measures are used. First, the number of tax returns filed as a share of the population in a year helps control for the number of people in a city who were part of the upper tail of the income distribution. This gives the number of families in a city with incomes above $5,000 (about $60,000 in 2007 dollars), and individuals with incomes over $2,000 in a city. After controlling for a measure of average income, as below, increases in the share of the population filing tax returns would be associated with lower shares of income for the population that was not earning enough to pay income taxes.

Average annual earnings in the manufacturing sector, compiled from the biannual Census of Manufactures were also included to help control for the overall wealth of a city. State per capita income, estimated by Robert Martin (1939), was used to help interpolate the missing years.[1] The interpolation formula used was $MW_{i,t} = SPCI_t \left(\frac{1}{2} \frac{MW_{i,t-1}}{SPCI_{t-1}} + \frac{1}{2} \frac{MW_{i,t+1}}{SPCI_{t+1}} \right)$, where $SPCI_t$ is state per capita income in year t. Average annual earnings per worker, calculated by dividing the average annual earning in manufacturing by the average number of wage earners employed, and the percentage of workers in polluting industries is given in middle panel of Table 1. One potential problem with using average annual earnings to measure the average wages in the different areas is that they may be highly correlated with the amount of pollution in that area (Ruhm 2000). For this reason I look at the number of persons employed in each industry, separating polluting industries from non-polluting industries. I then count the number of workers in polluting industries such as steel, coal, automotive, leather, rubber, smelting and wood pulp, and include this number in the estimation to both control for and test the impact of the extent of industrial pollution, a severe problem in many cities in the early 1900s. For a list of those industries classified as "polluting," see Table A1. The 1931 and 1933 Census of Manufactures lack city by industry level data, so estimates from a linear trendline will be included in the estimation. Although this will miss the variation between years, it should still pick up the variation between cities. Because of the large drop off in manufacturing jobs between 1927 and 1929, for some cities the trendline estimated negative values. For these, I set the observation to zero.

[1] Martin (1939) does not give a good description of how he came to his estimates. Fishback and Kachanovskaya (2009) ran regressions for each state with the BEA state income data as a function of the Martin data without an intercept over the period from 1929 to 1938 when the two sets of series. The R-squareds from each of the regressions were all above 0.98. When they ran correlations of the growth rates for the overlap periods, they are all over 0.6 and most are over 0.9.

Table A.1: List of Industries Classified as Polluting

Industries classified as "heavy" and "polluting"
Heavy
Brass, bronze and other nonferrous alloys, and manufactures of these alloys and of copper
Copper, tin and sheet-iron work, including galvanized-iron work
Forgings, iron and steel, not made in steel works or rolling mills
Foundry and machine shop products
Iron and steel: Blast furnaces
Iron and steel: Cast iron
Iron and steel: processed
Iron and steel: Steel works and rolling mills
Leather: Tanned, curried and finished
Motor vehicle bodies and motor vehicle parts
Motor vehicles, not including motorcycles
Rubber goods, other than tires or inner tubes
Rubber tires and inner tubes
Smelting and refining, metals other than gold, silver or platinum
Other polluting industries
Belting, leather
Lumber and Timber, not elsewhere classified
Lumber, planing-mill products
Paper and wood pulp
Tanning materials, natural dyestuffs, mordants and assistants

APPENDIX B

CHAPTER 3 FIRST STAGE IV RESULTS

B.1 First Stage IV Results for Regressions in Tables 3.9, 3.10 and 3.11

Table B.1: 1994 Logged Wages: IV First Stage Results

	Coefficient	**t stat**
Prop. of Women's Colleges	0.6725	(3.85)***
College variables		
Attended HBCU	-0.0656	(1.99)**
College selectivity ("Middle quality public" excl.)		
Missing	0.1076	(1.93)*
Low quality, public	-0.0067	(0.72)
Low quality, private	0.1308	(5.05)***
Middle quality, private	0.1203	(8.65)***
High quality, public	-0.0014	(0.11)
High quality, private	-0.0098	(0.70)
Special, public	0.9500	(50.22)***
Special, private	0.6372	(4.77)***
Parent variables		
Parent region ("Mid-east" excl.)		
New England	-0.0249	(0.93)
Great Lakes	-0.0224	(1.28)
Plains	-0.0414	(2.28)**
Southeast	-0.0125	(0.77)
Southwest	0.0218	(0.98)
Rocky Mountains	-0.0089	(0.32)
Far West	-0.0266	(1.50)
Parent age (40-45 excl.)		
Missing	0.0166	(0.40)
35-40	0.0221	(0.54)
45-50	-0.0001	(0.00)
50+	0.0051	(0.41)
Mother education (BA excl.)		
Missing	0.0031	(0.03)
High School	-0.0082	(0.59)
Some college	0.0006	(0.05)
MA	0.0104	(0.65)
PhD/Other professional	0.0419	(1.23)
Father education (BA excl.)		
Missing	-0.0109	(0.10)
High School	0.0064	(0.46)
Some college	0.0034	(0.27)
MA	-0.0044	(0.31)
PhD/Other professional	-0.0211	(1.27)
Mom had more edu. than dad ("No" excl.)		
Missing	0.0026	(0.17)
Yes	0.0026	(0.19)
Parent marital status ("Yes" excl.)		
Missing	-0.0942	(2.32)**
Married	-0.0053	(0.39)
Parents helped with college ("Yes" excl.)		
No	-0.0012	(0.11)
Missing	-0.0101	(0.28)
Parent income (20-60k excl.)		
0-20k	-0.0003	(0.03)
60k+	-0.0027	(0.24)
Family variables		
Spouse employed	-0.0013	(0.08)
Spouse's income	8.24E-08	(0.20)
Spouse's income sq.	3.10E-13	(0.59)
Number of dependents	-0.0205	(2.24)**
Demographics		
Earned another degree/license prior to BA	0.0054	(0.50)
Ethnicity ("Hispanic" excl.)		
Missing	-0.0707	(2.05)**
Native American	0.0033	(0.07)
Asian	0.0266	(0.76)
Black	0.0259	(0.75)
White	-0.0079	(0.34)
SAT/ACT merged quartile ("Quartile 3" excl.)		
Missing	-0.0011	(0.06)
Quartile 1	0.0160	(1.16)
Quartile 2	0.0075	(0.61)
Quartile 4	0.0071	(0.54)
High school type ("Private, Catholic" excl.)		
Missing	-0.0370	(0.51)
Public	-0.0219	(0.92)
Private, non-Catholic	-0.0470	(1.63)
Private, non-religious	-0.0356	(1.21)
Age in 1992	0.0107	(1.03)
Age in 1992 sq.	-0.0001	(0.53)
Undergrad decision making variables		
Crime rate important ("Yes" excl.)		
No	0.0097	(0.70)
Missing	0.0480	(0.68)
Graduation rate important ("Yes" excl.)	-0.0191	(1.24)
Placement rate important ("Yes" excl.)	-0.0082	(0.76)
Constant	-0.1547	(0.97)
Observations	3007	
Partial R^2 of excluded instrument	0.0063	

Absolute value of t statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table B.2: 1997 Logged Wages: IV First Stage Results

	Coefficient	t stat
Prop. of Women's Colleges	0.7014	(4.2)***
College variables		
Attended HBCU	-0.0820	(3.04)***
College selectivity ("Middle quality public" excl.)		
Missing	0.1081	(2.01)**
Low quality, public	-0.0035	(0.41)
Low quality, private	0.1353	(5.37)***
Middle quality, private	0.1313	(9.66)***
High quality, public	0.0064	(0.44)
High quality, private	-0.0051	(0.41)
Special, public	0.9602	(53.26)***
Special, private	0.6569	(4.64)***
Parent variables		
Parent region ("Mid-east" excl.)		
New England	-0.0604	(2.37)**
Great Lakes	-0.0317	(1.83)*
Plains	-0.0726	(4.39)***
Southeast	-0.0220	(1.40)
Southwest	0.00007	(0.00)
Rocky Mountains	-0.0258	(0.93)
Far West	-0.0438	(2.53)*
Parent age (40-45 excl.)		
Missing	0.0694	(1.54)
35-40	0.0595	(1.27)
45-50	0.0034	(0.28)
50+	0.0067	(0.57)
Mother education (BA excl.)		
Missing	0.0090	(0.11)
High School	-0.0046	(0.35)
Some college	0.0003	(0.03)
MA	0.0067	(0.43)
PhD/Other professional	0.0192	(0.62)
Father education (BA excl.)		
Missing	-0.0119	(0.16)
High School	0.0057	(0.43)
Some college	0.0012	(0.10)
MA	-0.0056	(0.41)
PhD/Other professional	-0.0017	(0.09)
Mom had more edu. than dad ("No" excl.)		
Missing	0.0024	(0.17)
Yes	0.0179	(1.31)
Parent marital status ("Yes" excl.)		
Missing	-0.0995	(2.54)**
Married	0.0041	(0.33)
Parents helped with college ("Yes" excl.)		
No	0.0056	(0.48)
Missing	0.0147	(0.35)
Parent income (20-60k excl.)		
0-20k	-0.0082	(0.80)
60k+	-0.0025	(0.23)
Family variables		
Spouse employment ("unemployed" excl.)		
Missing	-0.0037	(0.20)
Employed full time	0.0010	(0.05)
Spouse's income	4.81E-09	(0.01)
Spouse's income sq.	2.03E-13	(0.26)
Number of dependents	-0.0018	(0.23)
Demographics		
Earned another degree/license prior to BA	0.00573	(0.55)
Ethnicity ("Hispanic" excl.)		
Missing	-0.0950	(2.74)***
Native American	0.0059	(0.12)
Asian	0.0217	(0.65)
Black	-0.0043	(0.13)
White	-0.0065	(0.28)
SAT/ACT merged quartile ("Quartile 3" excl.)		
Missing	0.0254	(1.43)
Quartile 1	0.0353	(2.7)***
Quartile 2	0.0060	(0.54)
Quartile 4	0.0093	(0.75)
High school type ("Private, Catholic" excl.)		
Missing	-0.0837	(1.40)
Public	-0.0396	(1.59)
Private, non-Catholic	-0.0651	(2.17)**
Private, non-religious	-0.0680	(2.31)*
Age in 1992	0.0071	(0.71)
Age in 1992 sq.	-0.0001	(0.46)
Job variables		
Tenure	3.95E-05	(0.08)
Tenure sq.	2.65E-06	(0.66)
Undergrad decision making variables		
Crime rate important ("Yes" excl)		
No	0.0062	(0.47)
Missing	0.0473	(0.72)
Graduation rate important ("Yes" excl.)	-0.0163	(1.10)
Placement rate important ("Yes" excl.)	-0.0081	(0.77)
Constant	-0.0757	(0.50)
Observations	3144	
Partial R^2 of excluded instrument	0.0071	

Absolute value of t statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

Table B.3: 1997 Logged Wages: IV First Stage Results

	Coefficient	t stat
Prop. of Women's Colleges	0.5995	(3.82)***
College variables		
Attended HBCU	-0.0763	(2.76)***
College selectivity ("Middle quality public" excl.)		
Missing	0.1225	(2.11)**
Low quality, public	-0.0041	(0.52)
Low quality, private	0.1440	(5.69)***
Middle quality, private	0.1310	(10.11)***
High quality, public	-0.0004	(0.04)
High quality, private	0.0068	(0.48)
Special, public	0.9650	(62.13)***
Special, private	0.5765	(4.68)***
Parent variables		
Parent region ("Mid-east" excl.)		
New England	-0.0496	(2.08)**
Great Lakes	-0.0246	(1.52)
Plains	-0.0547	(3.34)***
Southeast	-0.0147	(1.01)
Southwest	0.0118	(0.58)
Rocky Mountains	-0.0073	(0.27)
Far West	-0.0349	(2.11)**
Parent age (40-45 excl.)		
Missing	0.0223	(0.54)
35-40	0.0209	(0.53)
45-50	-0.0043	(0.37)
50+	0.0007	(0.06)
Mother education (BA excl.)		
Missing	0.0294	(0.39)
High School	-0.0036	(0.28)
Some college	0.0033	(0.29)
MA	0.0155	(1.03)
PhD/Other professional	0.0256	(0.91)
Father education (BA excl.)		
Missing	-0.0325	(0.45)
High School	0.0102	(0.79)
Some college	0.0015	(0.13)
MA	-0.0011	(0.08)
PhD/Other professional	0.0014	(0.08)
Mom had more edu. than dad ("No" excl.)		
Missing	0.0047	(0.34)
Yes	0.0127	(0.98)
Parent marital status ("Yes" excl.)		
Missing	-0.1044	(2.88)***
Married	-0.0094	(0.77)
Parents helped with college ("Yes" excl.)		
No	0.0079	(0.71)
Missing	0.0481	(1.11)
Parent income (20-60k excl.)		
0-20k	-0.0003	(0.03)
60k+	0.0010	(0.10)
Family variables		
Spouse employment ("Unemployed" exlc.)		
Missing	0.0138	(0.73)
Employed full time	0.0038	(0.21)
Employed part-time	-0.0285	(1.12)
Dependents age 0-17	0.0059	(1.20)
Demographics		
Earned another degree/license prior to BA	0.0118	(1.17)
Ethnicity ("Hispanic" excl.)		
Missing	0.0090	(0.13)
Native American	0.0003	(0.01)
Asian	0.0761	(2.26)**
Black	0.0165	(0.56)
White	0.0059	(0.30)
SAT/ACT merged quartile ("Quartile 3" excl.)		
Missing	0.0149	(0.89)
Quartile 1	0.0210	(1.69)*
Quartile 2	0.0024	(0.22)
Quartile 4	0.0032	(0.27)
High school type ("Private, Catholic" excl.)		
Missing	-0.0371	(0.58)
Public	-0.0290	(1.29)
Private, non-Catholic	-0.0433	(1.57)
Private, non-religious	-0.0425	(1.50)
Age	-0.0074	(0.56)
Age sq.	0.0001	(0.78)
Job variables		
Tenure	-0.0013	(0.35)
Tenure sq.	0.0002	(0.68)
Undergrad decision making variables		
Crime rate important ("Yes" excl)		
No	-0.0021	(0.16)
Missing	-0.0301	(0.50)
Graduation rate important ("Yes" excl.)	-0.0228	(1.56)
Placement rate important ("Yes" excl.)	-0.0051	(0.50)
Constant	0.1427	(0.52)
Observations	3448	
Partial R^2 of excluded instrument	0.0052	

Absolute value of t statistics in parentheses
**significant at 10%; **significant at 5%; ***significant at 1%*

REFERENCES

Abbott, G. (1922). Federal Aid for the Protection of Maternity and Infancy. *American Journal of Public Health*, **12**(9), pp. 737–743.

Alley, W. (1984). The Palmer drought severity index: limitations and assumptions. *Journal of climate and applied meteorology*, **23**(7), pp. 1100–1109.

Antonovics, K., P. Arcidiacono, and R. Walsh (2004). Competing Against the Opposite Sex. *University of California at San Diego, Economics Working Paper Series*.

Antonovsky, A. and J. Bernstein (1977). Social class and infant mortality. *Social Science & Medicine (1967)*, **11**(8-9), pp. 453–470.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, **121**(1-2), pp. 343–375.

Arcidiacono, P. and J. Vigdor (2010). Does the River Spill Over? Estimating the Economic Returns to Attending a Racially Diverse College. *Economic Inquiry*, **48**.

Armstrong, D., C. Chapin, H. Emerson, A. Freeman, W. Frost, L. Thompson, C. Winslow, and L. Dublin (1924). Report of the Committee on Municipal Health Department Practice. *American Journal of Public Health*, **14**(3), p. 184.

Astin, A. (1977). *Four critical years*. Jossey-Bass.

Baker, J. (1918). Lessons from the Draft. *Transactions of the American Association for Study and Prevention of Infant Mortality*, pp. 161–178.

Bean, L. (1942). Crop yields and weather. *USDA Misc. Publ*, **471**.

Black, D. and J. Smith (2004). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics*, **121**(1-2), pp. 99–124.

Blank, R. (2002). Evaluating welfare reform in the United States. *Journal of Economic Literature*, **40**(4), pp. 1105–1166.

Bleakley, H. (2007). Disease and Development: Evidence from Hookworm Eradication in the American South. *The Quarterly Journal of Economics*, **122**(1), pp. 73–117.

Bleakley, H. and F. Lange (2009). Chronic disease burden and the interaction of education, fertility, and growth. *The Review of Economics and Statistics*, **91**(1), pp. 52–65.

Brodie, B. (1993). Children's Bureau: Guardian of American Children. *Nursing Research*, **42**(3), p. 190.

Brosco, J. (1999). The Early History of the Infant Mortality Rate in America:" A Reflection Upon the Past and a Prophecy of the Future" 1. *Pediatrics*, **103**(2), p. 478.

Clifford, W. and Y. Brannon (1978). Socioeconomic differentials in infant mortality: An analysis over time. *Review of Public Data Use*, **6**, pp. 29–37.

Company, L. (1922). Public Health Education (Chapter from Report of Committee on Municipal Health Department Practice). *American Journal of Public Health*, **12**(10), p. 815.

Cooley, T., S. Decanio, and M. Matthews (1977). An Agricultural Time Series-Cross Section Data Set. *NBER Working Paper*.

Costa, D. and M. Kahn (2004). Changes in the Value of Life, 1940–1980. *Journal of Risk and Uncertainty*, **29**(2), pp. 159–180.

Cutler, D. and G. Miller (2005). The role of public health improvements in health advances: the twentieth-century United States. *Demography*, **42**(1), pp. 1–22.

Daniel, K., D. Black, and J. Smith (2001). Racial Differences in the Effects of College Quality and Student Body Diversity on Wages. *Diversity challenged: evidence on the impact of affirmative action*, p. 221.

Deschenes, O. and M. Greenstone (2007). The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. *The American Economic Review*, **97**(1), pp. 354–385.

DOC (1925). *Financial Statistics of Cities, 1923*. United States Government Printing Office, Washington D.C.

DOC (1926a). *Biennial Census of Manufactures, 1923*. United States Government Printing Office, Washington D.C.

DOC (1926b). *Financial Statistics of Cities, 1924*. United States Government Printing Office, Washington D.C.

DOC (1927a). *Financial Statistics of Cities, 1925*. United States Government Printing Office, Washington D.C.

DOC (1927b). *Mortality Statistics, 1923*. United States Government Printing Office, Washington D.C.

DOC (1928a). *Biennial Census of Manufactures, 1925*. United States Government Printing Office, Washington D.C.

DOC (1928b). *Financial Statistics of Cities, 1926*. United States Government Printing Office, Washington D.C.

DOC (1928c). *Mortality Statistics, 1924*. United States Government Printing Office, Washington D.C.

DOC (1929a). *Financial Statistics of Cities, 1927*. United States Government Printing Office, Washington D.C.

DOC (1929b). *Mortality Statistics, 1925*. United States Government Printing Office, Washington D.C.

DOC (1930a). *Biennial Census of Manufactures, 1927*. United States Government Printing Office, Washington D.C.

DOC (1930b). *Financial Statistics of Cities, 1928*. United States Government Printing Office, Washington D.C.

DOC (1930c). *Mortality Statistics, 1926*. United States Government Printing Office, Washington D.C.

DOC (1931a). *Financial Statistics of Cities, 1928*. United States Government Printing Office, Washington D.C.

DOC (1931b). *Mortality Statistics, 1927*. United States Government Printing Office, Washington D.C.

DOC (1932a). *Biennial Census of Manufactures, 1929*. United States Government Printing Office, Washington D.C.

DOC (1932b). *Financial Statistics of Cities, 1930*. United States Government Printing Office, Washington D.C.

DOC (1932c). *Mortality Statistics, 1928*. United States Government Printing Office, Washington D.C.

DOC (1933a). *Financial Statistics of Cities, 1931*. United States Government Printing Office, Washington D.C.

DOC (1933b). *Mortality Statistics, 1929*. United States Government Printing Office, Washington D.C.

DOC (1934a). *Biennial Census of Manufactures, 1931*. United States Government Printing Office, Washington D.C.

DOC (1934b). *Financial Statistics of Cities, 1932*. United States Government Printing Office, Washington D.C.

DOC (1934c). *Mortality Statistics, 1930*. United States Government Printing Office, Washington D.C.

DOC (1935). *Mortality Statistics, 1931*. United States Government Printing Office, Washington D.C.

DOC (1936a). *Biennial Census of Manufactures, 1933*. United States Government Printing Office, Washington D.C.

DOC (1936b). *Mortality Statistics, 1932*. United States Government Printing Office, Washington D.C.

DOL (1914). *Reports of the Department of Labor, 1913*. United States Government Printing Office, Washington D.C.

DOL (1915). *Reports of the Department of Labor, 1914*. United States Government Printing Office, Washington D.C.

DOL (1916). *Reports of the Department of Labor, 1915*. United States Government Printing Office, Washington D.C.

DOL (1917). *Reports of the Department of Labor, 1916*. United States Government Printing Office, Washington D.C.

DOL (1921). *Reports of the Department of Labor, 1920*. United States Government Printing Office, Washington D.C.

Ferrell, J. and P. Mead (1936). History of County Health Organizations in the United States, 1908-33–Public Health Bulletin 222. Technical Report 222, United States Public Health Service.

Ferrell, J., W. Smillie, P. Covington, and P. Mead (1932). *Health departments of States and Provinces of the United States and Canada*. Washington, DC; US Government Printing Office.

Fishback, P., M. Haines, and S. Kantor (2007). Births, deaths, and New Deal relief during the Great Depression. *The Review of Economics and Statistics*, **89**(1), pp. 1–14.

Fishback, P. and V. Kachanovskaya (2009). In Search of the Multiplier for Net Federal Spending in the States During the New Deal: A Preliminary Report. Technical report, Working paper to be presented at the NBER-DAE Summer Institute.

Fishback, P., W. Troesekn, K. T., M. Haines, P. Rhode, and M. Thomasson (????). The Trials of Job: The Impact of Climate and Weather on Infant and Non-Infant Death Rates During the Great Depression. *Unpublished manuscript*.

Galloway, P. (1985). Annual variations in deaths by age, deaths by cause, prices, and weather in London 1670 to 1830. *Population Studies*, **39**(3), pp. 487–505.

Goldin, C., L. Katz, and I. Kuziemko (2006). The homecoming of American college women: The reversal of the college gender gap. *The Journal of Economic Perspectives*, **20**(4), pp. 133–4A.

Greene, D. (1922). Pamphleteering and Public Health. *American Journal of Public Health*, **12**(2), p. 148.

Haines, M., L. Craig, and T. Weiss (2003). The short and the dead: Nutrition, mortality, and the antebellum puzzle in the United States. *The Journal of Economic History*, **63**(02), pp. 382–413.

Hamilton, D. (1982). Herbert Hoover and the great drought of 1930. *The Journal of American History*, **68**(4), pp. 850–875.

Hannan, A. (1931). The Influence of Weather on Crops, 1900-1930: A Selected and Annotated Bibliography. *USDA Misc. Publ*, **118**.

Heckman, J., H. Ichimura, and J. Smith (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies*, **64**, pp. 605–54.

Heckman, J., H. Ichimura, and P. Todd (1998). Matching as an econometric evaluation estimator. *Review of Economic studies*, **65**(2), pp. 261–294.

Hirano, K. and G. Imbens (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes Research Methodology*, **2**(3), pp. 259–278.

Hirano, K., G. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, **71**(4), pp. 1161–1189.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. *NBER working paper*.

IRS (1933). Individual Income Tax Returns 1923-1932. Technical report, United States Treasury Department.

Karl, T., C. Williams Jr, P. Young, and W. Wendland (1986). A model to estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperatures for the United States. *Journal of Climate and Applied Meteorology*, **25**(2), pp. 145–160.

Komlos, J. (2009). The height and weight of West Point cadets: dietary change in antebellum America. *The Journal of Economic History*, **47**(04), pp. 897–927.

Lange, F., A. Olmstead, and W. Paul (2009). The Impact of the Boll Weevil, 1892-1932. *Journal of Economic History*, **69**(3), pp. 685–718.

Lathrop, J. (1919). Income and infant mortality. *American Journal of Public Health*, **9**(4), p. 270.

Lavy, V. and A. Schlosser (2006). Does Being with More Girls in School Improve Students Human Capital Outcomes and Behavior? Evidence on Effects and Mechanisms. *IZA Working Paper*.

Levy, E. (1920). Reduction of Deaths from Infantile Diarrhea by Care of the Bowel Discharges of Infants. *American Journal of Public Health*, **10**(5), pp. 400–404.

Lindenmeyer, K. (1995). The US Children's Bureau and Infant Mortality in the Progressive Era. *Journal of Education*, **177**(3), pp. 57–69.

Maddala, G. and S. Wu (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and statistics*, **61**(S1), pp. 631–652.

Martin, R. (1976). *National income in the United States, 1799-1938*. Ayer Co Pub.

Meckel, R. (1990). *Save the babies*. Johns Hopkins University Press.

Mendelsohn, R., W. Nordhaus, and D. Shaw (1994). The impact of global warming on agriculture: a Ricardian analysis. *The American Economic Review*, pp. 753–771.

Miller, G. (2008). Women's Suffrage, Political Responsiveness, and Child Survival in American History. *Quarterly Journal of Economics*, **123**(3), pp. 1287–1327.

Miller-Bernal, L. (1989). College Experiences and Sex-Role Attitudes: Does a Women's College Make a Difference? *Youth & Society*, **20**(4), p. 363.

Moehling, C. (2006). Mothers' Pension Legislation and the Cross-State Variation in Welfare Generosity. *Working paper*.

Mundlak, Y. and D. Larson (1992). On the transmission of world agricultural prices. *The World Bank Economic Review*, **6**(3), p. 399.

NCDC (2002). Data Set 9640: Time Bias Corrected Divisional Temperature–Precipitation–Drought Index.

Newmayer, S. (1911). The warfare against infant mortality. *The Annals of the American Academy of Political and Social Science*, **37**(2), p. 288.

Officer, L. (2008). The Annual Consumer Price Index for the United States, 1774-2007.

Palmer, G., P. Platt, W. Walker, A. Nicholl, and J. A. (1925). *A health survey of 86 cities*. American child health association.

Palmer, W. (1965). Meteorological drought. *Research paper*, **45**, p. 58.

Rosenbaum, P. and D. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**(1), p. 41.

Rosenbaum, P. and D. Rubin (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician*, **39**(1), pp. 33–38.

Routzahn, E. (1922). Symposium on How to Further Progress in Health Education and Publicity. *American Journal of Public Health*, **12**(4), pp. 279–289.

Rubin, D. (1977). Assignment to Treatment Group on the Basis of a Covariate. *Journal of Educational and Behavioral statistics*, **2**(1), p. 1.

Rude, A. (1920). Status of State Bureaus of Child Hygiene. *American Journal of Public Health*, **10**(10), p. 772.

Ruhm, C. (2000). Are Recessions Good for Your Health? *Quarterly Journal of Economics*, **115**(2), pp. 617–650.

Schlenker, W., W. Hanemann, and A. Fisher (2005). Will U.S. agriculture really benefit from global warming? Accounting for irrigation in the hedonic approach. *American Economic Review*, **95**(1), pp. 395–406.

Schlenker, W., W. Hanemann, and A. Fisher (2006). The impact of global warming on US agriculture: an econometric analysis of optimal growing conditions. *Review of Economics and Statistics*, **88**(1), pp. 113–125.

Skocpol, T., M. Abend-Wein, C. Howard, and S. Lehmann (1993). Women's associations and the enactment of mothers' pensions in the United States. *American Political Science Review*, **87**(3), pp. 686–701.

Smith, D. (1990). Women's Colleges and Coed Colleges: Is there a Difference for Women? *The Journal of Higher Education*, **61**(2), pp. 181–197.

Smith, D., L. Wolf, and D. Morrison (1995). Paths to success: Factors related to the impact of women's college. *The Journal of Higher Education*, pp. 245–266.

Snyder, T. (2005). *Digest of education statistics 2004*. Education Dept. Institute of Education Sciences.

Steckel, R. (1992). Stature and living standards in the United States. *American economic growth and standards of living before the Civil War*, pp. 265–308.

Stoian, A. and P. Fishback (2010). Welfare spending and mortality rates for the elderly before the Social Security era. *Explorations in Economic History*, **47**(1), pp. 1–27.

Thompson, S. (1921). Factors that Influence Infant Mortality. *American Journal of Public Health*, **11**, pp. 415–419.

Thorp, W., W. Mitchell, and H. Thorp (1926). *Business annals*. National Bureau of Economic Research, New York.

Tobey, J. (1926). The Children's Bureau: Its History, Activities and Organization. *The American Journal of Nursing*, **26**(1), p. 85.

Tobias, J. (2003). Are Returns to Schooling Concentrated Among the Most Able? A Semiparametric Analysis of the Ability–Earnings Relationships. *Oxford Bulletin of Economics and Statistics*, **65**(1), pp. 1–29.

Troesken, W. (2004). *Water, race, and disease*. The MIT Press.

Trotter, P., G. Johnson, R. Ricks, and S. D. (2009). Floods on the Lower Mississippi: An Historical Economic Overview.

USDA (????). Fluctuations in Crops and Weather: 1866-1948. *USDA Statistical Bulletin*, (101).

USDA (1923). *Annual Yearbook 1922*. USDA.

Vincent, G. (1918). The Rockefeller Foundation Annual Report.

Vincent, G. (1921). The Rockefeller Foundation Annual Report.

Waldmann, R. (1992). Income distribution and infant mortality. *The Quarterly Journal of Economics*, **107**(4), pp. 1283–1302.

Whitt, E. (1994). I can be anything! Student leadership in three womens colleges. *Journal of College Student Development*, **35**(3), pp. 198–207.

Woodbury, R. (1924). The Trend of Maternal-Mortality Rates in the United States Death-Registration Area, 1900-1921. *American Journal of Public Health*, **14**(9), p. 738.

Ziegler, C. (1922). How Can We Best Solve the Midwifery Problem? *American Journal of Public Health*, **12**(5), p. 405.