# SCHOOL QUALITY AND WAGES

A Dissertation

by

ROBERT B. SPEAKMAN, JR.

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2006

Major Subject: Economics

# SCHOOL QUALITY AND WAGES

A Dissertation

by

ROBERT B. SPEAKMAN, JR.

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

| | |
|---|---|
| Chair of Committee, | Finis Welch |
| Committee Members, | Donald Deere |
| | Manuelita Ureta |
| | Mark Fossett |
| Head of Department, | Amy Glass |

December 2006

Major Subject: Economics

## ABSTRACT

School Quality and Wages. (December 2006)

Robert B. Speakman, Jr., B.A., Brigham Young University

Chair of Advisory Committee: Dr. Finis Welch

This dissertation examines the literature that attempts to measure the relationship between school quality and earnings. I begin by developing a simple economic model that predicts that, everything else being equal and with comparisons being made within a market, workers from higher quality schools will have higher earnings among those with the same level of schooling and they will have steeper schooling-earnings gradients. The remainder of this dissertation explores problems that exist in this literature for which no solutions have been presented. These problems include: 1) there doesn't have to be a direct and positive relationship between school quality and earnings; 2) the data suggest that school quality measures are frequently mismatched to workers; 3) most school quality studies include college-trained labor while completely ignoring the quality of the college attended; 4) the omission of college quality from the estimation is especially problematic for studies that attempt to measure the school quality-earnings relationship through differences in schooling-earnings gradients for those educated in different systems; 5) state of birth wage rankings thought to capture a school quality effect are not invariant to the market (state of residence) in which they are evaluated; and 6) the evidence presented herein suggests that interstate migration is selective. These problems undermine the credibility of existing estimates of a school quality-earnings relationship.

This dissertation is dedicated to Tyler, Austen, Nathan, and J.R.
who teach me new things everyday.

## ACKNOWLEDGEMENTS

I would like to express my appreciation to all of my committee members. Each took the time to review and discuss my dissertation with me and each added helpful insights and suggestions that improved the final product. They have also been very accommodating with the scheduling of all of the requirements that must be completed prior to the awarding of a Ph.D. Among my committee members, I would like to single out the chairman, Finis Welch, who has encouraged me for some time to complete this thesis. I'm doing so after a mere sixteen years in the program. Finis has both dangled the carrot and cracked the whip. I'm not sure which worked, but I am grateful for his persistence. I have had the opportunity to work for Finis for the last eleven years and have learned from him everything I know about organizing and analyzing data. For this, I am deeply indebted.

I would also like to thank my wife Natie and my four sons, Tyler, Austen, Nathan, and J.R., who have been patient with me as I spent many nights in the office over the last several months attempting to complete this thesis. Each has been very supportive. In fact, I think that those who are old enough to understand are more excited about it than I am. I hope that my sons take away from this that a good thing is worth completing, even if it takes a lot longer than expected.

# TABLE OF CONTENTS

Page

ABSTRACT ........ iii

DEDICATION ........ iv

ACKNOWLEDGEMENTS ........ v

TABLE OF CONTENTS ........ vi

LIST OF FIGURES ........ viii

LIST OF TABLES ........ ix

CHAPTER

I INTRODUCTION ........ 1

Introduction ........ 1
A Brief Synopsis of the Literature ........ 3
Summary ........ 6

II AN ECONOMIC MODEL OF SCHOOL QUALITY AND WAGES ........ 9

Theory ........ 9
An Ad Hoc Index of School Quality ........ 14
Comparing the Theory to the Literature ........ 15

III THE MISIDENTIFICATION OF SCHOOLS ATTENDED ........ 19

Introduction ........ 19
Birth State and Attendance State ........ 20
School Measures from a School Actually Attended ........ 27
Conclusions ........ 28

CHAPTER Page

IV THE CONFOUNDING INFLUENCE OF COLLEGE-TRAINED WORKERS ........ 30

Introduction ........ 30
The Importance of Omitted College Quality: A Simple Model ........ 31
How Much Information Does $Q_H$ Carry About $Q_C$? ........ 33
The Influence of College Trained Labor on the Estimated Schooling-Earnings Gradients ........ 35
Alternative Estimates of School Quality Using the Card/Krueger Data ... 42
Quality Effects Measured Through Schooling-Earnings Gradients ... 43
Quality Effects Measured Through Wages at Each Schooling Level 48
Matching Measures of College Quality to Workers ........ 51
Conclusions ........ 53

V ARE THE DATA CONSISTENT WITH THE THEORETICAL PREDICTIONS? ........ 56

Introduction ........ 56
Wage Rankings by Birth States Across Markets ........ 57
But Have You Looked at Occupation? ........ 65
A Preliminary Investigation of Interstate Migration ........ 66
Why Do People Move from State to State? ........ 67
The Importance of Age and Years of Schooling ........ 71
Additional Observations Regarding Interstate Migration ........ 74
Conclusions ........ 83

VI CONCLUSIONS ........ 85

REFERENCES ........ 89

APPENDIX A ........ 94

APPENDIX B ........ 128

APPENDIX C ........ 132

VITA ........ 134

## LIST OF FIGURES

FIGURE Page

2.1 Illustrative wage-schooling profiles contrasting students from lower and higher quality schools. ........ 16

4.1 Revisiting Figure 2.1. ........ 42

4.2 An example of how excluding college-trained labor biases the pool of remaining workers. ........ 55

5.1 The wages of high school and college graduates for men from a given birth state. ........ 63

5.2 Percent of white, non-Hispanic men with an interstate move by age. ........ 71

5.3 Percent of white, non-Hispanic men with an interstate move in the last five years by education and age. ........ 74

5.4a Inflows, outflows, and the net migration of white, non-Hispanic prime-mover age men with a high school degree as a percent of the 1995 stock of those with a high school degree. ........ 76

5.4b Inflows, outflows and the net migration of white, non-Hispanic prime-mover age men with a college degree as a percent of the 1995 stock of college graduates. ........ 77

5.5 Do high school graduates and college graduates move to different markets? The percent of high school and college graduate interstate migrants moving from the same 1995 residence state to the same 2000 residence state. ........ 80

5.6 Movement between large cities for white, non-Hispanic, prime-mover age men with a college degree, 1995 to 2000. ........ 82

## LIST OF TABLES

TABLE Page


3.1 Elementary and Secondary Enrollment Rates in 1990 by School Type, Whether Living in Their Birth State, and Age .................................................. 21

3.2 Percent of Enrolled Children Not Attending Public School in Their Birth State by Age for the Years 1960 – 2000.............................................................. 22

3.3 Elementary and Secondary Enrollment Rates in 1990 by School Type, Whether Living in Their State of Birth, and Birth State .................................... 25

3.4 Elementary and Secondary Enrollment Rates in 1990 by School Type, Whether Living in Their State of Birth, and State of Residence......................... 26

3.5 Percent of Enrolled Children Moving Within the Last Five Years Categorized by the Type of Move and Age........................................................ 28

4.1 The Proportion of the Population Spanned by Two Observations, with $X_1>X_2$, to be Able to Conclude with a Given Confidence Level that $Y_1>Y_2$ ..... 34

4.2 The Probability that $Y_1>Y_2$ Given that $X_1>X_2$. ................................................ 35

4.3 National Average Weekly Wage Differences and Weights for Adjacent Levels of Schooling, 1980 Census ..................................................................... 37

4.4 Schooling-Earnings Gradients Calculated Using Workers of All Schooling Levels and the Amount of the Gradient Attributable to College-Trained Workers by Residence State ........................................................................... 39

4.5 The Amount of the Schooling-Earnings Gradient Attributable to College-Trained Workers in 1980 by Residence and Birth State Pairs ............................ 41

4.6 Alternative Estimates of the Contribution of State School System Characteristics to the Earnings-Schooling Gradient – Card and Krueger Revisited ........................................................................................................ 46

4.7 Alternative Estimates of Effects of School System Characteristics on Earnings ......................................................................................................... 49

4.8 Post-Secondary Enrollment by School Type and Whether Living in Their Birth State Reported by Age........................................................................... 52

TABLE Page

5.1 Log Wage Differentials For Men Living in a Given State from Different Birth States for High School and College Graduates Reported for Selected Residence and Birth States ........ 58

5.2 Percent of Male High School Graduates Ages 30 to 59 Within a Residence State Who Are Working in a Blue Collar Job Among Those from a Given Birth State ........ 66

5.3 Mobility Propensities Reported by Schooling Level ........ 68

5.4 The Main Reason Given For Moving by Interstate Migrants Reported by Schooling Level ........ 69

C.1 State Specific Schooling Gradients, the Return to a College Degree Over a High School Degree ........ 133

# CHAPTER I
# INTRODUCTION

## Introduction

During the 2000-2001 school year, $738 billion, or 7.5 percent of GDP, was spent on education in the United States with three-fifths being spent on elementary and secondary training ($442 billion) and the remainder producing college and graduate training. These levels represent a fifty percent increase in spending on education as a percent of GDP over just forty years earlier.[1] In international comparisons, the United States ranks high in educational spending. In 2000, only four nations spent more per pupil than the United States on elementary and secondary schooling and no other nation outspent the U.S. on college and graduate training per pupil.[2]

In 1992, Card and Krueger (1992a) published their study of the effect of school quality on earnings titled “Does School Quality Matter? …” Shortly thereafter, the proceedings of a Brookings Institute conference on the effect of school resources on student outcomes was published under a similar title, “Does Money Matter? …” In practice, the two titles are nearly synonymous as most studies that examine “school quality” focus on either expenditures per pupil or on teacher’s salaries, the largest component of expenditures. Given the expenditure figures above, the answer to the questions posed in these titles must be, “We hope so.” It is clear that we believe that spending matters and the presumption must be that extra spending buys something of value, i.e., higher quality schools.[3]

---

This dissertation follows the style and format of the *Journal of Labor Economics*.

[1] Digest of Education Statistics 2003, Table 29. The value for 1960, 4.97%, is linearly interpolated given the values of spending and GDP for 1959 and 1961.

[2] Condition of Education 2004, Table 36-1. The four countries with higher spending on elementary and secondary schooling per student are Switzerland, Austria, Denmark, and Norway.

[3] This “something of value” could be many things such as higher productivity (and earnings), better citizens, more ethical people, healthier people, or an increased ability to enjoy leisure time to name just a

The school quality literature, however, is much less conclusive on the matter. There are two main branches in this literature. The first examines the effects of quality on test scores. In general, these studies find that schools matter and that teachers matter, but are unable to correlate these findings with the measures of quality – expenditures, pupil-teacher ratios, teacher's salaries and educational attainment, and term lengths – that are available in practice. These studies have the advantage of timeliness; the impact of differences in or changes to any definable quality measure can be studied almost immediately. With this advantage also comes a major drawback; it's unclear what these tests measure and whether a higher test score is correlated with anything that we care about.

For example, in Speakman and Welch (2006), we used the NLSY79 to examine the ability of years of schooling and the AFQT (the Armed Forces Qualifying Test) to predict wages. We followed the sample selection criteria and variable definitions used by Johnson and Neal (1996). One the one hand, consider all that is omitted when only the number of years of schooling completed is included in the regression. Among those of the same age and years of schooling, it ignores, for example, differences in school and home resources, effort, innate ability, the effectiveness of teachers, classes available and classes taken, peer group, etc. AFQT score, on the other hand, is probably the most commonly used "ability" measure in the labor economics literature. That said, we found that the years of schooling that an individual attended is at least as good a predictor of his or her salary as is test measured achievement. Neither for men nor for women does the AFQT "explain" a higher fraction of the residual wage variation, after age and race are regressed out, than does the level of school completion. Moreover, we found the impact of AFQT on earnings to be small; a one percentile change in the AFQT score is predicted to increase wages by at most one-third of one percent.[4]

---

few. This dissertation, however, focuses on the literature that attempts to measure the effect of higher quality schooling through earnings.

[4] The predicted increase is 0.26 percent for men and 0.33 percent for women.

The second branch of the literature attempts to link these same schooling measures to earnings. Most of these studies find a positive and statistically significant relationship between earnings and the schooling measures thought to identify higher quality schools.[5] The advantage to these studies is that we know what is measured, the skills that are valued in the labor market. One drawback, of course, is that years or even decades must pass before the impact of quality measures can be observed. Other limitations in these studies will be discussed throughout this thesis.

**A Brief Synopsis of the Literature**

Appendix A contains a non-critical, encyclopedic summary of the school quality-earnings literature. Attention is restricted to published studies that examine earnings for those educated, often inferred by place of birth, in the United States. After each study is identified, the summary describes the earnings and schooling data, the population studied, and the school variables used to attempt to infer school quality. Next, the empirical model is outlined along with a description of variables other than the school ones. Finally, the author's conclusions are reported.

The studies are grouped by the earnings data source and listed chronologically. The data sources include studies that examine wages and individual characteristics taken from a variety of sources: the U.S. Decennial Censuses (1960-1990), various releases of the original cohort files of the National Longitudinal Surveys of Young Men and Young Women (NLS-YM and NLS-YW), the NBER-Thorndike data, the Panel Survey of Income Dynamics (PSID), the National Longitudinal Survey of Youth 1979 (NLSY79), High School and Beyond (HSB), the National Longitudinal Study of the High School Class of 1972 (NLS72), Project Talent, the Postcensal Survey of Professional and Technical Manpower (conducted in 1962), and a Minnesota Twin Registry survey.

---

[5] The exception is the studies that have examined the NLYS, NLS72, and HSB, which find that schools matter, i.e., the average earnings for those educated in different schools are not equal, but that the measured characteristics of schools available in the data are not statistically significant determinants of income.

Several of these sources – NLS-YM, NLS-YW, NLSY79, HSB, and NLS72 – provide school characteristics from the high schools or districts attended by the respondents during a given year. Available information that has been utilized in these studies includes district level expenditures per pupil in average daily attendance (both adjusted and unadjusted for cost-of-living differences), teachers per pupil, counselors per pupil, the starting pay of teachers with a bachelor's degree, school enrollment, library books per student, the percent of teachers with a master's degree, and the curriculum available.

The remainder of the studies combine the other earnings data with school characteristic data from alternative sources. Most have relied upon information reported in the *Biennial Survey of Education*[6] and the *Digest of Education Statistics*[7] matched to earnings data by state-of-birth (most Census studies) or state-of-residence at age 12 (all PSID studies) or location where attended high school (some NBER-Thorndike studies[8]). The *Biennial Survey* reports statewide averages of expenditures, teachers' salaries, instructional staff, pupils enrolled and pupils in average daily attendance, term lengths, and other variables in even-numbered school years from 1918 to 1958 for public elementary and secondary schools. Data is reported separately by race for white and black schools in 18 states[9] with segregated school systems from 1918 until the *Brown v.*

---

[6] U.S. Department of Health, Education, and Welfare. Office of Education. *The Biennial Survey of Education in the United States*. Washington, D.C. 1918-1958. In the earlier years, the *Survey* was published by the U.S. Department of the Interior (1918-1936) and the Federal Security Agency (1938-1948) and, in the first several years it was published by the Bureau of Education (part of the Department of the Interior).

[7] U.S. Department of Health, Education, and Welfare, Education Division. National Center for Education Statistics. *The Digest of Education Statistics.* Washington, D.C. 1962-2002. Publications through 1974 were titled *The Digest of Educational Statistics* and were published by the following organizations within the U.S. Office of Education: Division of Educational Statistics (1962-64), Bureau of Educational Research and Development (1965), the National Center for Educational Statistics (1966-73), and the National Center of Education Statistics (1974).

[8] This includes two studies by Wachtel (1975, 1976) and one by Link and Ratledge (1976). These authors cite the *Biennial Survey* and appear to use city level data published for large city school systems matched to the high school attended as identified in the NBER-Thorndike survey.

[9] They are Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia.

*Board of Education* decision in 1954. The *Digest of Education Statistics* continued where the *Biennial Surveys* left off and has been published annually since 1962.[10]

Although only six of the thirty-six (16.7 percent) studies in the literature examine census data, these are the more heavily cited ones (58.4 percent of citations) in recent academic publications (1990 forward). And, two out of every three citations (66.3 percent) to this literature are to studies that use the *Biennial Surveys* to measure school characteristics. One study in particular, Card and Krueger's "Does School Quality Matter? …" (1992a), which matches 1980 census data to the *Biennial Survey* data, accounts for 26.4 percent of these citations. As such, some of the calculations in the proceeding chapters focus on the census data or on Card and Krueger's data in particular. For the most part, however, the limitations of these published studies discussed in this dissertation apply to all them and not just the census studies nor to those that use the *Biennial Survey* data.

While almost all of the studies summarized focus exclusively on elementary and high school characteristics, there are a few that examine characteristics of the college attended such as expenditures per full-time equivalent student, faculty salaries, and faculty per full-time equivalent student. This information has been gleaned from publications by the U.S. Office of Education, the Integrated Postsecondary Education Data System (IPEDS), Barron's Guides and CASPAR[11] data. In addition, some have attempted to use rankings of colleges such as those produced by Gorman (1967), Carter (1966), Astin (1965), and Cass and Birnbaum (1964) as a proxy for college quality.[12]

Although it is clear that data is available for many inputs that might enter into the production of education, most studies examine primarily expenditures per pupil or teachers' salaries, a major component of total expenditures. This is peculiar given that the production function for education is essentially a mystery and the ways schools

---

[10] Except for combined editions that were issued in 1977-78, 1983-84, and 1985-86.
[11] The CASPAR database provides data on finances, enrollments and degrees beginning in 1967 for approximately 3,899 colleges and universities from the Higher Education General Information Survey (HEGIS).
[12] J. Gourman, *The Gourman Report*. Carter, *An Assessment of Quality in Graduate Education*. Astin, *Who Goes Where to College*. Cass, James and Birnbaum, Max, *Comparative Guide to American Colleges*.

spend money is likely to be as important as the amount spent. Equally important, if the purpose of the analysis is to inform public policy makers, a recommendation either to spend more or to spend less does not appear to be especially informative.

In reviewing the summary, note that if there is a consensus it is that wages are positively correlated with either school's expenditures, which are dominated by teacher costs, or teachers' salaries *per se*. This relation, however, may be spurious in the sense that in the cross-section, wages for one group may be positively correlated with the wages of others for reasons that supercede the causality of teacher to student.[13]

**Summary**

I begin in Chapter II by developing a simple economic framework in which the effect of school quality on earnings can be examined. The main assumptions are that higher quality schools produce education at a faster rate, a worker's earnings are the product of the amount of education possessed and the price of education, and the amount of education possessed is an increasing function of the quality of schooling and the years of schooling. The main theoretical implication of the model is that within well-defined labor markets, where the price of education is the same for all workers, higher quality schools produce workers that earn more. Workers from higher quality schools will have higher earnings among workers with the same level of schooling and they will have steeper schooling-earnings gradients, everything else being equal. Finally, I place the empirical models from the literature in the context of the theoretical model developed in this chapter.

---

[13] There are large interstate differences in wages and costs-of-living. The current differentials appear to have persisted throughout the range of the available data; wages are low in Mississippi, South Carolina and Arkansas and they are high in Pennsylvania, New Jersey and New York. Most wage earning adults reside in the state of their birth, so if—across states-we compare wages today with teacher salaries or educational expenditures in earlier periods, we expect positive covariances. A few papers, notably, Card and Krueger (1992a) and Heckman, et al. (1996), adjust teachers' salaries for area wide differences in wages, but the large majority of the papers ignore this confounding influence.

The next two chapters discuss problems generated by measurement error in the quality variables. In the literature, statewide average school quality measures from sources such as the *Biennial Survey* are either matched to birth state or school quality measures from a high school attended during a single academic year are collected with the data. Chapter III examines the misidentification of the school system attended using both of these methods and examines the impact of the measurement error. In the census type studies, about 20 percent of students never attend any of the schools included in the statewide averages to which they are matched and, at a minimum, 30 percent spend part of their academic career in a different state. Using school quality measures from a single school actually attended also leads to high error rates. A 30 percent error rate in matching school characteristics to students will reduce the school quality-earnings signal by 70 percent, greatly undermining the ability of these studies to observe an effect.

Moreover, even though almost all studies summarized in Appendix A include workers who have attended college, very few of them consider the impact of college quality on earnings. Unless college and elementary/high school quality are highly correlated, this will lead to measurement error in the school quality measures that capture the relevant academic career. Using figures presented by Wachtel (1976), it is estimated that the correlation between these two quality measures is 0.09, i.e., it is small. Moreover, schooling gradients that are estimated using workers of all schooling levels are highly dependent, 70 percent on average in the U.S., on those with college training. Studies that examine the impact of quality through gradients while ignoring the quality of the college attended risk finding a spurious relationship between school quality and earnings because an important component of quality is not considered.

An obvious solution to this problem is to exclude those with college training. However, doing so typically weakens, and sometimes reverses, the positive associations between school quality measures and earnings observed in the literature. Re-estimating Card and Krueger's gradient models without college trained labor changes the sign on the estimated impact of the pupil/teacher ratio (now positive) and the term length (now negative). Extending their analysis to an examination of the impact of these school quality measures among those with the same level of schooling reveals that their findings regarding pupil/teacher ratios and term lengths stem from perverse foundations; the impact of these factors is opposite the expected signs with the impact being larger for those with lower levels of education.[14] Their one robust finding is that teachers' wages positively affect earnings. As mentioned above, this finding may be the result of economic phenomena other than a school quality effect. Finally, omitting those with college training may also confound the comparisons if higher quality schools encourage increases in the level of schooling for students with higher quantities of some other factor (such as ability) that is positively correlated with earnings so that the comparisons made at a given level of schooling are among workers who are not the same on average in all other factors.

Chapter V makes the main comparisons implied by the theory in Chapter II, that workers from higher quality schools identified by birth states will earn more in all markets (or residence states). Unfortunately, these birth state wage rankings are not invariant across markets suggesting that migration is selective, i.e., workers migrating to different places do not have the same average quantities of other factors that determine earnings and that are omitted from the regression. Only one study, Heckman, et al. (1996), has suggested a correction for selective migration; however, it is not possible to implement their strategy in these comparisons. The remainder of Chapter V presents a preliminary exploration of interstate migration that suggests that migration is a complex topic that must be resolved before reliable school quality estimates can be obtained.

The final chapter presents concluding thoughts on the identification of a school quality effect.

[14] When measuring the effects of school quality on earnings through the schooling-earnings gradient, the impact is equal to the quality effect at the higher level relative to the quality effect at the lower level. In this case, a negative effect relative to a larger negative effect results in a positive effect, but one with perverse underpinnings.

# CHAPTER II
# AN ECONOMIC MODEL OF SCHOOL QUALITY AND WAGES*

## Theory

Define education as an unobserved bundle of skills that workers sell in the labor market that are valuable to firms in producing goods and services. Education is *not* the number of years that a student sat in a classroom, but years of schooling (or grades completed) is a component of education. So too are the quality of the school, the innate ability of students, the effort of students, family and community support, and many other factors.

The quality of a school is measured by its ability to produce units of education. In this section, I abstract from other factors that comprise education and focus on two components, years of schooling and quality of schooling. The amount of education that a worker possesses, $e(Q_s,s)$, is a function of the amount of time spent in school, $s$, and the quality of the school for those completing $s$ levels, $Q_s$. The quality of schooling is allowed to vary for those completing different levels of schooling. The first derivative with respect to each argument is non-negative, $e_s \geq 0$ and $e_{Qs} \geq 0$. Importantly, higher quality schools produce more education given the amount of time spent in school.

Suppose that aggregate output is a function of the stock of school-trained labor, which I refer to as aggregate education ($E$), and a numeraire input ($X$).

$$Y = f(E,X) \tag{2.1}$$

where education is the sum, over schooling levels, $s$, of the education of those who completed $s$ grades and no more; i.e.,

---

* Part of this Chapter is reprinted from Speakman, Robert and Welch, Finis, "Chapter 13. Using Wages to Infer School Quality" in *Handbook of the Economics of Education, Vol. 2* edited by Eric Hanushek and Finis Welch, Copyright 2007 with permission from Elsevier.

$$E = \sum_s N_s e(Q_s, s) \tag{2.2}$$

with $N_s$ being the number who completed $s$ grades with school quality $Q_s$ and let $e(Q_s,s)$ now measure their average school acquired job skills. This average $e(Q_s,s)$ depends only on school quality at level $s$ and the level of schooling.

There is a technological issue regarding the nature of the relation $e(Q_s,s)$. In much of the literature there is an implicit assumption that it is multiplicative, i.e.,

$$e(Q_s, s) = Q_s h(s) . \tag{2.3}$$

$h(s)$ is the effective units of "schooling" for one completing $s$ grades. This allows the education function to be non-linear in $s$, i.e., sixteen years of schooling is not necessarily twice as many as eight years of schooling. In this form, if the quality of schools in one system is twice that of another, then for students from the two systems matched on schooling, those from the superior school system will have twice as much education as students from the inferior system. In Mincer's (1974) specification, $h(s)$ is the exponential function, i.e., $e(Q_s,s)=Qe^s$.

An alternative specification used by Card and Krueger (1992a) is exponential in both $s$ and $Q$,

$$e(Q_s, s) = e^{s(a_0 + a_1 Q_s)} \tag{2.4}$$

so that for fixed quality differentials between school systems, the logarithmic educational differential increases with the level of schooling.

In what follows I use the simpler specification (2.3).[15] In this form, the marginal product of a representative individual with schooling $s$ is

---

[15] A similar solution, although not as clean, can be derived using specification (2.4).

$$\frac{\partial Y}{\partial N_s} = Q_s h(s) f_E \tag{2.5}$$

where $f_E$ is the derivative of the aggregate production function with respect to aggregate education. Wages, expressed relative to the numeraire, are

$$w(s) = Q_s h(s)(f_E / f_X)\,. \tag{2.6}$$

The ratio, $f_E/f_X$, is the unit price of education so that the wage is the quantity of education times its unit price.

Now assume that school quality is constant, $Q_s$=$Q$ for all values of $s$, and consider the effect of an increase in quality on wages.

$$\frac{\partial w(s)}{\partial Q} = h(s)(f_E / f_X) + Qh(s)\frac{\partial (f_E / f_X)}{\partial E}\frac{\partial E}{\partial Q} \tag{2.7}$$

The second term on the right hand side of equation (2.7) is the effect of increasing school quality on the stock of education and, therefore, on its price. Since, holding the distribution of school completion constant, an increase in school quality will increase the stock of education, the effect must be to lower its price. If the labor market were truly national, we can ignore the second term on the RHS of equation 5 since an increase in the quality of schools in the areas typically studied in the literature – a school, a district, or even a state – would not have an appreciable effect on the aggregate stock of education in the U.S. The simple evidence suggests, however, that the market is not national for all levels of workers. The returns to schooling, for example, are not equal across census divisions or states.[16] As such, I proceed to investigate the second term. First,

$$\frac{\partial(f_E / f_X)}{\partial E} = (f_X f_{EE} - f_E f_{EX}) f_X^{-2} \tag{2.8}$$

which, if *f(E,X)* is linear homogeneous, becomes

$$\frac{\partial(f_E / f_X)}{\partial E} = -\frac{f_E}{f_X}\frac{1}{E}\cdot\frac{1}{\sigma} \tag{2.9}$$

where $\sigma$ is the elasticity of substitution between $E$ and $X$ in $f$, i.e., $\sigma = f_E f_X / y f_{EX}$ by definition, and $y = Ef_E + Xf_X$ and $Ef_{EE} + Xf_{EX} = 0$ by Euler's Theorem.

Because $\frac{\partial(f_E / f_X)}{\partial Q} = \frac{\partial(f_E / f_X)}{\partial E}\frac{\partial E}{\partial Q}$ it follows that

$$\frac{\partial(f_E / f_X)}{\partial Q} = \frac{f_E}{f_X}\left(-\frac{1}{\sigma}\right)\frac{1}{E}\frac{\partial E}{\partial Q}. \tag{2.10}$$

With $Q$ constant over all levels of schooling, $E = Q\sum_s N(s)h(s)$, and it is obvious that holding the amount of schooling constant $\frac{Q}{E}\frac{\partial E}{\partial Q} \equiv 1$.

---

[16] For example, data for white, non-Hispanic males ages 20 to 64 from the 2000 Census reveal that the states that pay the least to college graduates (relative to high school graduates) are Wyoming (23.7%), Vermont (29.6%), South Dakota (29.7%), Montana (31.2%), and North Dakota (32.0%). At the other end of the spectrum are Texas (56.0%), New York (57.2%), California (57.3%), Georgia (58.2%), and Virginia (59.9%). At the extremes, the top five states pay college graduates nearly double what the bottom five states pay when compared to the earnings of high school graduates in the state. If labor is freely mobile, we expect to see migration occur until incomes, and hence college premia, equalize across states. While the interstate migration rate of college graduates is fairly high, perhaps sufficiently high to equalize wages, the interstate migration rate of high school graduates is much lower. (See Figure 2.3 in Chapter V.) To equalize wages for all levels of schooling, there must be sufficient migration at all schooling levels. Additional sample restrictions and the model specification are described in Appendix C.

Thus from equations (2.7) and (2.10),

$$\frac{Q\partial w(s)}{\partial Q} = w(s)\left(1 - \frac{1}{\sigma}\right) \tag{2.11}$$

and

$$\frac{\partial \ln w(s)}{\partial \ln Q} = \left(\frac{\sigma - 1}{\sigma}\right). \tag{2.12}$$

With a two-factor aggregate production function, the elasticity of substitution between *E* and *X* is (the negative of) the factor demand elasticity and we are reminded that if demand is of unit elasticity, $\sigma$=1, there is no relation between school quality and wages. If demand is inelastic, $\sigma$<1, holding other things constant, an increase in the quality of schooling will *lower* wages. This underscores the need for empirical investigations of the wage-school quality relation to consider the potential endogeneity of wages, i.e., wages must be examined within markets where the price of education is the same for all workers.

To illustrate, suppose that there are two markets, A and B, and that workers are born, educated, and work in these markets without any migration between markets. Let the schooling distribution be the same in both markets, but let the quality of schools in A be twice that of the schools in B. Further assume that the demand (function) for education by firms is the same in both markets. Under these assumptions, there will be twice as much aggregate education in A as in B, and the price of education will be lower.[17] Recall that wages are the amount of education times the price of education. For a given schooling level, who earns more, workers in A or workers in B? Workers in A have twice as much education as those in B, but the price of education is lower.

---

[17] That is, unless the demand for education is perfectly elastic.

Workers in A may earn less (inelastic demand, so that $P_E^A < .5P_E^B$), the same (unit elasticity, $P_E^A = .5P_E^B$), or more (elastic demand, $P_E^A > .5P_E^B$) than workers in B depending upon the price response to differences in the amount of education.

The purpose of this illustration is not to capture all of the complexities of the real world, incorporating elements such as migration and the endogeneity of school quality, but simply to show that if comparisons are not made within market, there does not have to be a positive relationship between wages and school quality. Aside from school quality, differences in the amount of education in the market due to differences in the amount of schooling (supply) will impact the price of education as will differences in the demand for education, perhaps driven by differences in available technologies or resources.

**An Ad Hoc Index of School Quality**

Returning to equation (2.6) with $e(Q_s,s)$ substituted for $Q_s h(s)$ yields

$$w(s) = e(Q_s, s)(f_E / f_X) \tag{2.13}$$

Suppose that two individuals competing in the same market with the same level of schooling, *s*, were educated in different school systems, A and B, with different quality levels, $Q_s^A \neq Q_s^B$. The relative wages of those educated in the two systems forms an index of the relative quality of the two school systems since the factor prices cancel.

$$\frac{w_s^A}{w_s^B} = \frac{e(Q_s^A, s)}{e(Q_s^B, s)}. \tag{2.14}$$

Given the specification in equation (2.3), this becomes,

$$\frac{w_s^A}{w_s^B} = \frac{Q_s^A}{Q_s^B} . \tag{2.15}$$

Under the Card-Krueger specification, equation (2.4), it is

$$\frac{w_s^A}{w_s^B} = e^{a_1 s (Q_s^A - Q_s^B)} . \tag{2.16}$$

If $Q_s^A > Q_s^B$, then the ratios in (2.15) and (2.16) will both be greater than one; if $Q_s^A < Q_s^B$, then the ratios will be less than one; and, if $Q_s^A = Q_s^B$, then the ratios will be equal to one. The relative wages of workers educated in different school systems forms an ad hoc index of school quality and this can be examined empirically given the wage comparisons in (2.15) and (2.16). Moreover, this comparison is completely detached from any debate regarding which measurable attributes of schools capture the quality of a school.

In addition, when workers from Schools A and B compete within more than one market, it is possible to form an ad hoc index of relative school quality within each market. If School B is better than School A, then workers who attended School B should earn more than those who attended School A in *all* markets, *ceteris paribus*. These relationships are explored empirically in Chapter V.

**Comparing the Theory to the Literature**

The theory is illustrated graphically in Figure 2.1. The comparison is between the products of two schools where quality is higher in one than in the other. In each case the wage increases with the level of schooling, but the increase is faster for students of the higher quality school where, by assumption, a year of schooling resulted in greater learning (more education). Wages are higher for students from the higher quality school at all positive levels of schooling. As the wage profiles are projected to zero schooling, they converge because at zero schooling there should be no school quality differential. In addition, the schooling gradient is steeper. In empirical work, higher quality schools can be identified as those whose students have a higher return to schooling or higher wages at each level of schooling. Comparisons of either gradients or wages at given schooling levels are both valid and should agree in identifying the relative quality of schools.

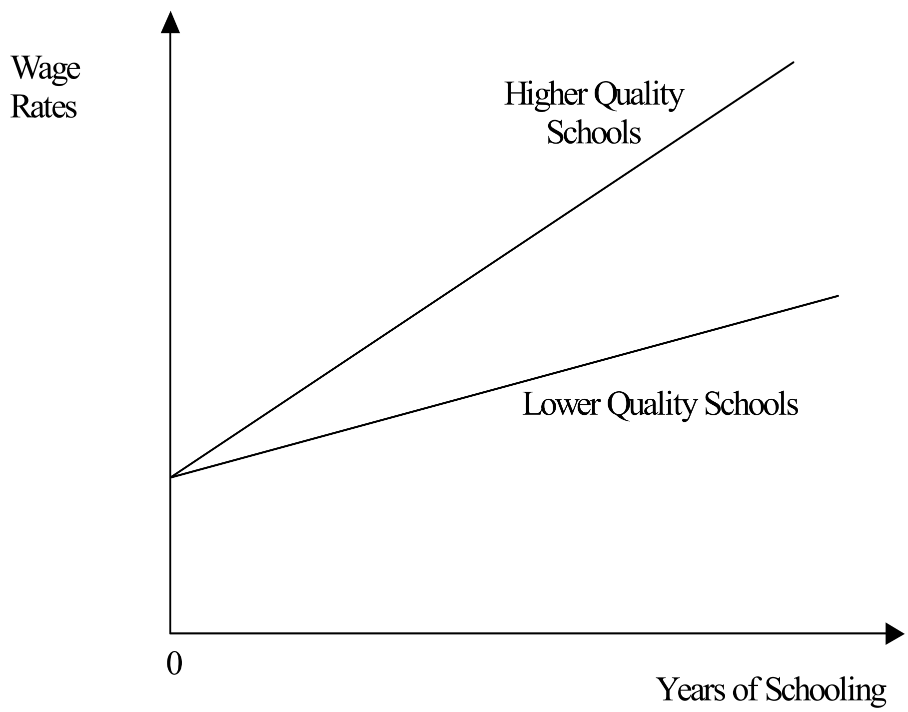


FIG. 2.1.—Illustrative wage-schooling profiles contrasting students from lower and higher quality schools.

Turning to the literature, the estimated models of wages that attempt to capture the impact of school quality on earnings can be categorized by how quality and years of schooling enter the earnings equation. The variations can be justified by differences in the specification of $e(Q_s,s)$. These estimates nearly always examine the log of wages[18] and can be categorizes as follows, with the labels in parentheses used to link the model to the discussion below:

$$
\begin{array}{ll}
(I) & \ln(w) = X\beta + \alpha Q + \varepsilon \\
(I+) & \ln(w) = X\beta + \alpha Q + \gamma S + \varepsilon \\
(S) & \ln(w) = X\beta + \delta QS + \varepsilon \\
(S+) & \ln(w) = X\beta + \gamma S + \delta QS + \varepsilon \\
(IS) & \ln(w) = X\beta + \alpha Q + \delta QS + \varepsilon \\
(IS+) & \ln(w) = X\beta + \alpha Q + \gamma S + \delta QS + \varepsilon
\end{array}
$$

Estimates that include no interaction between quality and schooling might be referred to as intercepts models (I) as the quality effect shifts the (log) earnings-schooling equation with no impact on the schooling gradient. The "+" designation identifies models that have an independent schooling effect, i.e., one that is unaffected by the quality component.[19] Specifications such as equation (2.3) in Section I justify the functional form (I+).[20] Models that contain a school quality-years of schooling interaction term might be referred to as slope models (S). Equation (2.4) above justifies this functional form. Finally, some estimates include both intercept and slope components (IS or IS+) and can be justified by the function $e(Q_s,s) = e^{\alpha Q + \gamma s + \delta Q s}$.

The theory itself does not have anything to say regarding which specification is preferred; it depends on the true functional form of $e(Q_s,s)$ and the data must determine which is correct. It is curious that the forms (I), (I+), (IS), and (IS+) predict higher wages even for workers with no schooling, provided that the parameter estimates are

---

[18] Some of the earlier studies examined hourly/weekly/annual earnings rather than the log of earnings.

[19] Because the wage specification is in logs, the difference between expected earnings levels for those attending different quality schools grows as schooling increases.

[20] If the function $e(Q_s,s)$ is separable in Q and s, then specification (I+) is justified. Q may be a measurable characteristic such as the pupil-teacher ratio or it might be a monotonic transformation, such as the logarithm, of the characteristic. As such, without loss of generality, either $e(Q_s,s) = (\alpha Q)e^{\gamma s}$ or $e(Q_s,s) = e^{\alpha Q + \gamma s}$ justify (I+).

positive. This may be justified if working alongside better trained workers raises the productivity of uneducated workers, perhaps through better management and supervision. Alternatively, Q may be set to zero for workers with no schooling. Regardless of the form, however, all of the models dictate that within a market, those educated in schools with higher quality will earn more and those from higher quality schools will have steeper wage-schooling gradients.

# CHAPTER III
# THE MISIDENTIFICATION OF SCHOOLS ATTENDED

## Introduction

The census based studies that attempt to find a link between earnings and school quality invariably match statewide average school characteristics from the *Biennial Surveys* and the *Digest of Education Statistics* to the birth state under the assumption that workers are educated in their birth state.[21] The PSID based studies follow a similar procedure, matching statewide *Biennial Survey* data to the state in which workers lived when they were 12 years old, whereas the NBER-Thorndike and Hagen data based studies match workers to school characteristics from the *Biennial Surveys*, but at the district or city level.

The remainder of the school quality-earnings literature relies upon data from the NLSY79, NLS-YM and NLS-YW, HSB, and NLS72, (or NLS-based studies) all of which include characteristics of a school attended by the respondent during a single academic year for those in these surveys.[22] The NLSY79, for example, surveyed the last high school attended as of 1979 for those ages 17 or more.

The assumption in these studies is that birth state or residence state at a given age or school attended during a single year can be used to identify school quality during the entirety of a student's academic career. The purpose of this chapter is to determine whether the assumption adopted in the studies that use the *Biennial Survey* data, that birth state is the same as the school attendance state, is valid. I also examine the prevalence of students changing schools during their academic careers so that, for these students, the last high school attended or the school attended during a single academic year may be a poor proxy for the quality of elementary and secondary schooling.

---

[21] Welch (1966) is the exception, linking school characteristics to the residence state and restricting the population to rural farm workers.
[22] For the sophomore cohort of the HSB survey, however, data was collected during the sophomore and senior years.

**Birth State and Attendance State**

The statewide averages reported in the *Biennial Survey* and the *Digest of Education Statistics* include data from public elementary and secondary schools in each state. Table 3.1 presents data from the 1990 U.S. Census for white, non-Hispanic boys ages three to eighteen that are enrolled in elementary and secondary school. These sample restrictions are imposed to mimic the populations studied in most of the school quality-earnings literature and to focus on the levels of schooling that are included in measuring the characteristics of schools. These students are categorized into (1) those attending public schools in their birth state, (2) those attending private school in their birth state, and (3) those attending public or private school outside their birth state. The *Biennial Survey* data is directly applicable to those in the first group, but not the latter two.[23]

After the break at age six, when many children attending private preschools and kindergartens transfer to public elementary schools, the percent attending public school in their birth state is remarkably stable at 70 percent. This figure balances a gradual 6 percent reduction in those attending private school between the ages six and eighteen with an offsetting gradual increase in those moving outside their birth state. Because the data is a cross-section rather than a panel, it is not clear what fraction of students complete their high school degree entirely at public schools within their birth states, but it is clear that a significant portion of students attain some of their training elsewhere suggesting that the use of birth state as an identifier for schooling state results in a non-trivial amount of error. At a minimum, 30 percent don't attend public elementary or high school for some part of their academic career in their birth state, and the true percentage may be significantly higher given enrollments after age five in private schools and moves after age five from the student's birth state. It is remarkable that

---

[23] Students in the first category who are born in a state, move away, and then return to their birth state are an exception; a part of their academic careers will be spent at other schools. The percentage of students in this category, identified by the residence state five years ago, is between 1.5 – 2.0% for the ages presented in the table.

nearly one in five (17.5 percent) students have left their birth state by the time they begin elementary school and likely never attend school there.

**Table 3.1**
**Elementary and Secondary Enrollment Rates in 1990 by School Type, Whether Living in Their Birth State, and Age**

| Age | Percent Attending Public School In Their Birth State | Percent Attending Private School In Their Birth State | Percent Attending School Outside Their Birth State | Observations |
|---|---|---|---|---|
| 3 | 29.4 | 56.6 | 14.0 | 10,979 |
| 4 | 33.9 | 51.1 | 15.0 | 22,767 |
| 5 | 55.6 | 28.4 | 16.0 | 39,797 |
| 6 | 70.6 | 11.9 | 17.5 | 54,684 |
| 7 | 71.3 | 10.8 | 17.8 | 57,197 |
| 8 | 71.1 | 10.5 | 18.4 | 55,767 |
| 9 | 70.8 | 10.0 | 19.2 | 58,531 |
| 10 | 71.1 | 9.4 | 19.5 | 57,395 |
| 11 | 70.3 | 9.2 | 20.6 | 56,004 |
| 12 | 70.0 | 8.9 | 21.1 | 54,590 |
| 13 | 69.8 | 8.5 | 21.6 | 53,459 |
| 14 | 69.5 | 8.1 | 22.4 | 52,551 |
| 15 | 70.1 | 7.4 | 22.5 | 53,076 |
| 16 | 69.4 | 7.2 | 23.5 | 50,758 |
| 17 | 69.6 | 6.7 | 23.7 | 48,353 |
| 18 | 70.0 | 5.8 | 24.2 | 26,806 |

SOURCE.-1990 U.S. Census of Housing and Population, PUMS 5% Sample.

NOTE.-Figures include white, non-Hispanic males ages 3 to 18 that are born in the United States, enrolled in school, have not completed high school, are not living in group quarters, and do not have allocated values for any of the relevant variables. School enrollment information is not available for children under age three.

**Table 3.2**
**Percent of Enrolled Children Not Attending Public School in Their Birth State by Age for the Years 1960 – 2000**

| | Census Year | | | | |
|---|---|---|---|---|---|
| **Age** | **2000** | **1990** | **1980** | **1970** | **1960** |
| 3 | 66.2 | 70.6 | 80.4 | 81.5 | – |
| 4 | 59.8 | 66.1 | 74.6 | 74.3 | – |
| 5 | 43.5 | 44.4 | 41.4 | 34.4 | 27.6 |
| 6 | 30.0 | 29.4 | 29.9 | 28.5 | 30.2 |
| 7 | 29.5 | 28.7 | 29.9 | 29.1 | 31.0 |
| 8 | 29.4 | 28.9 | 30.4 | 28.7 | 30.8 |
| 9 | 29.5 | 29.2 | 30.9 | 30.3 | 31.3 |
| 10 | 30.3 | 28.9 | 31.3 | 29.7 | 30.9 |
| 11 | 30.4 | 29.8 | 31.9 | 30.7 | 31.2 |
| 12 | 30.4 | 30.0 | 32.3 | 30.2 | 30.3 |
| 13 | 30.9 | 30.2 | 31.8 | 30.1 | 30.6 |
| 14 | 30.8 | 30.5 | 31.1 | 28.8 | 29.3 |
| 15 | 30.3 | 29.9 | 30.2 | 27.9 | 28.7 |
| 16 | 30.1 | 30.6 | 30.0 | 28.2 | 26.9 |
| 17 | 29.5 | 30.4 | 29.9 | 28.3 | 27.2 |
| 18 | 29.8 | 30.1 | 28.5 | 27.3 | 26.0 |
| Total | 32.7 | 32.3 | 32.8 | 30.1 | 30.1 |

SOURCE.-The 1960 to 2000 U.S. Census of Housing and Population, PUMS files.

NOTE.-Figures include white, non-Hispanic males ages 3 to 18 who are born in the United States, enrolled in school, have not completed high school, are not living in group quarters, and do not have allocated values for any of the relevant variables. School enrollment information is not available for children under the age of three in 1970 to 2000 and for children under the age of five in 1960. For 1960 and 1970, those of Hispanic origin are not identified. Available allocation indicators change over time. I use all allocation information in a given year to make the populations as comparable as possible.

Although Table 3.1 examines 1990 data, the overwhelming majority of the literature focuses on students who completed schooling prior to 1990. As such, Table 3.2 presents the same tabulations for 1960 to 2000 using the U.S. decennial censuses. Because the focus is on the number of mismatches created when data for public schools from the Biennial Survey is matched to the birth state, I present only the percentage of children attending school outside of their birth state or attending private school. The samples are restricted to match those in Table 3.1 as closely as possible.

Over the forty-year period presented in Table 3.2, the percent of children for whom the *Biennial Survey* data is an improper match is surprisingly stable at 30 percent. In fact, if the data is restricted to match the ages with available information in 1960, the percentages are 30.1%, 29.4%, 31.2%, 30.5%, and 30.9% for the years 1960 to 2000, respectively.

To illustrate the magnitude of the problem generated by the misidentification of the state where a worker attended school, suppose that there are only two states with school quality constant within each state but differences between the two states and the quality of schools in State A is twice that of those in State B. Suppose that 30 percent of those born in State A are educated in State B and 30 percent of those born in State B are educated in State A. Assume that all workers attend for exactly twelve years and then sell their skills in a single market. Workers educated in State A have twice the education of those educated in State B and, controlling for other differences in skills, will earn *twice* as much. What if birth state is used as a proxy for attendance state? Workers born in State A will earn only *1.31* times as much as those born in State B.[24] The noise created by a 30 percent error rate in identifying the state where a worker attended reduces the school quality-earnings signal by 70 percent. Schooling inputs will be

---

[24] Let the price of education be one with all other skills the same on average for those born in the two states. Recall that wages are equal to the price of education times the amount of education. Since all workers complete the same number of years of schooling in the example, only the quality difference matters. Seventy percent of workers born in State A are educated in State A and thirty percent are educated in State B. The average school quality for those born in State A is 1.7. A similar calculation for those born in State B yields a value of 1.3. Thus, the relative wages of those born in State A to those born in State B is 1.31.

undervalued and may be statistically insignificant and deemed to be unimportant, when in fact they have a large impact on earnings.

Returning to the data, while the percentage of students for whom birth state fails to identify the attendance state and school type is stable at thirty percent at each age between six and eighteen and between 1960 and 2000, there may be differences in the error rate across birth states or residence states. Table 3.3 presents the same percentages as in Table 3.1 differentiated by birth state. The population follows the same restrictions as above, except it is restricted to those aged six to eighteen years old. Again, the first percentage calculation reported, those enrolled in public school in their birth state, are those for whom attendance state and school type match the specification in the statewide quality data. The three calculations reported in Table 3.3 contain a considerable amount of variation across states. For example, only 1.4 percent of children born in D.C. are attending public school in D.C. whereas 81.6 percent of children born in Maine attend public school in Maine and the private school enrollment rate in Louisiana is nearly eight times as high as private school enrollment rate in Idaho. The bottom line is that the accuracy of the identification of quality varies by birth state. Although not presented here, it is also true that these numbers are fairly stable over time in that states that are poor matches in 1990, such as D.C., are poor matches for all years between 1960 and 2000.

Table 3.4 examines the same information as in Table 3.3 from a different angle. It presents the percentage of students attending school outside their birth state in the last column for each state. For example, 49.3 percent of students attending public schools in Nevada where born elsewhere and their quality measures will not reflect the time spent in attendance in Nevada schools.

**Table 3.3**
**Elementary and Secondary Enrollment Rates in 1990 by School Type, Whether Living in Their State of Birth, and Birth State**

| Birth State | Attending Public School In Their Birth State (Pct) | Attending Private School In Their Birth State (Pct) | Attending School Outside of Their Birth State (Pct) | Birth State | Attending Public School In Their Birth State (Pct) | Attending Private School In Their Birth State (Pct) | Attending School Outside of Their Birth State (Pct) |
|---|---|---|---|---|---|---|---|
| AL | 74.0 | 7.3 | 18.7 | MT | 68.1 | 3.1 | 28.9 |
| AK | 51.1 | 3.3 | 45.6 | NE | 66.6 | 8.5 | 25.0 |
| AZ | 63.9 | 4.7 | 31.4 | NV | 59.0 | 4.0 | 37.0 |
| AR | 73.6 | 4.9 | 21.6 | NH | 66.2 | 6.2 | 27.6 |
| CA | 70.2 | 9.7 | 20.1 | NJ | 67.6 | 12.0 | 20.4 |
| CO | 62.0 | 4.7 | 33.4 | NM | 53.5 | 4.3 | 42.3 |
| CT | 70.4 | 9.9 | 19.8 | NY | 69.1 | 13.0 | 17.9 |
| DE | 51.1 | 15.0 | 33.9 | NC | 76.5 | 4.9 | 18.6 |
| DC | 1.4 | 4.0 | 94.6 | ND | 64.9 | 3.5 | 31.5 |
| FL | 65.1 | 10.1 | 24.9 | OH | 73.1 | 10.6 | 16.3 |
| GA | 72.6 | 6.9 | 20.5 | OK | 70.7 | 3.9 | 25.4 |
| HI | 28.0 | 9.5 | 62.5 | OR | 69.8 | 5.5 | 24.7 |
| ID | 66.7 | 2.2 | 31.1 | PA | 71.3 | 14.6 | 14.1 |
| IL | 67.3 | 12.4 | 20.3 | RI | 64.3 | 10.2 | 25.5 |
| IN | 74.9 | 7.0 | 18.1 | SC | 69.0 | 7.4 | 23.7 |
| IA | 71.7 | 7.0 | 21.3 | SD | 66.7 | 3.7 | 29.6 |
| KS | 69.1 | 5.5 | 25.4 | TN | 72.5 | 6.3 | 21.3 |
| KY | 73.3 | 6.8 | 19.9 | TX | 74.3 | 6.0 | 19.7 |
| LA | 61.8 | 16.9 | 21.3 | UT | 76.9 | 1.6 | 21.6 |
| ME | 81.6 | 3.9 | 14.6 | VT | 73.9 | 4.5 | 21.6 |
| MD | 64.7 | 11.4 | 23.9 | VA | 66.7 | 6.1 | 27.2 |
| MA | 70.4 | 10.6 | 19.0 | WA | 74.0 | 6.2 | 19.8 |
| MI | 76.9 | 9.5 | 13.6 | WV | 71.5 | 2.8 | 25.8 |
| MN | 77.1 | 8.4 | 14.5 | WI | 74.0 | 12.9 | 13.1 |
| MS | 63.4 | 10.7 | 26.0 | WY | 55.4 | 1.2 | 43.4 |
| MO | 68.0 | 11.6 | 20.4 | Total | 70.3 | 9.0 | 20.7 |

SOURCE.-1990 U.S. Census of Housing and Population, PUMS 5% Sample.

NOTE.-Figures include white, non-Hispanic males ages 6 to 18 that are born in the United States, enrolled in school, have not completed high school, are not living in group quarters, and do not have allocated values for any of the relevant variables.

**Table 3.4**
**Elementary and Secondary Enrollment Rates in 1990 by School Type, Whether Living in Their State of Birth, and State of Residence**

| Residence State | Attending Public School in Their Birth State (Pct) | Attending Private School (Pct) | Attending Public School Outside Their Birth State (Pct) | Residence State | Attending Public School in Their Birth State (Pct) | Attending Private School (Pct) | Attending Public School Outside Their Birth State (Pct) |
|---|---|---|---|---|---|---|---|
| AL | 70.4 | 9.4 | 20.2 | MT | 68.9 | 4.8 | 26.3 |
| AK | 48.3 | 5.3 | 46.4 | NE | 71.1 | 11.2 | 17.7 |
| AZ | 52.8 | 6.5 | 40.7 | NV | 45.5 | 5.2 | 49.3 |
| AR | 68.3 | 6.1 | 25.6 | NH | 57.1 | 9.3 | 33.6 |
| CA | 72.1 | 11.6 | 16.4 | NJ | 67.4 | 15.1 | 17.5 |
| CO | 60.4 | 6.5 | 33.2 | NM | 49.0 | 7.2 | 43.8 |
| CT | 69.5 | 12.8 | 17.7 | NY | 76.4 | 15.4 | 8.2 |
| DE | 53.8 | 21.1 | 25.1 | NC | 71.5 | 6.7 | 21.8 |
| DC | 15.9 | 64.8 | 19.3 | ND | 77.6 | 4.9 | 17.6 |
| FL | 51.7 | 11.6 | 36.8 | OH | 76.3 | 12.4 | 11.3 |
| GA | 65.5 | 8.8 | 25.7 | OK | 69.2 | 5.3 | 25.5 |
| HI | 39.3 | 21.0 | 39.7 | OR | 67.2 | 7.1 | 25.7 |
| ID | 66.7 | 3.5 | 29.8 | PA | 72.5 | 16.4 | 11.1 |
| IL | 72.5 | 15.0 | 12.5 | RI | 69.1 | 14.0 | 16.9 |
| IN | 76.1 | 8.9 | 15.0 | SC | 65.1 | 9.7 | 25.2 |
| IA | 77.7 | 8.9 | 13.4 | SD | 72.2 | 5.2 | 22.6 |
| KS | 67.5 | 7.7 | 24.8 | TN | 70.8 | 8.1 | 21.1 |
| KY | 75.6 | 8.9 | 15.5 | TX | 72.2 | 7.3 | 20.5 |
| LA | 67.4 | 20.8 | 11.9 | UT | 81.2 | 2.5 | 16.4 |
| ME | 72.5 | 4.9 | 22.6 | VT | 66.9 | 6.4 | 26.8 |
| MD | 59.0 | 15.9 | 25.1 | VA | 61.7 | 8.5 | 29.8 |
| MA | 76.7 | 13.0 | 10.3 | WA | 65.2 | 7.5 | 27.3 |
| MI | 79.3 | 11.0 | 9.7 | WV | 76.1 | 3.9 | 20.1 |
| MN | 77.1 | 9.7 | 13.1 | WI | 73.6 | 14.3 | 12.0 |
| MS | 65.0 | 13.8 | 21.2 | WY | 56.7 | 2.3 | 41.0 |
| MO | 68.6 | 13.5 | 17.9 | Total | 70.3 | 10.9 | 18.8 |

SOURCE.-1990 U.S. Census of Housing and Population, PUMS 5% Sample.

NOTE.-Figures include white, non-Hispanic males ages 6 to 18 who are born in the United States, enrolled in school, have not completed high school, are not living in group quarters, and do not have allocated values for any of the relevant variables.

**School Measures from a School Actually Attended**

The decennial census includes information about each person's residence – both state and PUMA code – on April 1st five years prior to the census year. The PUMA code is a census designated geographic area based on county or parish boundaries with more populous counties being divided into sub-counties and less populous counties being aggregated. The PUMA code of residence five years ago does not, however, identify the sub-divisions for larger counties. Table 3.5 presents information on the migration of enrolled students ages six to eighteen using the 1990 Census. Those that moved across state or PUMA code boundaries likely changed schools. Nearly one in five (18.5%) twelve year olds, for example, moved at least once to a new residence within the last five years that almost surely involved a change in schools. The next category, intra-PUMA code moves, may involve a change in schools as they could include moves across county boundaries for sparsely populated counties or to the far side of a large city, but they could also include moves to the house next door. The last category, non-movers, would not have changed schools unless they transferred between private and public schools. What is clear from Table 3.5 is that many school-aged children move during their academic career and the characteristics of the last high school attended may not accurately capture the quality of pre-college education.[25]

[25] If students move between neighborhoods with similar quality schools, then the identification of the quality of schools attended from the one school observed may be adequate.

**Table 3.5**
**Percent of Enrolled Children Moving Within the Last Five Years Categorized by the Type of Move (Geographically) and Age**

| Age | Moving to Another State (Pct) | Moving to Another PUMA Code (Within State) (Pct) | Moving within PUMA Code (Pct) | Residing in the Same House (Pct) | Obs. |
|---|---|---|---|---|---|
| 6 | 11.7 | 13.4 | 30.1 | 44.8 | 52,998 |
| 7 | 11.3 | 12.6 | 28.8 | 47.4 | 55,846 |
| 8 | 10.5 | 11.7 | 27.9 | 50.0 | 54,527 |
| 9 | 10.2 | 11.1 | 26.9 | 51.8 | 57,348 |
| 10 | 9.7 | 10.4 | 26.2 | 53.7 | 56,348 |
| 11 | 9.4 | 9.9 | 24.8 | 55.9 | 55,048 |
| 12 | 9.2 | 9.3 | 23.2 | 58.3 | 53,660 |
| 13 | 9.0 | 8.8 | 22.6 | 59.6 | 52,600 |
| 14 | 8.8 | 8.5 | 21.3 | 61.4 | 51,706 |
| 15 | 8.2 | 8.1 | 20.8 | 62.9 | 52,313 |
| 16 | 7.8 | 7.6 | 19.8 | 64.8 | 50,098 |
| 17 | 7.1 | 6.9 | 18.8 | 67.2 | 47,732 |
| 18 | 6.5 | 6.5 | 18.8 | 68.2 | 26,465 |
| Total | 9.3 | 9.8 | 24.2 | 56.7 | 666,689 |

SOURCE.-1990 U.S. Census of Housing and Population, PUMS 5% Sample.

NOTE.-Figures include white, non-Hispanic males ages 6 to 18 that are born in the United States, enrolled in school, have not completed high school, are not living in group quarters, and do not have allocated values for any of the relevant variables.

**Conclusions**

Some of the debate in the literature has focused on potential problems with the school quality data such as how statewide averages ignore within state variance or how school-specific quality data may be more prone to measurement error. What has not been emphasized is that both statewide averages and school specific data from a single year may inadequately reflect a student's full elementary and secondary career.

Using census data, I show that at a minimum 30 percent of those born in a state do not attend public elementary or high school for some fraction of their pre-college school career in that state and that nearly one in five never attend school in their birth state. Using high school characteristics from a school attended during a single academic year does not eliminate the measurement error problem, as the fraction of students switching elementary or high schools also appears to be high. Importantly, the measurement error in school quality created by the misidentification of the state or school attended greatly weakens the ability of the data to identify a school quality effect on earnings.

This is not to say that other sources of measurement error – such as those created by ignoring intrastate variation in quality, not to mention intra-school variations in quality – are not important. They may be. But, the first step is to identify the schools (or even the states) where the student attended during their full academic career. The next chapter extends the discussion of measurement error focusing on the confounding influence of college attendance when the quality of the college education is completely ignored.

The debate over which school characteristics correctly capture the influential components of quality is also an important one, especially to policy makers. But, because the ability to match correctly the characteristics of schools to workers who attended them is so poor, any discussion regarding which characteristics of schools impact the earnings capacity of workers is premature.

# CHAPTER IV
# THE CONFOUNDING INFLUENCE OF COLLEGE-TRAINED WORKERS

## Introduction

The literature review contained in Appendix A provides a synopsis of thirty-six studies that examine the link between school quality and earnings. Of these, twenty-seven include elementary and/or high school quality measures only, although just one (Welch, 1966) excludes those with college training. Of the remaining nine studies, five focus solely on the quality of the colleges attended, leaving four – Wachtel (1975, 1976), Morgan and Sirageldin (1968), and Tremblay (1986) – that include quality measures for both the high school and college attended. Wachtel's two studies examine the NBER-Thorndike data, which identifies the high school and college attended by name and location, and use primarily expenditures per pupil at the high school and college levels to measure quality. Morgan and Sirageldin examine the PSID and match elementary and high school expenditures per pupil from the *Biennial Surveys* to the state where the worker grew up (residence at age 12) and a ranking of colleges to capture quality.[26] Tremblay examines the NLS-YM, which includes expenditures per pupil in the district where respondents attended high school and expenditures per student at the most recent college attended. The main point here is that, although almost all studies include those who have attended college, very few examine the impact of college quality on earnings jointly with the impact of elementary and high school quality and that the omission of college quality can lead to inconsistent and biased results.

This chapter begins by discussing the importance of the measurement error introduced when college quality is ignored in obtaining estimates of school quality effects on earnings. I then present evidence that the magnitude of the error is non-trivial as high school quality provides very little information about college quality. Moreover,

---

[26] Cass and Birnbaum, *Comparative Guide to American Colleges* (1964). Morgan and Sirageldin do not describe how the college index is matched to the worker.

when the gradient method is used to evaluate school quality, the estimated gradients are dominated by the impact of those who have attended college. I then revisit the work of Card and Krueger (1992a) to examine whether their estimates are robust to the exclusion of those with college training. Finally, I discuss a major obstacle in matching college quality to workers in census-based studies. The work in this chapter emphasizes models such as Card and Krueger's, that examine the impact of school quality on school-earnings gradients using census data; however, the confounding influence of including college trained labor while ignoring college quality creates problems whether a slope model or an intercept model is estimated (see Chapter II) and it also creates problems whether Census-style (state level quality measures matched to birth state) data or NLS-style (school level quality measures from a school actually attended) data are analyzed.

**The Importance of Omitted College Quality: A Simple Model**

Returning to the model in Chapter II, recall that wages are the product of the amount of education possessed (*e*) and the price of education ($P_e$) in the market where education is sold and that education is equal to school quality (*Q*) times the number of years attended (*s*). Let the econometric model be

$$y = \alpha + P_e e + u = \alpha + (P_e Q)s + u \tag{4.1}$$

for illustrative purposes. For those working in the same market, $P_e$ is constant so that, if earnings are regressed on schooling, differences in the relative coefficients on *s* for those educated in different systems reflect differences in school quality. The true amount of a worker's education is

$$e = \begin{cases} sQ_H & \text{if } s <= 12 \\ 12Q_H + (s-12)Q_C & \text{if } s > 12 \end{cases} \tag{4.2}$$

where $Q_H$ is elementary and high school quality and $Q_C$ is college quality. Suppose that equation (4.1) is estimated substituting $e^* = sQ_H$ for $e$ for all levels of schooling. If the population is restricted to those with no college training, then the estimated coefficient, $P_eQ$, is consistent and unbiased since e*=e. There is no measurement error in e* for this group. Suppose, however, that those with college training are included in the regression. For those extending their schooling beyond high school, the measurement error in e* is $\upsilon = (s-12)(Q_C - Q_H)$. If $Q_C$ and $Q_H$ are perfectly (and positively) correlated (and can be rescaled to have similar means and variances), then there is no measurement error. However, if the quality measures are not perfectly correlated, then the estimates of $P_eQ$ are inconsistent and downward biased in the single variable regression model. Clearly, the correlation between $Q_H$ and $Q_C$ determines the importance of ignoring college quality. The amount of bias in $\hat{\beta}$ (=$P_eQ$) is asymptotically equal to

$$\text{plim}\,\hat{\beta} = \frac{\beta\sigma_e^2}{\sigma_e^2 + \sigma_\upsilon^2} \leq \beta \qquad (4.3)$$

and $\sigma_\upsilon^2$ depends on the correlation between $Q_H$ and $Q_C$.

Wachtel's (1975) is the only study in the quality-earnings literature to include any information on the relationship between $Q_H$ and $Q_C$. He regresses the total expenditures for post-secondary schooling (per student in the undergraduate and graduate institution attended) on high school expenditures (per student in average daily attendance for the school district attended). Also included in the regression are post-high school achievement test scores (an aggregate index whose creation is not described), the number of siblings, mother's education, father's education, and age. He finds that high school quality is a statistically significant predictor of college quality. He also presents standardized regression (beta) coefficients for all of the independent variables in a path diagram. The beta coefficients may be derived by multiplying the OLS estimates, $\hat{\beta}$, by $\sigma_x/\sigma_y$ (using the standard regression notation). In a one variable model, the standardized regression coefficient is the simple correlation between *x* and *y*. In a multiple regression, such as Wachtel's, they are similar to a conditional correlation and Wachtel (1975) reports the beta coefficient of $Q_C$ on $Q_H$ as 0.09.[27]

*How Much Information Does $Q_H$ Carry About $Q_C$?*

What does a (statistically significant) correlation of 0.09 mean? One interpretation is that if $Q_H$ is one standard deviation higher for Student 1 than for Student 2, then it is expected that $Q_C$ will be 0.09 standard deviations higher for Student 1. Alternatively we might ask, how much higher must $Q_H$ be for Student 1 so that we are 95% confident that $Q_C$ for Student 1 will also be higher? The difference in $Q_H$ for Students 1 and 2 can be expressed as the proportion of the population spanned by the two observations. Table 4.1 presents these proportions, using X and Y as the variables under the distributional assumptions given at the bottom of the table. For example, if the correlation coefficient is 0.5 and $X_1 > X_2$, then $X_2$ needs to be at the 6th percentile and $X_1$ needs to be at the 94th percentile before we are 90% confident that $Y_1$ will be higher than $Y_2$. At a correlation of 0.10, the entire population must be spanned before we can be confident at any of the levels presented that college quality for Student 1 will be higher than college quality for Student 2.

---

[27] The beta coefficients are not technically the conditional correlation because the variances used to adjust $\hat{\beta}$ are not conditional on the other factors in the regression. The unconditional correlation between $Q_H$ and $Q_C$ could be computed given the beta coefficients between the other independent variables in the regression and the dependent variable ($Q_C$), which he reports, and given the unconditional correlations between the other independent variables and $Q_H$, which he does not report. Wachtel reports an $R^2$ of 0.1323 for the regression described in the text, which implies an upper bound on the correlation between $Q_H$ and $Q_C$ of 0.3637; however, as all of the other variables in the regression are significant, the simple correlation is probably much lower and 0.09 likely serves as a good approximation.

**Table 4.1**
**The Proportion of the Population Spanned by Two Observations, with $X_1>X_2$, to be Able to Conclude with a Given Confidence Level that $Y_1>Y_2$**

| Correlation Coefficient ($\rho$) | Confidence Level | | |
|---|---|---|---|
| | **75.00%** | **90%** | **95%** |
| 0.00 | – | – | – |
| 0.10 | 100.00% | 100.00% | 100.00% |
| 0.25 | 93.50% | 100.00% | 100.00% |
| 0.36 | 78.40% | 98.10% | 99.70% |
| 0.50 | 59.10% | 88.30% | 95.60% |
| 0.75 | 32.60% | 57.60% | 69.50% |
| 0.90 | 18.30% | 33.90% | 42.70% |
| 1.00 | 0.00% | 0.00% | 0.00% |

NOTE.-Assumes that X and Y are distributed as a bivariate normal with $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$, and correlation coefficient $\rho$ .

Another question we could ask is, given that $Q_H$ is higher for Student 1, what is the probability that $Q_C$ will also be higher? Under the same distributional assumptions as in Table 4.1, the calculation can be made given the correlation coefficient and the size of the difference between $X_1$ and $X_2$, measured as the proportion of the population spanned by the population. These probabilities are reported in Table 4.2. If there is zero correlation between X and Y, then the probability that $Y_1>Y_2$ is always 50 percent and if they are perfectly (and positively) correlated, it is always 100 percent. The 38.3 percent of the population spanned by $X_1$ and $X_2$ calculation is of particular interest as this is equal to a one standard deviation difference in $X_1$ and $X_2$ (around the center of a normal distribution). With a correlation of 0.10, $Y_1$ will be greater than $Y_2$ only 52.8% of the time, a very marginal difference between this and the completely uncorrelated situation.

**Table 4.2**
**The Probability that $Y_1>Y_2$ Given that $X_1>X_2$.**

| Correlation Coefficient ($\rho$) | Proportion of the Population Spanned By $X_1$ and $X_2$ | | | | | |
|---|---|---|---|---|---|---|
| | **10%** | **25%** | **38.3%** | **50%** | **75%** | **90%** |
| 0.00 | 50.00% | 50.00% | 50.00% | 50.00% | 50.00% | 50.00% |
| 0.10 | 50.70% | 51.80% | 52.80% | 53.80% | 56.50% | 59.20% |
| 0.25 | 51.80% | 54.60% | 57.20% | 59.70% | 66.30% | 72.60% |
| 0.36 | 52.70% | 56.90% | 60.80% | 64.40% | 73.50% | 81.50% |
| 0.50 | 54.10% | 60.30% | 65.80% | 70.90% | 82.60% | 91.00% |
| 0.75 | 58.00% | 69.50% | 78.90% | 86.00% | 96.70% | 99.60% |
| 0.90 | 64.30% | 82.40% | 92.80% | 97.60% | 100.00% | 100.00% |
| 1.00 | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |

NOTE.-Assumes that X and Y are distributed as a bivariate normal with $\mu_X = \mu_Y = 0,\ \sigma_X = \sigma_Y = 1$, and correlation coefficient $\rho$.

In answer to the question posed at the beginning of this sub-section, "How much information does $Q_H$ carry about $Q_C$?", the answer is, "Not much." The quality of the elementary and high school attended is nearly independent of the quality of the college attended.

*The Influence of College Trained Labor on the Estimated Schooling-Earnings Gradients*

Consider again the simple regression of wages on school completion rewritten from equation (4.1) above

$$y = \alpha + \beta s_l + u \tag{4.4}$$

with *y* being income or wage or its logarithm and $s_l$ being the school years associated with level *l*. In this case the calculated gradient, $\hat{\beta}$, is a measure of the incremental value of a year of schooling and, given the theory, is thought to be positively related to the quality of the schools attended.

The regression calculation is $\hat{\beta} = \sum_{i,l} y_{il}(s_{il} - \overline{s}) / \sum_{i,l}(s_{il} - \overline{s})^2$, which simplifies to $\hat{\beta} = \sum_{l} N_l y_l (s_l - \overline{s}) / \sum_{l} N_l (s_l - \overline{s})^2$ where $N_l$ is the number of observations at level $l$ and $y_l$ is the mean of $y$ at $l$ and $\overline{s}$ is the population mean of $s$.

Now consider the differences between the adjacent values of $y_l$, normalized by the school year differences between them:

$$\begin{cases} \tilde{y}_1 = (y_1 - y_0)/(s_1 - s_0) \\ \tilde{y}_2 = (y_2 - y_1)/(s_2 - s_1) \\ \cdots \\ \tilde{y}_N = (y_N - y_{N-1})/(s_N - s_{N-1}) \end{cases} \tag{4.5}$$

Each value in this series measures the average incremental value of a year of schooling over the completion interval indicated. Since $\hat{\beta}$ measures an average increment over the full span, $s_0$ to $s_N$, it must be an average of the components in (4.5). After some manipulation, presented in Appendix B, this calculation can be rewritten as

$$\hat{\beta} \equiv \sum_{j=1}^{N} \omega_j \tilde{y}_j \tag{4.6}$$

where $\omega_j = (s_j - s_{j-1}) \sum_{i=j}^{N} N_i (s_i - \overline{s}) / \Delta$ with $\Delta = \sum_{\ell=0}^{N} N_\ell (s_\ell - \overline{s})^2$. It is also true that $\omega_j > 0$ and $\sum_{j=1}^{N} \omega_j = 1$ so that the average in (4.6) is a simple one and the weights depend exclusively on the distribution of completion levels.

**Table 4.3**
**National Average (Log) Weekly (Experience-Adjusted) Wage Differences and Weights for Adjacent Levels of Schooling, 1980 Census**

| Grades | Weight ($\omega_j$) | Wage Difference ($\tilde{y}_j$) | Grades | Weight ($\omega_j$) | Wage Difference ($\tilde{y}_j$) |
|---|---|---|---|---|---|
| 1-0 | 0.0014 | 0.0546 | 11-10 | 0.0766 | 0.0443 |
| 2-1 | 0.0020 | 0.0216 | 12-11 | 0.0846 | 0.0933 |
| 3-2 | 0.0032 | -0.0366 | 13-12 | 0.1221 | 0.0805 |
| 4-3 | 0.0051 | 0.0914 | 14-13 | 0.1225 | 0.0437 |
| 5-4 | 0.0076 | 0.0020 | 15-14 | 0.1145 | 0.0391 |
| 6-5 | 0.0108 | 0.0624 | 16-15 | 0.1085 | 0.1911 |
| 7-6 | 0.0168 | 0.0599 | 17-16 | 0.0707 | 0.0286 |
| 8-7 | 0.0261 | 0.0698 | 18-17 | 0.0552 | 0.0587 |
| 9-8 | 0.0486 | 0.0680 | 19-18 | 0.0359 | 0.0088 |
| 10-9 | 0.0630 | 0.0451 | 20-19 | 0.0248 | 0.0713 |

SOURCE.-1980 U.S. Census of Population and Housing, PUMS (A Sample) 5 percent file.

NOTE.-The sample is restricted to white, non-Hispanic males ages 30 to 59 that were born in the U.S., are not living in group quarters, are not in school and do not have any allocated values for the sample restriction criteria or the variables used to compute the weights or wage differences. Wage differences are experience adjusted, where experience is defined as the minimum of age minus years of schooling minus 6 and age minus 16. Wages are weekly wages and are bottom coded and top coded at the $2^{nd}$ and $98^{th}$ percentiles.

Table 4.3 provides the national average (experience-adjusted) log wage differentials associated with an additional year of schooling ( $\tilde{y}_j$ ) and the weights ( $\omega_j$ ) for each year of schooling calculated using the 1980 Census. Figures are provided from 1980 for two reasons: 1) this examines the same data analyzed by Card and Krueger (1992a) in their highly cited paper, and 2) the 1980 decennial census was the last to report schooling by years completed for the full schooling distribution.[28] First, notice that wage growth is not uniform across the various levels. For example, there is a clear "sheepskin" effect with completion of high school (the 12-11 grade differential) adding 9.3 percent to weekly wages and with completion of college (the 16-15 differential) adding 19.1 percent. Since the regression method of averaging depends on the schooling distribution, it follows that if regressions were run on wage increments like those listed in Table 4.3, the calculated averages will vary with differences in distributions of completion levels. For the data presented in Table 4.3, 43.8 percent completed at least one year of college and the sum of the weights ( $\omega_j$ ) for those completing at least one year is 65.4 percent. Most importantly, the fraction of the U.S. schooling-earnings gradient attributable to those that attended college is 69.0 percent.[29] In other words, although 69 percent of Card and Krueger's wage growth is attributable to college-trained workers included in their analyses, they do not include any measure of college quality and college quality is virtually unrelated to high school quality. This suggests that the amount of downward bias in their estimated impact of school quality may be substantial.

The discussion in Chapter II stressed the importance of examining school quality effects within markets. In Table 4.4, I present the schooling-earnings gradients and the percent of the gradient due to college trained workers for each residence state (assuming that residence states and markets are synonymous).

---

[28] In 1990, the years of schooling variable reports the degree received for those with more than a high school degree and combines all those with some college training, but no degree, into a single group.

[29] The overall U.S. gradient is $\hat{\beta} \equiv \sum_{j=1}^{N} \omega_j \tilde{y}_j$ and the fraction attributable to college-trained labor is

$$\hat{\beta} \equiv \sum_{j=13}^{N} \omega_j \tilde{y}_j .$$

**Table 4.4**
**Schooling-Earnings Gradients Calculated Using Workers of All Schooling Levels and the Amount of the Gradient Attributable to College-Trained Workers by Residence State**

| State | Linear Gradient | Percent Of Gradient Attributable to Those Completing At Least One Year of College | State | Linear Gradient | Percent Of Gradient Attributable to Those Completing At Least One Year of College |
|---|---|---|---|---|---|
| AL | 0.0617 | 51.91 | MT | 0.0560 | 77.18 |
| AK | 0.0900 | 89.23 | NE | 0.0591 | 74.21 |
| AZ | 0.0638 | 80.20 | NV | 0.0647 | 78.76 |
| AR | 0.0530 | 53.84 | NH | 0.0577 | 67.55 |
| CA | 0.0711 | 86.38 | NJ | 0.0745 | 79.78 |
| CO | 0.0657 | 86.34 | NM | 0.0620 | 84.58 |
| CT | 0.0714 | 81.36 | NY | 0.0684 | 77.13 |
| DE | 0.0669 | 77.14 | NC | 0.0560 | 58.84 |
| DC | 0.0719 | 100.00 | ND | 0.0568 | 66.25 |
| FL | 0.0590 | 73.42 | OH | 0.0709 | 67.43 |
| GA | 0.0625 | 57.99 | OK | 0.0604 | 68.26 |
| HI | 0.0629 | 96.86 | OR | 0.0637 | 75.88 |
| ID | 0.0581 | 80.18 | PA | 0.0663 | 73.11 |
| IL | 0.0747 | 72.96 | RI | 0.0600 | 66.89 |
| IN | 0.0689 | 60.59 | SC | 0.0585 | 59.53 |
| IA | 0.0628 | 72.68 | SD | 0.0512 | 69.28 |
| KS | 0.0614 | 76.30 | TN | 0.0592 | 50.34 |
| KY | 0.0623 | 44.94 | TX | 0.0693 | 74.91 |
| LA | 0.0710 | 60.41 | UT | 0.0649 | 82.98 |
| ME | 0.0484 | 61.10 | VT | 0.0489 | 57.71 |
| MD | 0.0728 | 75.67 | VA | 0.0670 | 66.02 |
| MA | 0.0650 | 76.52 | WA | 0.0697 | 83.43 |
| MI | 0.0763 | 73.24 | WV | 0.0643 | 44.50 |
| MN | 0.0678 | 73.58 | WI | 0.0657 | 71.99 |
| MS | 0.0562 | 58.52 | WY | 0.0682 | 85.07 |
| MO | 0.0637 | 66.78 | U.S. | 0.0688 | 69.03 |

SOURCE.-1980 U.S. Census of Population and Housing, PUMS (A Sample) 5 percent file.
NOTE.-See notes for Table 4.3.

First, note the large amount of variation in gradients across states. Increasing schooling by one year for workers in Maine only increases earnings by 4.8 percent on average whereas the same increase in schooling will produce a 9.0 percent change in earnings in Alaska. The difference in gradients across states is statistically significant. (The F-Statistic is 435.30, with 50 and 965,490 degrees of freedom, with a p-value of less than 0.0001.) This suggests that the U.S. is not a single market or that there are systematic, uncaptured differences in the amount of education possessed by workers across states. In addition, alongside the wide variation in the gradients is the variation in the weight given to college attendees – in West Virginia 44.5 percent of the gradient is due to college trained labor whereas 100 percent of the gradient in D.C. is due to this higher educated group.[30]

The final table in this section, Table 4.5, extends the calculations of Table 4.4. The Card and Krueger maintained hypothesis is that we can learn something about school quality by observing workers educated in difference states (identified by birth states) working in the same market (identified by residence states). Table 4.5 presents the percent of the schooling-earnings gradient attributable to those who attended college for those in a given residence state (market) from a given birth state (school system) for the five states with the largest white, non-Hispanic male population among those ages 30 to 59. In the top panel, we again see large variances across states, and we also see large variances for those born in different states living in the same state and for those born in the same states and living in different states. Examining those born in Texas (reading across the row), for example, 62.7 percent of the schooling gradient for those remaining in Texas is due to college trained labor whereas 98.5 percent of the schooling gradient for those moving to New Jersey is due to college trained labor. The bottom panel presents the number of male workers in the sample including all levels of schooling.

---

[30] Because the wage differences do not have to be strictly positive, it is possible for the "fraction" attributable to college trained workers to be above one. This is particularly a problem where very few workers have given levels of schooling as is true of those with a high school degree or less living in D.C. In instances where the percent is above 100, I report a value of 100.0.

Turning the comparison on its head, for those living in Ohio (reading down the column), we see the same kind of variation across birth states in the fractions of the gradient attributable to college workers ranging from a low for those born in Ohio (57.2 percent) to a high for those born in Florida (100 percent). These results are suggestive of selectivity in interstate migration; those born in Texas and remaining have less schooling or smaller wage increments than those moving from Texas to New Jersey. The selectivity of migration is discussed further in the next chapter.

**Table 4.5**
**The Amount of the Schooling-Earnings Gradient Attributable to College-Trained Workers in 1980 by Residence and Birth State Pairs**

| | Residence State | | | | |
|---|---|---|---|---|---|
| **Birth State** | **CA** | **FL** | **NJ** | **OH** | **TX** |
| | | | *Percentages* | | |
| **CA** | 83.3 | 100.0 | 100.0 | 92.9 | 81.8 |
| **FL** | 83.7 | 57.1 | 76.6 | 100.0 | 84.4 |
| **NJ** | 89.1 | 78.7 | 69.7 | 59.4 | 99.1 |
| **OH** | 91.3 | 69.1 | 88.6 | 57.2 | 91.7 |
| **TX** | 88.7 | 76.9 | 98.5 | 93.6 | 62.7 |
| | | | *Observations for All Schooling Levels* | | |
| **CA** | 27,942 | 368 | 140 | 187 | 838 |
| **FL** | 411 | 6,557 | 80 | 114 | 369 |
| **NJ** | 1,281 | 1,336 | 18,741 | 310 | 360 |
| **OH** | 2,969 | 2,189 | 320 | 36,753 | 924 |
| **TX** | 2,237 | 366 | 94 | 135 | 28,777 |

SOURCE.-1980 U.S. Census of Population and Housing, PUMS (A Sample) 5 percent file.
NOTE.-See notes for Table 4.3 and the text for a description of the sample.

It should be noted that the states in Table 4.5 are selected only for illustrative purposes due to the number of observations in each state. They were *not* selected because the weights that are associated with college attendance are high or highly varied. The calculations in Tables 4.3, 4.4, and 4.5 hint at the amount of measurement error created by including college trained workers in the analyses while ignoring college quality when the impact of quality is measured through the schooling-earnings gradients. The importance of college-trained labor in estimated schooling-earnings gradients strains the credibility of school quality effects estimated through this avenue.

**Alternative Estimates of School Quality Using the Card/Krueger Data**

This section examines the robustness of Card and Krueger's (1992a) findings of a positive school quality effect on earnings to the exclusion of college-trained workers. This is not to imply that their estimates are robust when the gradient method is used as they report estimates that vary significantly given differences in their specification.[31]

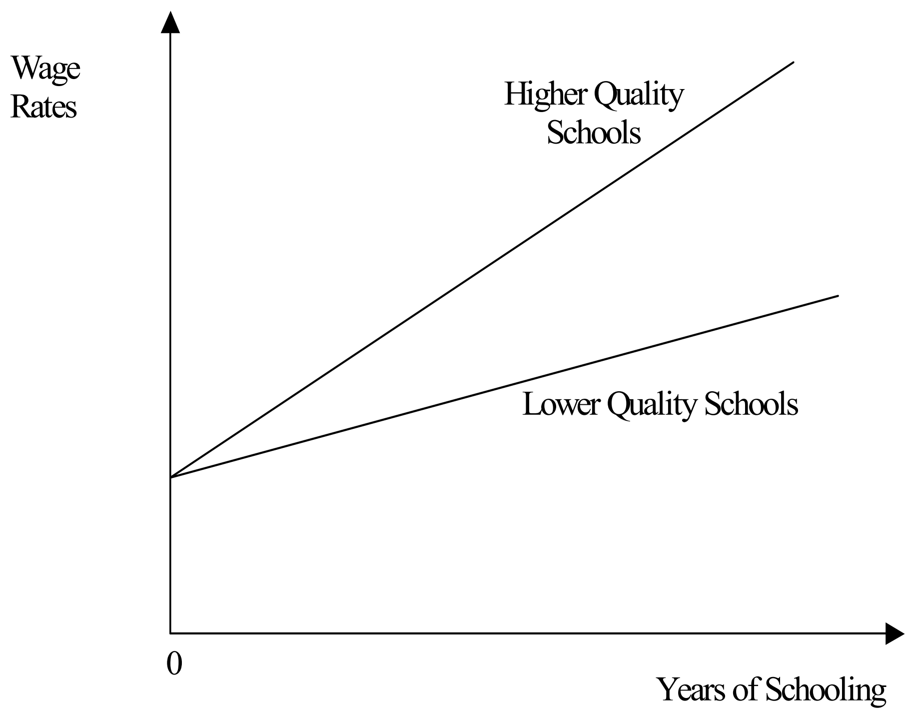


FIG. 4.1.—Revisiting Figure 2.1 (Illustrative wage-schooling profiles contrasting students from lower and higher quality schools.)

[31] Heckman, et. al. (1996) further examine the robustness of the Card/Krueger findings to changes in the empirical specification and also find that their results are not robust.

Figure 4.1 (copied from Chapter II) is presented here for illustrative purposes. As discussed in Chapter II, there are two ways to examine differences in school quality that are manifest through earnings. The first way is through differences in the schooling-earnings gradient, those attending higher quality schools will gain more with each additional year of schooling completed, i.e., the gradient is steeper. The other method observes that at each level of schooling, earnings are higher for those from the higher quality school. Inclusion of college attendees can confound the comparisons using either method and the simplest solution to the problem is to exclude those with college training.

*Quality Effects Measured Through Schooling-Earnings Gradients*

Card and Krueger examine 1980 Census data for three birth cohorts of white men partitioned by year of birth: 1920-29, 1930-39, and 1940-49. Their estimates of school quality effects are obtained in two steps. The first involves a separate regression for each birth cohort. The observations are individuals and the model is

$$\begin{aligned}\log(w) = \delta_b{}'\alpha_{bc} + \delta_r{}'\alpha_{rc} + \delta_c{}'\alpha_c + \alpha_{1c}EXP + \alpha_{2c}EXP^2 \\ +\alpha_{3c}SMSA + \alpha_{4c}MSP + (\delta_b{}'\beta_{bc} + \delta_d{}'\beta_{dc})S^* + \varepsilon\end{aligned} \quad (4.7)$$

Their calculation examines log weekly wages (*w*) and the model includes indicators for state of birth ($\delta_b$), state of residence ($\delta_r$), cohort ($\delta_c$), residence within an SMSA, married with spouse present (MSP), a quadratic in potential work experience (EXP—the approximate years out of school)[32], state of birth and division of residence (nine groups of states) interactions with years of schooling ($S^*$) and an error term ($\varepsilon$). The implicit assumption in step one is that the division of residence is the relevant market and the state of birth schooling-earnings gradients are estimated within these markets. The

[32] There are 49 birth states (excluding Alaska and Hawaii and including the District of Columbia) and 51 residence states (including the District of Columbia).

asterisk indicates that years of schooling is a transform of the usual variable and is defined as $S^*=\max(0,S-S_{.02})$, where S refers to years of school completed and $S_{.02}$ refers to the second centile in the completed schooling distribution within the birth cohort and birth state. The "c" subscripts on the parameters α and β indicate that they shift between cohorts.

The three first step regressions provide 147 (49x3) birth state parameter estimates for the schooling interactions. Only 144 of these parameters are identified since each estimate refers to an implicitly arbitrary residence division. The second step uses the 147 birth state schooling gradient effects as dependent variable observations and regresses them on indicators of birth cohort and on school system characteristics. There are two sets of second step estimates, one with and the other without indicators for birth state. The authors indicate a preference for the version of the model that includes the birth state indicators. Notice in that case that the school characteristic effects are identified only through cohort-birth state interaction, i.e., through variation within states in cohort differentials. Observations are weighted by the inverse of the variance of the birth state schooling gradients obtained in the first step. The second step model, including birth state indicators, is

$$\hat{\beta}_{bc} = \delta_b{}'\gamma_{bc} + \delta_c{}'\gamma_c + \theta Q_{bc} + \eta \qquad (4.8)$$

The schooling gradient is seen as depending on cohort, birth state (in some specifications), and school characteristics in the birth state for the cohort ($Q_{bc}$). Within each birth state and cohort, the school characteristics are averages computed under the assumptions that each person attends public elementary and high schools[33] for twelve years beginning at age six and that the number of people born during each year of the birth cohort is a constant.

---

[33] The *Biennial Survey* data used by Card and Krueger reports statewide average characteristics of public elementary and secondary schools.

The three main school quality measures studied by Card and Krueger are statewide averages of teachers' wages, pupil/teacher ratios, and term lengths. Teachers' wages are normalized using average state-level wages (due to cost of living differences across states) and then rescaled within cohort to have a mean of one (due to cost of living differences over time).

A problem with the two-step procedure is that the weights to be used in the second step, i.e., the inverse variances of the birth state gradients, are not well defined and the second step estimates are sensitive to the weights to be chosen. In my attempt to replicate their work, weights that produced estimates that more or less match the ones reported by Card and Krueger were found as were weights that give somewhat different results. Alternatively, the estimates could be derived in a single step, inserting equation (4.8) into equation (4.7), and this resolves the weighting problem.

A second problem involves the schooling truncation; it is anomalous. The specification implies that schooling has no effect on earnings until the second centile value is passed. In Texas for the oldest cohort, that value is 2 years of schooling while in California for the youngest cohort it is 9 years. Thus, Texas-born third grade graduates are treated as school-transformed-equivalents to tenth grade graduates from California schools. The idea seems to be to identify those who are uneducated; the problem is that this doesn't do it.

Table 4.6 provides the main results for the models Card and Krueger estimate.[34] Panel A lists their estimates for purposes of comparison. All remaining panels include one-step estimates where state school system characteristics are substituted directly into the model. Panel B parallels the Card-Krueger procedure exactly; Panel C repeats the Panel B calculations with years of school completed substituted for the truncated measure. Panel D restricts observations to men who have not attended college and uses

---

[34] I define experience as minimum(age-years of schooling-6, age-16). This may differ from Card and Krueger's experience variable, which appears to be age-years of schooling-6. In addition, birth quarter is used to adjust age to December 31, 1979 and top-coded annual earnings ($75,000) are multiplied by 4/3 of the top value in calculating weekly wages.

the truncated education measure while Panel E uses the same observations but substitutes years of schooling for the truncated measure.

**Table 4.6**
**Alternative Estimates of the Contribution of State School System Characteristics to the Earnings-Schooling Gradient – Card and Krueger Revisited (Regression Coefficients with Standard Errors in Parentheses)**

| | Characteristics | | |
|---|---|---|---|
| **Model:** | **Pupils Per Teacher** | **Term Length** | **Relative Teacher Wage** |
| ***Including All Education Levels*** | | | |
| A. Card-Krueger reported estimates | | | |
| without birth state effects | -2.38 (1.64) | 1.93 (0.94) | 1.35 (0.33) |
| with birth state effects | -9.35 (3.18) | -0.02 (0.99) | 0.97 (0.44) |
| B. Card-Krueger one step estimates | | | |
| without birth state effects | -2.05 (0.75) | -0.04 (0.40) | 1.75 (0.14) |
| with birth state effects | -4.40 (2.44) | -1.86 (0.75) | 0.67 (0.35) |
| C. Card-krueger one step estimates using non-truncated schooling | | | |
| without birth state effects | -2.46 (0.73) | 0.72 (0.39) | 1.59 (0.14) |
| with birth state effects | -4.71 (2.38) | -1.20 (0.74) | 0.85 (0.34) |
| ***Excluding Men Who Have Attended College (One Step Estimates)*** | | | |
| D. Card-Krueger truncated schooling | | | |
| without birth state effects | -3.94 (1.71) | -3.28 (0.83) | 2.05 (0.34) |
| with birth state effects | 1.85 (0.56) | -4.39 (1.56) | -0.54 (0.89) |
| E. Non-truncated schooling | | | |
| without birth state effects | -1.22 (1.59) | -0.34 (0.78) | 0.90 (0.31) |
| with birth state effects | 1.25 (0.51) | -1.94 (1.45) | 1.58 (0.80) |

SOURCE.-1980 U.S. Census of Population and Housing, PUMS (A Sample) 5 percent file.

NOTE.-See text for a description of the models. Coefficients are recoded. The pupil/teacher ratio is divided by 10,000 as is the number of days in the school term. The relative teacher wage is scaled within cohort to have a mean value (unweighted across states) of 0.01.

In comparing Panel A and B (the Card-Krueger two step estimates to the identical specification with one step) it becomes clear that the estimated effects of the pupil/teacher ratio and the teacher's relative wage are substantially the same. Estimates of effects of longer school terms, that are unstable between the Card-Krueger alternative specifications in any case, are different, however. In particular, the estimated effect is perverse in the model with independent birth state effects.

In comparing Panels B and C it appears that the truncation issue is moot. If men who attended college are included, as in the Card and Krueger calculations, it simply does not matter whether the observed school completion level is used or whether the truncated version is substituted for it. The estimates in Panels D and E, where men who attended college are excluded, are sensitive to the truncation.

Overall, the estimated school term effects are generally perverse; longer terms *decrease* the value of schooling, but the estimates are, at best (or at worst), marginally significant. If attention is restricted to men who have not attended college, the estimates for effects of the pupil/teacher ratio are unstable in the sense that signs differ depending on whether the Card-Krueger preferred specification is used. The preferred specification allows independent birth state effects so that school effects are identified only by cohort-birth state interaction. In these models higher pupil/teacher ratios detract (with marginal statistical significance) from the earnings power of schooling if observations include men who have attended college, but when observations are restricted to those with at most a secondary education, the estimated effect is perverse: higher numbers of students per teacher are estimated as *increasing* the value of schooling.

Over all of the estimates there is one robust finding, namely that higher relative wages for teachers are associated with higher schooling gradients. It is unclear, however, whether this reflects a school quality phenomena or some other relationship. Recall that teachers' wages are divided by state-level average wages. During the time period when these cohorts of men were in school, most teachers were college graduates and most workers were at most high schools graduates. In this case, the teachers' wage variable may be thought of as a simple schooling gradient – the average wage of those with a college degree divided by the average wage of those with a high school degree. Moreover, most workers continue to live in their birth state. As such, the teachers' wage effect may merely reflect that states with high returns to schooling in the past continue to have high returns to schooling. The evidence presented in the next sub-section supports this idea.

*Quality Effects Measured Through Wages at Each Schooling Level*

Table 4.7 considers a completely different partition of the data in which observations are divided on the basis of years of school completed. Within each schooling level, the estimated model is of the form:

$$\begin{aligned}\ln(\textit{weekly wages}) = {} & \delta_c{}'\alpha_c + \delta_r{}'\alpha_r + \alpha_1 SMSA \\ & + \alpha_2 MSP + \alpha_3 X + \alpha_4 X^2 + Q'\beta + u\end{aligned} \tag{4.9}$$

To determine whether an increased level of one of the school characteristics is associated with an increased schooling gradient, choose two levels of schooling and subtract the coefficient corresponding to the lower of the two levels from the coefficient for the higher of the two. For example, in the interval between the eighth and tenth grades the estimate is that increased numbers of pupils per teacher will reduce the income gained from the two years of additional schooling.

**Table 4.7**
**Alternative Estimates of Effects of School System Characteristics on Earnings**

| Years of School Completed | Characteristic: Pupils/Teacher | Term Length | Relative Teacher Wage |
|---|---|---|---|
| 8 | 14.00 (12.41) | -5.19 (6.05) | -3.97 (2.48) |
| 10 | 3.92 (12.71) | 4.09 (6.04) | -6.62 (2.24) |
| 12 | -9.17 (4.27) | -7.04 (2.28) | 1.61 (0.79) |
| 16 | -14.55 (7.19) | 0.36 (3.96) | 2.40 (1.29) |

SOURCE.-1980 U.S. Census of Population and Housing, PUMS (A Sample) 5 percent file.
NOTE.-See text for a description of the models. The scaling of school characteristics is the same as in Table 4.6.

On inspection, however, we see that this intuitively appealing result stems from perverse underlying relations. Both at 8 and again at 10 years of schooling higher pupil/teacher ratios are associated with higher wages. In this case the effect of the pupil/teacher ratio on the schooling 10/schooling 8 wage ratio is as would be predicted on the basis of a school quality argument, but as is clear from the underlying component estimates, the "right" result is derived from the "wrong" underlying relations.[35]

If we take the school quality idea seriously, we expect that exposure to an improved environment enhances productivity and that longer exposure results in greater enhancement. Thus increased school quality will augment worker productivity at each school completion level *and* the augmentation will increase with the level of schooling– if the level of schooling proxies the length of exposure to the enhanced school environment. The problem is more complicated if we consider individuals who have attended college, however. Their exposure to elementary and secondary school characteristics presumably tops at 12 years just as it does for high school graduates who have not attended college.

Since there is no information concerning the quality of colleges, the measured wage response to variation in elementary and secondary school characteristics should be viewed as potentially biased due to the omission of the college data. If elementary and

[35] For this comparison, the net effect is –10.08, which is 3.92 (tenth grade effect) minus 14.00 (the eighth grade effect).

secondary school characteristics are uncorrelated with college characteristics, then the expected wage response to elementary and secondary school characteristics should be somewhat smaller for college than for high school graduates. This due both to the extra lapsed time between exposure in school and subsequent use on the job and also to the competition on the job between skills learned in college and those acquired earlier.

In Table 4.7, the wage responses estimated for the pupil/teacher ratio are uninterpretable when viewed over the four completion levels. Two of the four response estimates are perverse (8 and 10 years of schooling). The largest response that is not perverse is the one estimated for college graduates for whom the estimates are ambiguous at best. Note, however, that the estimated responses decrease with school completion. Thus, increases in the pupil/teacher ratio unambiguously reduce the wages of those with more relative to those with less schooling. An analysis that considered schooling gradients only—such as the Card-Krueger study—finds these results unobjectionable.

The erratic, negative this time positive next, estimates of wage responses to increased length of school terms require no further comment. That leaves the relative wage of teachers.

The schooling level regressions in Table 4.7 show perverse teachers' wage responses for 8 and 10 years of schooling. The response at 12 years is small and statistically insignificant, but the sign is what one might expect if the teachers' wage variable proxies teacher quality. The estimate for college graduates suggests, however, that the teachers' wage variable does not proxy quality. As discussed above, this variable is a ratio, teachers' salaries normalized by statewide average wages. The positive covariance with wages of college graduates along with the negative covariance with wages of high school dropouts and the near zero covariance with wages of high school graduates is suggestive of a serial correlation argument. According to it, the covariance between the teacher wage construct and the wages of men with more-relative to those with less- schooling is nothing more than an observation that state differentials in gradients tend to persist for a long time. In states where wages of teachers (and, presumably, other college graduates) have historically been high, wages of college graduates continue to be high. Regardless of whether the serial correlation argument appeals, the pattern of wage responses to the teacher relative wage is hardly convincing that what resulted from the Card-Krueger analysis of schooling gradients is a meaningful estimate of the wage gains to be derived from increased investment in superior teachers. If attention is restricted to those who have not attended college, the positive gradient response results from perverse underpinnings, the negative response at lower grades.

**Matching Measures of College Quality to Workers**

This last section examines a major obstacle with the census-based studies in matching the characteristics of colleges to workers. With the NLSY and the NBER-Thorndike data, and perhaps several other non-census databases, it is possible to obtain the actual college attended by a worker and several measures of college quality are readily available in sources such as IPEDS and many college rankings. Using these data sources, it is possible to include both high school and college quality in the estimation. The census data, however, does not contain any information about the college attended. It might be tempting to apply the same methodology used for elementary and high school characteristics, i.e., match average statewide college characteristics to workers based on their birth state. However, the main problem associated with this methodology for pre-college characteristics, that many workers have mismatched school characteristics, extends to college-trained workers. Table 4.8 presents an extension of Table 3.1 for those enrolled in college.

**Table 4.8**
**Post-Secondary Enrollment by School Type and Whether Living in Their Birth State Reported by Age**

| | | | Of Those Enrolled | | |
|---|---|---|---|---|---|
| **Age** | **Not Enrolled (Percent)** | **Enrolled (Percent)** | **In Public School in Birth State (Percent)** | **In Private School In Birth State (Percent)** | **Outside of Birth State (Percent)** |
| 18 | 36.34 | 63.66 | 63.84 | 12.53 | 23.62 |
| 19 | 44.85 | 55.15 | 61.14 | 12.89 | 25.97 |
| 20 | 51.07 | 48.93 | 58.40 | 12.84 | 28.76 |
| 21 | 54.33 | 45.67 | 56.17 | 13.36 | 30.47 |
| 22 | 63.26 | 36.74 | 53.82 | 12.24 | 33.94 |
| 23 | 73.15 | 26.85 | 52.43 | 12.35 | 35.22 |
| 24 | 78.82 | 21.18 | 48.30 | 13.00 | 38.70 |
| 25 | 81.91 | 18.09 | 46.68 | 12.78 | 40.54 |
| 26 | 85.02 | 14.98 | 43.80 | 12.64 | 43.55 |
| 27 | 86.74 | 13.26 | 42.44 | 12.63 | 44.94 |
| 28 | 88.12 | 11.88 | 42.18 | 12.88 | 44.93 |
| 29 | 89.18 | 10.82 | 42.82 | 12.38 | 44.80 |
| 30 | 90.36 | 9.64 | 42.60 | 12.14 | 45.26 |
| 31 | 91.21 | 8.79 | 43.32 | 12.51 | 44.17 |
| 32 | 91.82 | 8.18 | 42.91 | 12.16 | 44.94 |
| 33 | 92.24 | 7.76 | 43.22 | 11.37 | 45.41 |
| 34 | 92.74 | 7.26 | 43.64 | 11.42 | 44.94 |
| 35 | 92.90 | 7.10 | 43.24 | 11.83 | 44.94 |
| 36 | 93.72 | 6.28 | 44.19 | 11.00 | 44.81 |
| 37 | 94.06 | 5.94 | 42.35 | 12.48 | 45.17 |
| 38 | 94.23 | 5.77 | 43.61 | 13.18 | 43.21 |
| 39 | 94.50 | 5.50 | 44.00 | 10.21 | 45.79 |
| 40 or More | 97.30 | 2.70 | 39.38 | 11.02 | 49.59 |

SOURCE.-1990 U.S. Census of Housing and Population, PUMS 5% Sample.

NOTE.-Figures include white, non-Hispanic males ages 18 or more, not living in group quarters, born in the United States, that have completed high school and do not have allocated values for any of the relevant variables.

Given the figures presented in Chapter III, these results should not be surprising. At age eighteen, 64 percent of white men are enrolled in school. Of those enrolled, 63.8% are enrolled in a public school in their birth state, 12.5 percent are attending a private school in their birth state, and 23.6 percent are attending in another state. As age increases, the percent enrolled in a public school in their birth state declines as the percent enrolled out-of-state increases with the percent attending a private, in-state institution remaining constant. Using IPEDS data matched to workers based on birth state would accurately map statewide average public college characteristics to, at best, 64 percent of those with college training. Datasets that identify the actual college(s) attended have a decisive advantage over those that do not, but only if that information is exploited in the analysis.

**Conclusions**

The measurement error model developed at the beginning of this chapter is very simplistic and is meant for illustrative purposes only. It is possible, perhaps probable, that elementary/high school and college and graduate school quality cannot be aggregated. And, it is not clear how elementary/high school quality impacts earnings for college graduates. It may be that higher quality high schools raise the probability of attending college (or how hard you have to work when you get there) and that for college attendees, only the college quality matters. A similar argument could be made for graduate school so that only the quality of the institution for the highest relevant degree matters. In hiring new assistant professors, how much value does the quality of the undergraduate institution carry? Maybe a lot, but not necessarily in the expected direction. Consider a job candidate in the economics department who's undergraduate degree is from the University of Chicago and their Ph.d. is from the University of Houston. This candidate might be viewed very differently than one who reversed the order of the institutions. In this example, however, the higher quality undergraduate institution (coupled with the lower quality graduate institution) sent a signal about the quality of the candidate rather than the quality of the schooling and is meant to illustrate that maybe the quality of earlier training plays only a minor role in identifying the quality of relevant schooling. Measuring the total amount of education possessed by an individual, which includes both years of schooling and the quality of that schooling, may be a complicated specification exercise. Although the argument that only the quality of the highest degree matters has some intuitive appeal, it is ironic that the impact of elementary/high school characteristics in the Card/Krueger levels comparisons are strongest, and sometimes only correctly signed, for those with college training.

In this chapter, calculations were performed that excluded those with college training so as to avoid the confounding influence of omitted college quality. However, restricting the data to those with at most twelve years of schooling might create other problems. These calculations assumed that all other things are equal. If elementary/high school quality increases the probability of attending college, then more workers will become college educated in states with higher elementary and high school quality. Suppose that the probability of attending college is also (positively) dependent on the ability of the student (or some other factors that are positively correlated with earnings) and that when the quality of pre-college schooling is increased, the marginal students who now attend college are of higher average ability than those that still do not attend college. The ability levels of those remaining without any college training will be lower on average and the comparisons are now confounded by the "other things" that are not equal and not included in the regressions.

Alternatively, suppose that the within state variance of school quality is higher in State B than in State A and that a line is drawn such that those who attend schools of a given quality or more continue their schooling beyond high school. This scenario is depicted in Figure 4.2. In the diagram, the distributions for the two states are drawn so that the average quality levels are equal. If college-trained workers are excluded from the comparisons, then those remaining from State B will have attended lower quality schools than those from State A. This relationship, however, is not captured in the statewide averages. It is easy to see that the distribution for State B could be shifted to the right, so that the average quality of schools in State B is higher than the average quality of schools in State A, but among those with at most a high school degree the opposite is true.

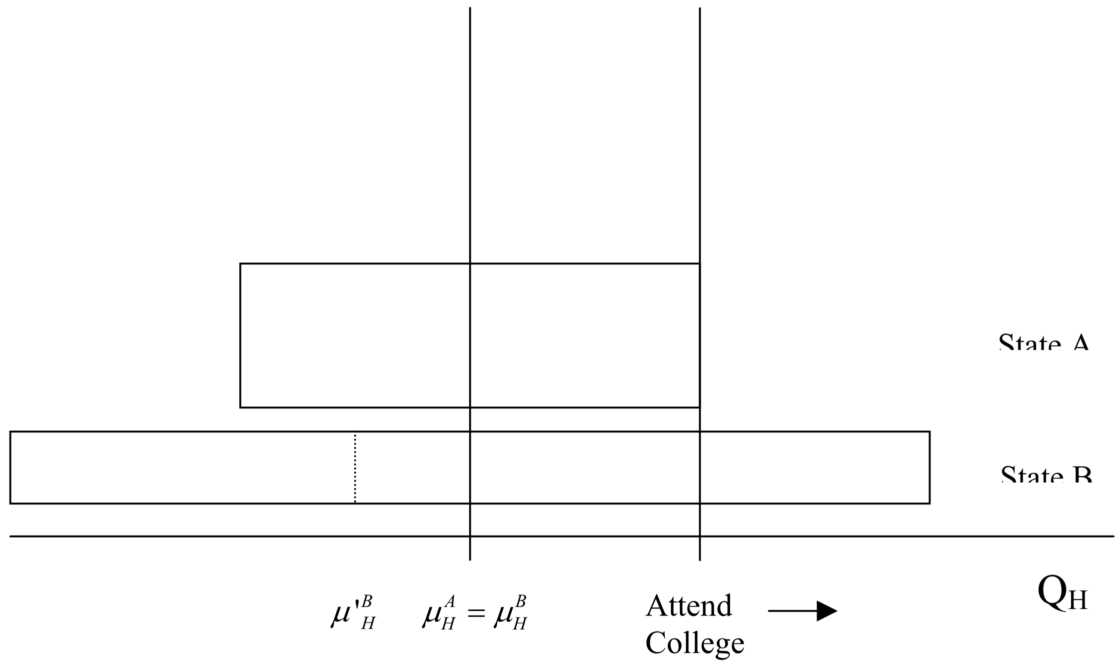


FIG. 4.2.—An example of how excluding college-trained labor biases the pool of remaining workers.

The point of these examples is not to describe all of the possible scenarios that might exist, but to make the point that when college-trained labor is excluded from the analysis, it is not clear whether the remaining workers are the same on average for all relevant characteristics.  As such, these comparisons may be no more meaningful than those that include workers with college training while ignoring the quality of that training.

# CHAPTER V
# ARE THE DATA CONSISTENT WITH THE THEORETICAL PREDICTIONS?

## Introduction

The economic theory formulated in Chapter II has clear empirical predictions. Within a well-defined labor market[36], where the price of education is the same for all workers, those educated in higher quality school systems (or states) will earn more among those with the same level of schooling, all other things equal. If Texas has higher quality schools than Oklahoma, then workers educated in Texas with twelve years of schooling will earn more than workers educated in Oklahoma with the same level of schooling. Moreover, this ranking will hold in Texas and in Oklahoma and in every other state, assuming that a state defines a market and that all other things are held equal. This is empirically testable.

If the data do not support the theory, then either the markets are poorly defined or other things are not equal, i.e., there is something selective about migration. Focusing on the second explanation, if workers educated in Texas earn more than those from Oklahoma when the comparison is made in California but the ranking is reversed when the comparison is made in New York, then this would suggest that those moving from Texas to New York are different than those moving from Texas to California or that those moving from Oklahoma to New York are different than those moving from Oklahoma to California or, perhaps, both. In this circumstance, wage rankings are meaningless unless skill differences in workers can be observed and comparisons can be adjusted for these differences.

In this chapter, I first present relative wage rankings for selected levels of schooling including white, non-Hispanic men residing in the same state who were born

[36] Card and Krueger (1992a) defined the market as census division (nine groups of states). In Speakman and Welch (2006), the market is defined as the residence state. I adopt the latter convention.

in different states. The assumption being that residence state defines the relevant labor market and differences in wages for men from different birth states reflect differences in the quality of schooling. Because wage rankings for those born in a given state vary tremendously across markets, the remainder of the chapter presents a preliminary examination of migration issues. The purpose is not to resolve the problem, but simply to point out that this is a complicated issue that has only been addressed in passing in the literature, yet never resolved. An issue, however, that undermines the ability of the school quality-earnings studies to identify a believable, causal link.

**Wage Rankings by Birth States Across Markets**

I begin this section by making the comparisons discussed in the introduction. They are of average log wages for white, non-Hispanic men (after adjusting to the mid-point of the age interval) for those from different birth states evaluated within each birth cohort, schooling level and residence state. The data is taken from the 2000 U.S. Census and the sample restrictions are discussed in the note below Table 5.1. The birth cohorts are defined as those ages 30 to 39, those ages 40 to 49, and those ages 50 to 59, mimicking Card and Krueger (1992a).

**Table 5.1**
**Log Wage Differentials (Measured Relative to a Reference Wage) For Men Living in a Given State from Different Birth States for High School and College Graduates Reported for Selected Residence and Birth States**

*High School Graduates*

| Ages: 30 to 39 | | | | | Birth State | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Residence State | CA | IL | IN | MA | MI | NJ | NY | OH | PA | TX |
| AZ | -12.1 | 8.6 | -3.3 | 1.5 | -5.7 | -21.2 | 4.7 | -16.3 | -1.4 | 2.1 |
| CA | -0.8 | -2.4 | 11.0 | 4.8 | 3.6 | 18.4 | 6.5 | 2.6 | -6.2 | -2.6 |
| CO | -1.0 | -9.5 | 6.9 | 5.8 | 20.4 | -8.0 | 6.9 | 11.5 | -1.1 | -2.9 |
| FL | -9.8 | -0.2 | -1.1 | -5.0 | 2.0 | -3.3 | 7.6 | -0.5 | -0.7 | 12.4 |
| GA | -9.4 | 11.5 | 1.0 | 3.7 | 1.4 | -13.7 | -7.4 | -5.9 | 10.0 | -1.2 |
| NC | 1.7 | -9.9 | 20.0 | 10.5 | 6.1 | -2.8 | 5.7 | -3.9 | -1.7 | 7.0 |
| OH | -3.3 | 2.3 | 7.1 | -2.9 | -12.0 | 10.1 | -0.8 | 0.2 | 2.4 | 7.4 |
| TX | -1.2 | 0.2 | 3.1 | 5.4 | 1.7 | -0.4 | 1.2 | 0.4 | -0.6 | -4.5 |
| VA | -1.5 | 15.0 | -1.8 | 18.9 | -17.7 | 0.3 | -6.9 | 5.7 | -0.8 | -8.5 |
| WA | 9.3 | -8.2 | 5.7 | 16.7 | -14.5 | 29.4 | 1.3 | 3.4 | -4.3 | 10.8 |

*Within Birth State SE: 0.06248* *Within Residence State SE: 0.06390*

| Ages: 40 to 49 | | | | | Birth State | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Res. State | CA | IL | IN | MA | MI | NJ | NY | OH | PA | TX |
| AZ | -10.0 | -2.8 | 13.5 | 16.7 | 4.5 | -20.1 | 6.5 | -6.2 | 2.9 | -6.3 |
| CA | -5.5 | -7.4 | -5.7 | -16.1 | -12.0 | -1.3 | 5.4 | -1.6 | 3.6 | -10.4 |
| CO | 4.4 | 5.1 | -18.7 | -17.4 | -5.8 | 21.3 | 1.7 | -0.8 | -10.3 | -26.2 |
| FL | -8.6 | 1.3 | -1.6 | -5.2 | 3.8 | 10.1 | 2.9 | -0.9 | 3.2 | 3.0 |
| GA | -1.5 | 19.2 | -16.9 | -15.0 | 3.0 | -6.4 | -4.4 | -6.0 | -7.3 | -15.1 |
| NC | -6.5 | -4.2 | 27.0 | -20.1 | -6.9 | 6.1 | -0.2 | 4.9 | 9.6 | -1.7 |
| OH | 3.8 | -1.0 | -4.2 | 2.0 | 11.6 | -8.3 | -6.6 | 4.4 | 5.9 | 4.5 |
| TX | -16.8 | -6.6 | -8.0 | -13.4 | 5.9 | -16.0 | 4.1 | -9.4 | 15.4 | -8.4 |
| VA | -0.3 | 6.4 | -7.1 | 28.4 | 12.3 | -6.2 | -7.9 | -5.4 | 5.9 | 10.6 |
| WA | -4.8 | -0.5 | -11.7 | -25.7 | -1.0 | 18.2 | -6.2 | 16.5 | 13.2 | -20.8 |

*Within Birth State SE: 0.07300* *Within Residence State SE: 0.07869*

| Ages: 50 to 59 | | | | | Birth State | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Res. State | CA | IL | IN | MA | MI | NJ | NY | OH | PA | TX |
| AZ | -11.6 | -7.9 | -10.2 | -3.1 | 3.1 | -35.5 | 10.7 | -6.8 | 2.7 | 12.9 |
| CA | 0.6 | 6.7 | -20.8 | 3.5 | 4.5 | 9.7 | 0.5 | -1.4 | -6.3 | -4.7 |
| CO | -19.2 | 17.2 | -9.7 | -13.1 | 8.0 | 37.5 | -11.0 | 12.7 | 19.6 | -4.5 |
| FL | -13.9 | 0.6 | 8.4 | 10.3 | 10.1 | 5.3 | 1.4 | -2.6 | 10.7 | 20.1 |
| GA | -3.5 | -2.0 | -0.5 | -1.6 | 10.8 | -16.3 | -3.3 | -12.5 | 9.3 | -7.0 |
| NC | -5.4 | 4.9 | 1.0 | 20.6 | -17.8 | -27.4 | -8.7 | -8.9 | 13.0 | 2.7 |
| OH | -4.9 | 13.1 | 7.4 | 8.5 | 6.3 | -30.5 | -2.6 | -3.3 | -2.0 | 10.0 |
| TX | -2.9 | -10.0 | 4.7 | -23.5 | -11.3 | 24.2 | 6.8 | 4.1 | 1.2 | -2.0 |
| VA | 30.5 | -19.3 | -11.1 | 0.1 | -20.1 | -5.3 | -7.2 | 8.6 | -4.3 | 14.4 |
| WA | 1.2 | -10.8 | 2.7 | -29.8 | -4.2 | -9.1 | 7.6 | 19.7 | -3.1 | 1.5 |

*Within Birth State SE: 0.08638* *Within Residence State SE: 0.08647*

**Table 5.1 (Continued)**

***College Graduates***

**Ages: 30 to 39** — **Birth State**

| Res. State | CA | IL | IN | MA | MI | NJ | NY | OH | PA | TX |
|---|---|---|---|---|---|---|---|---|---|---|
| **AZ** | 0.0 | 3.8 | 11.4 | 6.9 | 14.7 | 13.6 | -0.1 | 13.4 | -3.3 | 24.8 |
| **CA** | -10.8 | -0.1 | -7.9 | 3.1 | 0.4 | 2.4 | 3.3 | -3.0 | 2.3 | -9.6 |
| **CO** | -3.2 | 13.3 | 11.6 | 9.3 | -1.2 | 7.5 | -5.6 | 17.9 | -1.2 | 6.4 |
| **FL** | -7.9 | 8.2 | -3.4 | -5.8 | -5.1 | 2.1 | 0.0 | -7.5 | -2.1 | -6.3 |
| **GA** | -6.9 | -2.5 | 3.8 | -7.1 | -2.6 | 2.5 | -2.1 | 2.9 | 10.3 | -1.1 |
| **NC** | -3.2 | -5.3 | 5.5 | 7.0 | -7.7 | 1.2 | 1.6 | 5.7 | 4.6 | -3.3 |
| **OH** | -15.9 | 9.1 | -5.0 | -14.6 | 6.8 | 16.3 | 3.4 | -10.7 | -2.9 | -10.8 |
| **TX** | -3.4 | -4.1 | 1.9 | 12.7 | -1.5 | -6.8 | -1.1 | 13.6 | 7.9 | -16.9 |
| **VA** | 6.0 | -15.3 | 2.1 | 2.0 | -8.5 | 5.9 | 4.2 | 13.0 | 3.4 | -11.5 |
| **WA** | -5.5 | -12.8 | 29.7 | -8.4 | 3.3 | 1.2 | 11.4 | -10.8 | -2.2 | 5.6 |

*Within Birth State SE: 0.06510* *Within Residence State SE: 0.07399*

**Ages: 40 to 49** — **Birth State**

| Res. State | CA | IL | IN | MA | MI | NJ | NY | OH | PA | TX |
|---|---|---|---|---|---|---|---|---|---|---|
| **AZ** | 13.9 | 5.5 | -8.0 | -5.6 | 4.1 | 13.0 | -12.5 | 14.5 | 5.0 | 2.9 |
| **CA** | -8.3 | -2.1 | 1.2 | 2.0 | -2.1 | -1.4 | 1.9 | -1.5 | 4.8 | 8.8 |
| **CO** | -2.3 | -5.7 | -26.0 | -9.0 | -6.7 | -3.5 | 5.9 | -9.0 | -2.4 | 11.6 |
| **FL** | 9.4 | -11.4 | -0.2 | -5.0 | 3.7 | 5.6 | 3.3 | -3.0 | -1.1 | 5.4 |
| **GA** | -12.2 | 3.7 | -5.8 | 0.3 | -15.6 | -3.4 | -1.5 | 2.6 | 7.3 | -9.1 |
| **NC** | -13.9 | 4.2 | 8.9 | -15.3 | 7.5 | -15.1 | -2.7 | -12.2 | 9.8 | -11.1 |
| **OH** | -5.4 | 9.4 | -25.0 | -12.7 | 4.2 | 10.9 | -3.9 | -19.4 | 1.1 | -2.1 |
| **TX** | -5.9 | -1.9 | -6.9 | 9.6 | -14.4 | 3.6 | 7.1 | -1.3 | -5.3 | -21.1 |
| **VA** | 7.9 | 0.6 | -31.6 | 10.8 | 16.1 | 12.4 | -3.7 | 7.9 | -0.2 | -8.5 |
| **WA** | -2.3 | 13.4 | 17.5 | 7.5 | 1.0 | 8.8 | -1.8 | -7.5 | -8.0 | -1.0 |

*Within Birth State SE: 0.07838* *Within Residence State SE: 0.08296*

***Ages: 50 to 59*** — **Birth State**

| Res. State | CA | IL | IN | MA | MI | NJ | NY | OH | PA | TX |
|---|---|---|---|---|---|---|---|---|---|---|
| **AZ** | 4.6 | 7.2 | 5.2 | -17.7 | 10.6 | 12.7 | -2.7 | 3.6 | -5.6 | -6.5 |
| **CA** | -7.8 | 2.5 | -4.5 | 0.9 | -4.4 | 5.4 | 2.2 | 5.2 | -0.4 | 7.1 |
| **CO** | -9.7 | -6.0 | -3.3 | -8.7 | -7.3 | -9.8 | -1.6 | 2.0 | 14.6 | 2.6 |
| **FL** | 6.0 | -11.1 | 6.6 | -16.5 | -4.3 | 1.0 | 6.9 | -7.8 | -4.6 | -19.9 |
| **GA** | -7.3 | 13.3 | -19.9 | -6.2 | -4.0 | -1.0 | -1.4 | 9.7 | -5.0 | 16.5 |
| **NC** | 15.6 | 14.0 | 18.4 | -0.9 | -10.7 | -5.2 | -11.7 | 19.5 | -5.2 | -16.3 |
| **OH** | -5.2 | -0.1 | -22.1 | -13.2 | -19.6 | -6.6 | 10.2 | -22.0 | -12.3 | -14.9 |
| **TX** | 0.2 | -2.2 | 8.4 | -9.0 | -6.5 | -16.7 | -5.6 | -5.9 | 10.6 | -25.9 |
| **VA** | 3.5 | 7.5 | 0.7 | -0.1 | 4.1 | -5.3 | 1.1 | -8.2 | -11.0 | -1.8 |
| **WA** | -10.8 | -1.2 | 33.9 | 7.8 | -15.9 | 4.1 | -3.1 | -12.8 | 13.3 | -37.1 |

*Within Birth State SE: 0.08559* *Within Residence State SE: 0.09115*

SOURCE.-2000 Census of Population and Housing, PUMS 5-Percent Sample.

NOTE.-The sample includes white, non-Hispanic men ages 30 to 59 that are born in the U.S., not enrolled in school and not living in group quarters. I exclude those with no earnings or no weeks worked last year as well as those with allocated values for any of the sample inclusion criteria or the variables used to make the comparisons. The ten residence states reported are the largest recipients of sample members born in another state. The ten birth states are those with the largest number of sample members living outside their birth state. Average log weekly wages for each birth state are estimated separately within each schooling level, birth cohort, and residence state and are adjusted via regression to the mid-point of the age interval. Sample members from all birth states are included in each regression, but I only present the ten birth states listed due to space considerations. Weekly earnings are top coded at the 98th percentile and bottom coded at the 2nd percentile.

To simplify the comparisons, (log) wages are measured relative to the earnings of a reference group consisting of a fixed weight average for workers from the four states with the largest number of men in the sample living outside of their birth state. These states are California, New York, Pennsylvania, and Illinois and the fixed weights are the state's fraction of workers living in another state among migrants from these four states. The weights are calculated separately by schooling level. A reference wage is specific to a birth cohort, a schooling level, and a residence state.[37] Results are only presented for a select group of states due to space limitations. The ten residence states are those with the largest number of men in the sample who were born in another state (in-migrants). The ten birth states are those with the largest number of men in the sample living outside their birth state (out-migrants).

The purpose of Table 5.1 is to see whether we can comfortably infer school quality by examining wage rankings within residence state. Ideally, if the controls work, then aside from sampling error, if one reads down a column to hold birth state constant for any cohort and schooling level panel, but across residence states (the rows), the numbers should remain constant. Because they are log wages relative to a log reference wage, they are interpreted as the percentage difference in wages from the reference group for those from a given birth state. To see where this takes us, let's compare the wages of high school graduates who were born in Michigan to those born in New Jersey among men ages 30-39 (the first panel). For Colorado residents, those from Michigan earn 20.4 percent more than the reference group and those from New Jersey earn 8.0 percent less. Clearly Michigan schools are superior to New Jersey schools. But, if we instead evaluate schools based on the earnings of workers living in Washington, those

---

[37] Among high school graduates from these four states who are living outside of their birth state, for example, the fixed weights are: California (23 percent), Illinois (19 percent), New York (33 percent), and Pennsylvania (25 percent). Within each residence state, the reference wage is computed as the sum over these four states of the fraction for each state times the logarithm of the average weekly wage of those born in that state who are living in the residence state. Among Arizona residents ages 30 to 39, those from California earn 6.34 (the log weekly wage), those from Illinois earn 6.55, those from New York earn 6.51, and those from Pennsylvania earn 6.45. The reference wage is (.23*6.34)+(.19*6.55)+(.33*6.51)+(.25*6.45) = 6.42. In this example, the numbers are rounded to show how the calculation is done. In computing the numbers in Table 1, the components are not rounded.

from New Jersey earn 29.4 percent *more* than the reference group and those from Michigan earn 14.5 percent *less*. Now, which state has the higher quality schools?

For high school graduates ages 30 to 39 born in these two states, Michigan products earn more in Arizona, Colorado, Florida, Georgia, North Carolina, and Texas; New Jersey products earn more in California, Ohio, Virginia, and Washington. Michigan dominates in six of ten comparisons, are Michigan schools better than New Jersey schools? Next examine the first panel of the second page of Table 5.1 for college graduates ages 30 to 39 born in the same two states. Michigan products earn more in Arizona, Texas, and Washington; New Jersey products earn more in California, Colorado, Florida, Georgia, North Carolina, Ohio, and Texas. Six of the states flipped outcomes, i.e., one state dominates for high school graduates and the other state dominates for college graduates. Which state has higher quality schools? Would any sane person seriously try to answer this question from this data?

The example discussed here, involving Michigan and New Jersey, is solely for illustrative purposes. Many other comparisons showing the same type of reversals could have been discussed instead, using any of the birth cohorts and either schooling level. A more difficult task would be to find a comparison where one birth state dominates another in all comparisons for all cohorts and both schooling levels; there are none. What about for any cohort and schooling level? There is exactly one, Massachusetts beats California in every comparison for the youngest cohort of high school graduates.

From just a casual observation of the wage comparisons in Table 5.1 it appears that there may be as much wage variation for those born in the same state and living in different states (reading down the columns) than for those living in a common residence state from different birth states (reading across the columns). At the bottom of each panel, I present within birth state (or between residence states) and within residence state (or between birth states) standard errors for each schooling level and birth cohort. These standard errors are estimated using the relative wages presented in the table weighted by their variances.[38] For this subset of residence and birth states, it is not quite the case that there is more variation within birth state than residence state, but it is close.[39]

Another comparison that can be made from the calculations in Table 5.1 is between the relative wages of high school and college graduates paired on the residence state, birth state, and birth cohort. If California-born high school products ages 40 to 49 do well in Arizona, then so should California-born college products of the same age. After all, they were presumably educated in the same place at the same time; if quality is higher for one, then it must be higher for the other given the implicit assumptions in the literature that quality can be inferred from elementary and high school quality alone. The paired observations, plotted in Figure 5.1, should have a strong, positive relationship. The line is drawn for reference purposes and shows where the points would lie if high school and college relative wages were equal. Upon visual inspection, it is obvious that the correlation between the wages of high school and college graduates presumably trained in the same state at the same time and working in the same market are weak; the unconditional correlation is actually negative (–0.0789).

---

[38] Variances of these ratios are estimates using the delta method. Specifically,

$$Var(\frac{\bar{X}}{\bar{Y}}) = (\frac{\mu_X^2}{\mu_Y^2})(\frac{\sigma_X^2}{\mu_X^2} + \frac{\sigma_Y^2}{\mu_Y^2} - 2\frac{\sigma_{XY}}{\mu_X \mu_Y})$$

[39] Variances computed using all residences states and not just these ten show a similar pattern; within residence state variances are higher than within birth states variances, but they are close.

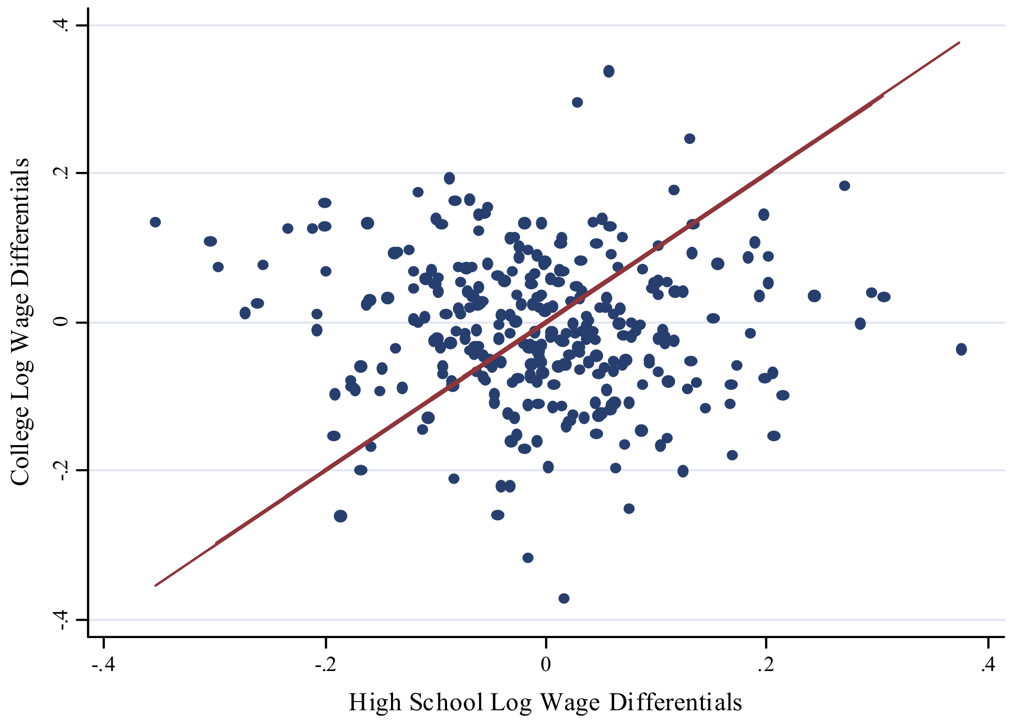


FIG. 5.1.—The wages of high school and college graduates (relative to a school level and cohort specific reference wage within residence state) for men from a given birth state.

The lack of consistency in Table 5.1 for men of the same schooling level suggests that those migrating from a given birth state to different markets possess different skill sets, i.e., that migration is selective in a way that undermines the ability of these types of analyses to inform about school quality. The lack of correlation in Figure 5.1 either reinforces this observation or highlights the importance of college quality in determining wages for college graduates coupled with the lack of correlation between college and high school quality documented in Chapter IV. In Heckman, et al.'s (1996) criticism of the Card/Krueger study, they discuss the problem of selective migration and, in an admittedly oversimplified attempt to correct for this, allow the impact of school characteristics to vary by residence division and include terms for the distance (and its square) between division of birth and division of residence.[40] They find that the effects of the pupil-teacher ratio vary by division, but could not reject a common teacher salary effect. Moreover, the distance terms, thought to capture the cost of migration (both actual and psychic), are both statistically significant. Card and Krueger, treating this modification in the specification as having corrected the selective migration problem, responded by noting that the average of the pupil-teacher effects over divisions is roughly equal to their pooled estimates using the 1970 and 1980 censuses, but is nearly three times larger than their estimate using the 1990 census. They conclude that selective migration might "lead to downward biases in the estimated effect of school quality." (Card and Krueger, 1996)

Unfortunately, the simple correction procedure implemented by Heckman, et al. cannot be implemented for the relative wage comparisons in Table 5.1. The average age-adjusted wages of men from each birth state are computed within residence state for each cohort and schooling level. The distance between division of birth and division of residence is a constant for all men born in a given state and living in a given state. The distance terms and the average wage within a birth state are not uniquely identified. Using distance calculated from the birth state to the residence state, of course, does not solve the problem.[41] Extending the correction beyond that suggested by Heckman, et al., one might want to control for other factors such as a worker's location within the

---

[40] The Card/Krueger (1992a) study estimated the impact of school quality on earnings for white men born between 1920 and 1949 in two stages using 1980 Census data. In the first stage, (log) weekly wages were regressed on state-of-residence and state-of-birth indicators, whether living in an SMSA, marital status, experience, and years of schooling interacted with state-of-birth and division-of-residence indicators. This is done separately for three ten-year birth cohorts, producing 147 state-of-birth by cohort gradients. In the second stage, these gradients are regressed on statewide average school characteristics for the birth state for each cohort, cohort indicators, and (in their preferred specification) state-of-birth indicators. The school characteristics are the pupil-teacher ratio, teacher's salaries, and term lengths. The effects of school characteristics are constrained to be constant across all residence divisions. The Heckman, et al. (1996) revisions allow the birth state gradients to vary by cohort and by division of residence yielding 1,323 estimates. The implicit assumption is that the census division is the market.

[41] The smallest geographic area identified in the PUMS census files for birth location is the state and the finest geographic area identified for current residence is either the metropolitan area (if living in a large metro area) or an aggregation of counties (if living in sparsely populated counties). If the variation within residence state were used to identify the distance effects, we would have to wonder what is being measured.

schooling distribution in their birth state, but all measures that are constant within birth state and schooling level (and cohort) cannot be uniquely identified along with the birth state effects and cannot be used to correct for selective migration. Heckman, et al.'s procedure is the only published correction for selectivity in assessing the impact of school quality on earnings, so that effectively none of the studies in the literature have overcome (or even addressed) this problem.

*But Have You Looked at Occupation?*

Because everyone's first impulse is to suggest that occupation is very important whenever I discuss the issue of migration, I will begin there before turning to other migration topics. The intuition is simple and reasonable. If workers who move from Texas to California are different from those who move from Texas to New York, the occupational choice of workers who migrate to different places might capture how they are different. Those moving to New York are more likely to be stockbrokers and those moving to California are more likely to be waiters (although sometimes known as actors).

To examine this contention, I grouped occupations at the broadest level possible, white collar and blue collar.[42] The comparisons presented in Table 5.2 are restricted to high school graduates in the youngest cohort from Table 5.1. For workers born in California who move to other states, do blue collar workers tend to go in one direction while white collar workers go in the other? Those moving to Florida (78.3 percent) and Ohio (75.0 percent) are more likely to be blue collar while those moving to Georgia (52.5 percent) and Virginia (58.8 percent) are less likely to be so. For those from

---

[42] White-collar workers include those in Management, Professional, and Related Occupations (codes 001-359), Service workers in Healthcare Support Occupations (360 – 369), and Sales and Office Occupations (470-599). Blue-collar workers include all other occupations: Service workers in Protective Service Occupations, Food Preparation and Serving Related Occupations, Building and Ground Cleaning and Maintenance Occupations, Personal Care and Service Occupations (370-469); Farming, Fishing, and Forestry (600-619); Construction, Extraction, and Maintenance (620-769); and Production, Transportation and Material Moving (770-979).

California, the difference in the proportions that are blue collar across destination states is statistically significant at the 0.05 level. The proportions are also statistically significantly different for men from Pennsylvania and Texas, but this is not true for men from the remaining seven states. Results for the other birth cohorts and for college graduates also reveal mixed results. Of the sixty comparisons (10 destinations, 3 cohorts, 2 schooling levels), fifteen are statistically significant at the 0.05 level. These results suggest that occupation might help in solving the selectivity problem, but it is not the magical elixir that will completely cure the symptoms of selective migration.

**Table 5.2**
**Percent of Male High School Graduates Ages 30 to 59 Within a Residence State Who Are Working in a Blue Collar Job Among Those from a Given Birth State**

| | Birth State | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Residence State** | **CA** | **IL** | **IN** | **MA** | **MI** | **NJ** | **NY** | **OH** | **PA** | **TX** |
| **AZ** | 65.70 | 60.56 | 67.74 | 66.67 | 67.86 | 51.72 | 62.50 | 67.69 | 74.29 | 73.53 |
| **CA** | NA | 57.98 | 58.14 | 68.00 | 61.54 | 57.14 | 61.27 | 61.63 | 55.36 | 48.65 |
| **CO** | 68.22 | 73.58 | 64.29 | 64.29 | 56.25 | 83.33 | 68.09 | 64.29 | 70.83 | 69.23 |
| **FL** | 78.26 | 68.26 | 74.36 | 68.53 | 69.76 | 64.98 | 65.20 | 67.97 | 67.87 | 68.57 |
| **GA** | 52.54 | 70.45 | 75.00 | 70.00 | 68.42 | 65.22 | 64.38 | 72.50 | 56.41 | 75.61 |
| **NC** | 66.67 | 48.00 | 72.22 | 80.00 | 78.38 | 65.22 | 68.14 | 72.84 | 70.59 | 70.97 |
| **OH** | 75.00 | 62.79 | 74.47 | 60.00 | 80.00 | 52.63 | 67.39 | NA | 79.03 | 75.00 |
| **TX** | 66.05 | 61.73 | 68.09 | 52.38 | 73.13 | 67.86 | 53.33 | 68.18 | 63.08 | NA |
| **VA** | 58.82 | 63.64 | 57.89 | 56.52 | 65.71 | 65.96 | 66.67 | 65.08 | 61.39 | 79.17 |
| **WA** | 68.95 | 63.64 | 81.82 | 90.91 | 78.26 | 64.29 | 72.22 | 66.67 | 64.00 | 65.62 |
| **Total** | 67.75 | 63.83 | 70.62 | 67.80 | 70.03 | 63.88 | 64.59 | 67.84 | 67.41 | 67.34 |
| **P-Value** | 0.036 | 0.399 | 0.569 | 0.485 | 0.160 | 0.610 | 0.623 | 0.887 | 0.048 | 0.035 |

SOURCE.-2000 Census of Population and Housing, PUMS 5-Percent Sample.

NOTE.-See the note to Table 5.1 for the sample description. The P-Values are taken from Pearson Chi-Squared tests calculated for each birth state testing the null hypothesis that the proportion of workers who are blue collar is constant across residence states. Only movers are included in the calculation.

## A Preliminary Investigation of Interstate Migration

The economic theory behind models of interstate migration in its simplest terms is as follows: (1) an individual is faced with various alternatives, each with a set of attributes that provide varying levels of utility to the individual (or family); (2) the individual selects the choice that maximizes the present value of expected utility (V)

$$\max_i \{V_1(A_1|X), V_2(A_2|X), \ldots V_i(A_i|X), \ldots , V_n(A_n|X)\}$$

where n is the number of alternatives, $A_i$ is a set of alternative specific attributes that affect utility and X is the individual's (or family's) set of characteristics. The $A_i$'s reflect the costs and benefits of living in a particular area and the X's reflect both differences in the costs and benefits to particular groups as well as differences in how those costs and benefits are evaluated. For example, an alternative where education is highly valued relative to other possible alternatives is expected to provide different incentives to college graduates than to high school dropouts. Similarly, the health benefits of the dry Arizona desert air will provide different incentives for those with asthma or chronic bronchitis. Migration decisions depend upon both alternative specific attributes and individual characteristics.

*Why Do People Move from State to State?*

Before examining any factors that might affect migration decisions, it might be informative to see why people say they move. Beginning in 1998, respondents to the March Supplement of the CPS who are at least one year old and who lived at a different residence on March 1st of the preceding year were asked, "What was [your] main reason for moving?" The data also include the state of residence on March 1st of the preceding year so that it is possible to identify those that moved to another state.

**Table 5.3**
**Mobility Propensities Reported by Schooling Level**

| Schooling Level | Not Moving (Pct) | Intrastate Move (Pct) | Interstate Move (Pct) | Observations |
|---|---|---|---|---|
| Did Not Complete High School | 77.8 | 19.8 | 2.4 | 10,240 |
| High School Degree | 82.6 | 15.1 | 2.3 | 43,985 |
| Some College | 82.4 | 14.9 | 2.7 | 34,546 |
| College Degree | 81.5 | 14.3 | 4.1 | 26,768 |
| Graduate Degree | 83.5 | 11.2 | 5.3 | 10,176 |
| Total | 82.0 | 14.9 | 3.0 | 125,715 |

SOURCE.-March Supplement to the Current Population Survey, 1998-2005.

NOTE.- The sample includes white, non-Hispanic men ages 18 to 45 who were living in the United States on March 1st of last year, not currently enrolled in school or in the armed forces. Those with allocated information for sex, race, age, Spanish ethnicity, enrollment status, educational attainment, whether moved, and main reason for the move are excluded from these calculations. Information on the main reason for moving is available from 1998 to 2005 (the last year of available data). Any changed in residence state is considered a move.

Table 5.3 presents the percent of white, non-Hispanic men ages 18 to 45 not enrolled in school or in the military that moved within state or to another state by schooling level.[43] Not surprisingly, the percent moving to another state within the last year increases with schooling. This is not true, however, of intrastate moves and this finding may reflect lower costs of a local move for those with less education. For example, 38 percent of the sample included in Table 5.3 with a high school degree reported living in rental housing as compared to 22 percent of those with a college degree. Moreover, the relationship between schooling level and the percent renting is perfectly ranked.[44] The difference in the percent living in rental housing across schooling levels is at least suggestive of the reason that intrastate moves are higher for those with less schooling, but this is not the focus of this study and will not be explored further here.

---

[43] The age restrictions are meant to isolate men with higher mobility rates.
[44] For this sample, the percentage of men renting is 47 percent of those with less than a high school degree, 38 percent of high school graduates, 30 percent of those with some college training, 22 percent of college graduates and 16 percent of those with graduate training.

**Table 5.4**
**The Main Reason Given For Moving by Interstate Migrants Reported by Schooling Level**

| Main Reason for Moving | Did Not Complete High School (Pct) | High School Degree (Pct) | Some College (Pct) | College Degree (Pct) | Graduate Degree (Pct) |
|---|---|---|---|---|---|
| | **Schooling Level** | | | | |
| Change in Marital Status | 5.2 | 4.7 | 3.6 | 2.7 | 1.5 |
| To Establish Own HH | 8.0 | 3.7 | 2.5 | 2.4 | 1.3 |
| Other Family Reason | 23.0 | 19.6 | 13.6 | 8.1 | 4.6 |
| All Family Reasons | 36.3 | 28.0 | 19.7 | 13.1 | 7.4 |
| New Job Or Job Transfer | 15.7 | 30.8 | 38.1 | 56.6 | 71.0 |
| To Be Closer to Work | 2.0 | 1.7 | 1.4 | 1.7 | 2.2 |
| To Look For Work Or Lost Job | 12.9 | 8.2 | 7.6 | 4.4 | 1.7 |
| Other Job Related Reason | 5.2 | 5.6 | 5.4 | 3.7 | 3.9 |
| All Job Reasons | 35.9 | 46.3 | 52.4 | 66.5 | 78.7 |
| Wanted Better Neighborhood/Less Crime | 3.8 | 2.3 | 1.8 | 0.5 | 0.6 |
| Wanted Cheaper Housing | 4.0 | 2.8 | 1.4 | 0.6 | 0.6 |
| Wanted New/Better House/Apartment | 2.4 | 2.0 | 1.3 | 1.5 | 1.7 |
| Wanted Own Home, Not Rent | 1.2 | 1.4 | 1.7 | 2.1 | 1.5 |
| Other Housing Reason | 4.8 | 2.5 | 2.3 | 1.0 | 0.2 |
| All Housing Reasons | 15.3 | 11.0 | 8.5 | 5.7 | 4.4 |
| Change Of Climate | 4.8 | 3.5 | 3.6 | 2.2 | 0.6 |
| Health Reasons | 2.8 | 1.5 | 1.2 | 0.4 | 0.2 |
| Retired | 0.4 | 0.7 | 0.9 | 0.5 | 0.2 |
| To Attend Or Leave College | 0.4 | 3.2 | 8.3 | 8.8 | 6.5 |
| Other Reasons | 4.0 | 5.9 | 5.5 | 2.8 | 2.0 |
| All Other Reasons | 12.5 | 14.8 | 19.5 | 14.7 | 9.4 |
| Observations | 248 | 1,001 | 935 | 1,104 | 541 |

SOURCE.-March Supplement to the Current Population Survey, 1998-2005.

NOTE.-The sample is restricted to white, non-Hispanic, males, ages 18-45, living in the United States last year who are not in the military and not enrolled in school and who lived in a different state on March 1st of the preceding year. Observations with allocated values for the sample restriction parameters, the schooling variables, or the migration questions are excluded.

Table 5.4 documents the main reasons for moving given by interstate migrants. Results are presented for the same schooling categories as above and the percentages reported are column percentages. In addition, reasons have been categorized into four main groups – family reasons, job reasons, housing reasons, and other reasons – with sub-percentages reported at the bottom of the groupings. The motivation for moving varies substantially by schooling. High school dropouts are more likely to move for family reasons (36.3 percent) than any other whereas those with a graduate degree are much more likely to move for job reasons (78.3 percent). Moreover, within the job-related reasons category, the detailed reasons also vary considerably by schooling category; nearly 40 percent of high school dropouts moved to look for a job and nearly all with graduate training moved for a new job or job transfer. This may suggest that high school dropouts are more likely to move to find a job and those with higher education are more likely to find a job and then move.[45] In addition, high school dropouts are nearly four times as likely to move for housing related reasons as men with graduate degrees. The tabulations in Tables 5.3 and 5.4 convey that not only does the propensity to move vary by schooling level, but so do the reasons for the move and models that attempt to measure the determinants of interstate migration should be sufficiently complex to capture these differences. It is also comforting to see that economic and family reasons play such a prominent role in the motivation to move since these factors are highlighted by many interstate migration studies.

---

[45] Alternatively, it may just reveal that these two groups answer questions differently.

*The Importance of Age and Years of Schooling*

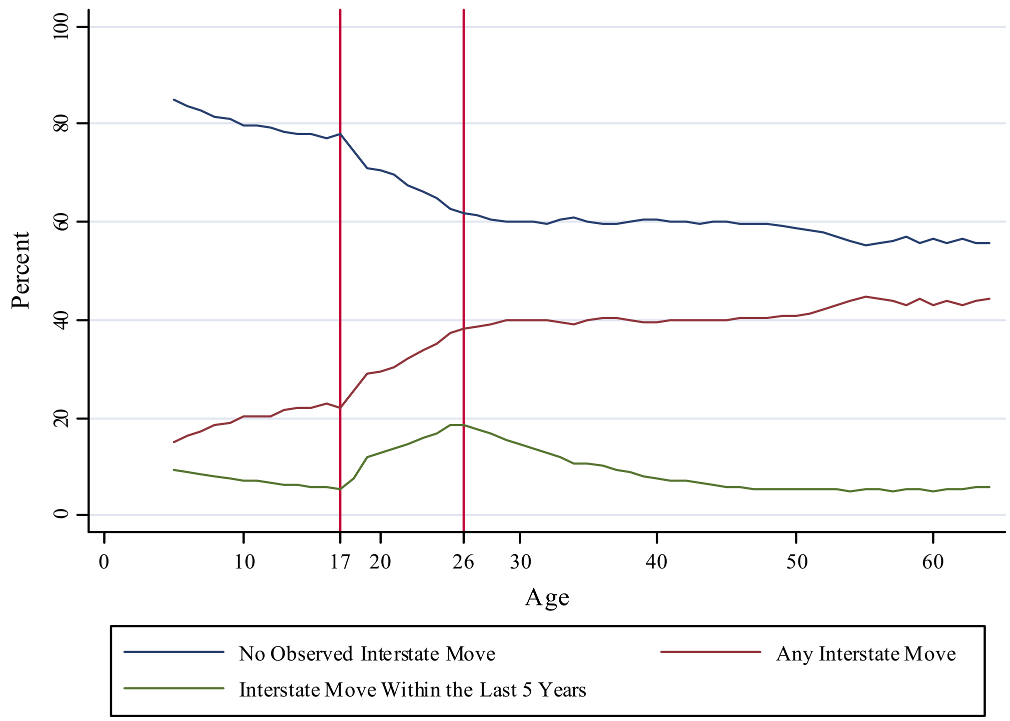


FIG. 5.2.—Percent of white, non-Hispanic men with an interstate move by age.

I begin this sub-section by examining migration propensities from the 2000 U.S. Census by single year of age for white, non-Hispanic men ages 5 to 64.[46] Figure 5.2 graphs the percent of men with an interstate move within the last five years, the percent with any interstate move (as captured by state-of-birth and interstate moves within the last five years), and the percent without an identifiable interstate move.[47,48] Given the

---

[46] Additionally, those included in these calculations had to be born in the U.S. and living in the U.S. in 1995 and could not have any allocated values for any of the variable used to define the sample or to compute the percentages in the tables and figures.

[47] I define an interstate move as a change in state of residence within the last five years, between April 1995 and April 2000, that is not within the same large metropolitan area (using the Census Bureau's Consolidated Metro Statistical Area, CMSA) and that involves a move of at least one hundred miles. A CMSA is loosely defined by the Census Bureau as a metropolitan area with at least one million inhabitants and with close economic and social ties. The prime-mover age migration sample includes 89,433 moves

data limitations, this last group may include too many men at the expense of those that moved since birth as those that moved from their birth state to another state and then returned to their birth state more than five years ago will show no evidence of a move.[49]

Focusing first on the percent with an interstate move within the last five years, it should not be surprising to see that at schooling completion ages, there are a larger percentage of interstate moves than at any other ages. A second finding, that reinforces the calculations presented in Chapter III, is the high percentage (22 percent) of men who have moved from their birth state by age seventeen and for whom birth state will not accurately capture the full pre-college academic career. Moreover, in studying interstate migration it is important to remember that these early decisions to move were probably not those of the individual and do not reflect any optimizing behavior on his part.

Before proceeding, I want to isolate two sub-groups included in Figure 5.2 that are very different in their migration behavior.[50] The first group is those in the military at the time the data was collected. Among this group, eighty-two percent of those with at most a high school degree and seventy-five percent of those with at least some college had an interstate move in the five years prior to the survey. The second group is those living in group quarters (after excluding those in the military). White, non-Hispanic men in institutionalized group quarters (those with formally authorized, supervised care or custody such as correctional institutions, nursing homes, and juvenile institutions) are

---

that cross state boundaries between 1995 and 2000, of which 10,018 (11.2%) are within the same large metro area (CMSA) or are of 100 miles or less. Given that the number of men in the data living in D.C. is small, the District of Columbia is combined with Virginia due to close economic ties with that state. Those living in Alaska or Hawaii in either 1995 or 2000 are also excluded. Distances are computed from the population centroids of the PUMA code of residence in 1995 and 2000 using data from Missouri Census Data Center (the "MABLE/Geocorr*2K*: Geographic Correspondence Engine with Census 2000 Geography") and the "Great Circle" formula, which calculates the shortest distance between two points along a sphere

[48] Figure 1 includes those enrolled in school or in the military or living in group quarters.

[49] The last group, however, may also be too low as it is not possible to exclude within large metro area moves and short interstate moves from the place of birth given the geographic information available in the census.

[50] All percentages reported in this paragraph exclude those enrolled in school, those with a Ph.D., those outside of the prime-mover ages defined below, and those with allocated values for enrollment, whether in the military, or education.

two to three times as likely to experience an interstate move as those in housing units and those in non-institutionalized groups quarters (college dormitories, military quarters, group homes, etc.) are three to four times as likely to move to another state.[51] Because these men probably did not move to capture economic rents, although their migration may suggest something selective, I exclude them from the calculations presented in the remainder of this chapter.

Focusing exclusively on interstate moves within the last five years and excluding those enrolled in school, those in the military, and those living in group quarters, Figure 5.3 extends the presentation of Figure 5.2.[52] First, combining all education categories into an undifferentiated whole masks large differences in the probability of moving across schooling levels. Reinforcing the results from the CPS, those with higher levels of education are more likely to move to another state within the last five years at all ages. There is very little interstate migration for those who dropped out of high school, never reaching 10 percent at any age. Those with Ph.D.'s, however, reach a peak at age thirty with nearly sixty percent moving across state lines within the past five years. The aggregation also masks that the peak migration ages vary by years of schooling and that they peak at roughly expected graduation ages, or at zero years of potential experience, and then decline steadily until age 64.[53]

---

[51] The range in values is because comparisons are made within the schooling categories high school degree or less, some college without a degree, college degree, and graduate degree.
[52] Those ages with few observations within schooling level are also excluded from Figure 2 due to the high variability in the percentages reported.
[53] The exception to the steady decline is those with a Ph.d. who experience a spike at age 34, the age around which many have tenure denied.

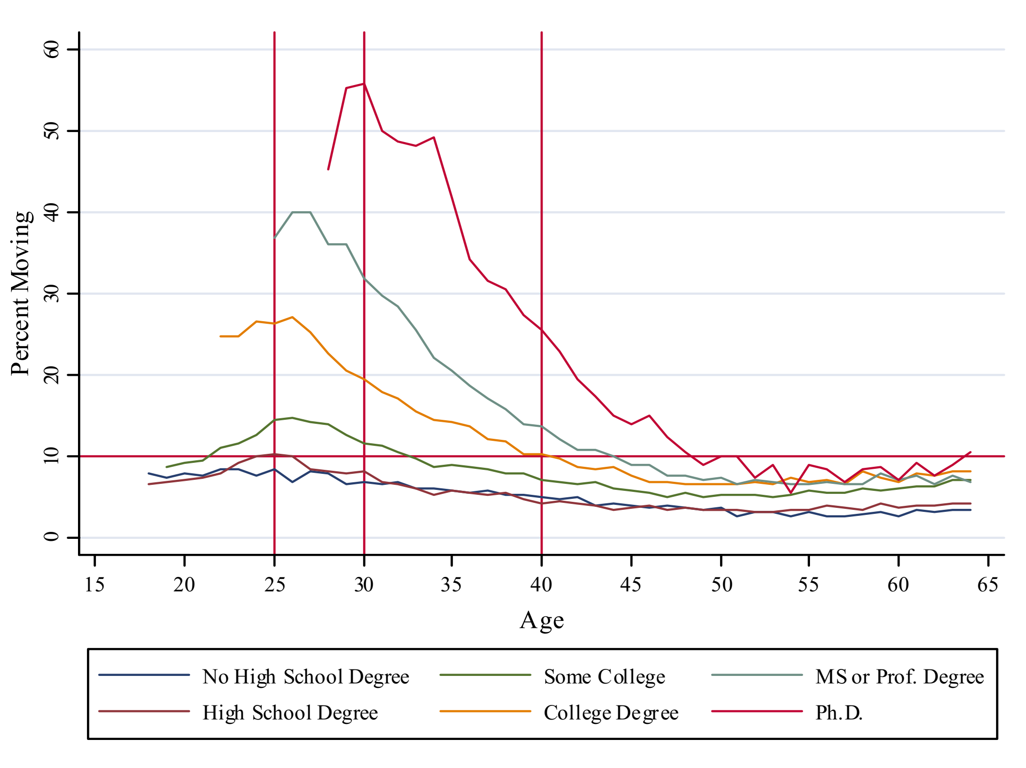


FIG. 5.3.—Percent of white, non-Hispanic men with an interstate move in the last five years by education and age.

In the figures below, I focus on those most likely to react to the economic incentives provided by the market. Rather arbitrarily, I selected the age at which the probability of an interstate move dips below ten percent for the first time at each level of schooling as the age cutoff point. The remaining charts in this section present data for high school graduates (with no college training) ages 18 to 26 and college graduates (with no graduate training) ages 22 to 40.

*Additional Observations Regarding Interstate Migration*

This final sub-section presents several descriptive figures that examine the migration patterns of white, non-Hispanic men between 1995 and 2000 for the prime-mover ages defined above. These calculations have important implications for analyses that would attempt to identify school quality effects and suggest that selective migration is a difficult hurdle for empirical work to clear.

I begin by examining gross and net migrant flows for high school graduates and for college graduates by 1995 residence state for forty-eight "states."[54] The boxes in Figures 5.4a and 5.4b show net migration gains or losses as a percent of the size of the 1995 labor market of white, non-Hispanic men in the state for workers with the same schooling level. The range plots with capped spikes, or "whiskers", present gross worker inflows to and outflows from the state. It is not necessarily surprising that the biggest gainers and losers are the smaller states, such as New Mexico, Wyoming, Colorado and Nevada, where small changes in the population can have a large impact on percentage changes. However, some of the larger states such as Florida, New York, California and Texas also appear in the extremes of the distribution.

There are some encouraging findings contained in these two figures. If we suppose, like those that adopt gradient models to measure school quality effects, that economic outcomes are captured by the returns to schooling, then we expect those with high levels of schooling to migrate to states with high returns to schooling and those with low levels of schooling to migrate to states with low returns to schooling. The five states with the lowest returns to education (college graduate earnings relative to high school graduate earnings) – Wyoming (23.7 percent), Vermont (29.6 percent), South Dakota (29.7 percent), Montana (31.2 percent), and North Dakota (32.0 percent) – are clustered at the low end of the college graduate graph; they are the largest losers of those with college degrees.[55] Unfortunately, this observation is tempered by their also being

---

[54] Alaska and Hawaii are omitted and the District of Columbia is combined with Virginia.
[55] The sample restrictions and model specification used to compute these schooling gradients are documented in Appendix C.

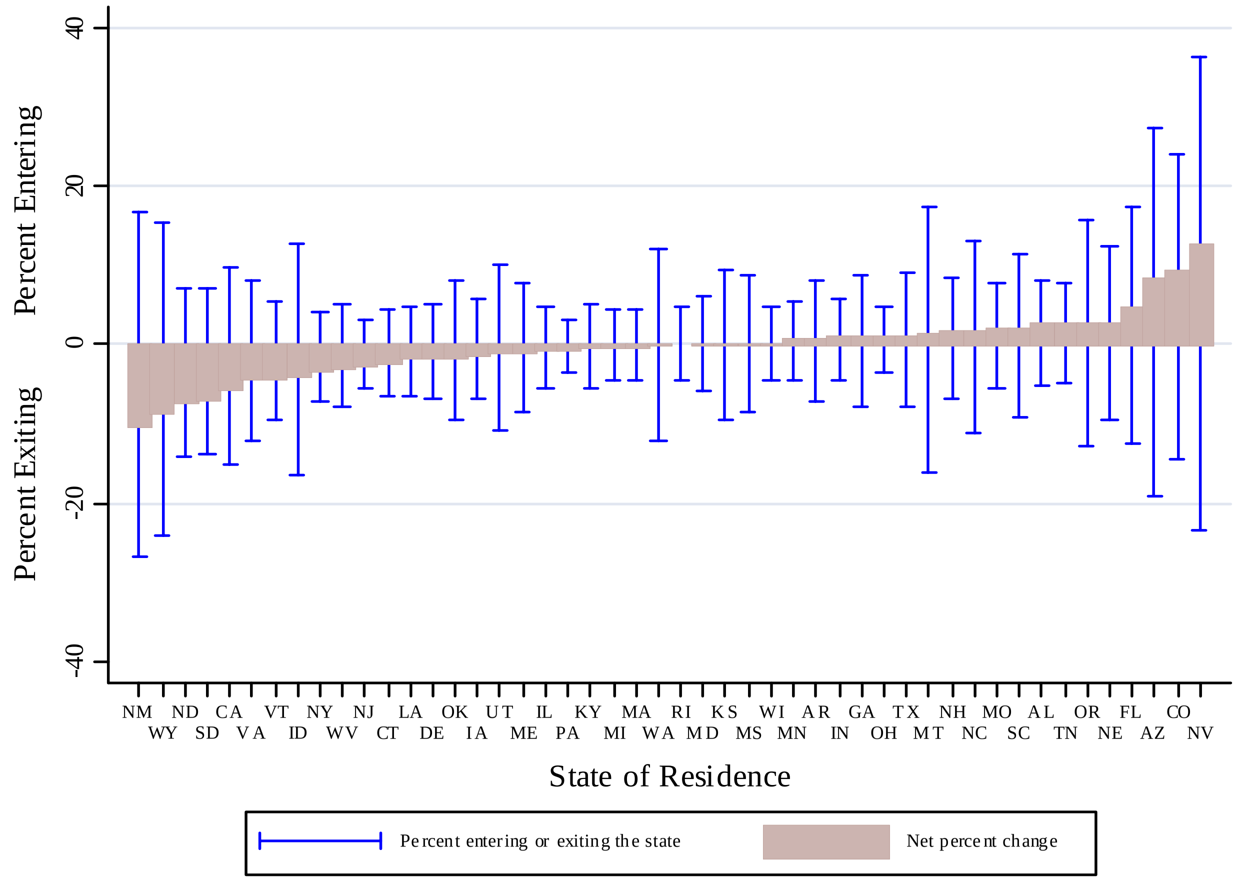


FIG. 5.4a.—Inflows, outflows, and the net migration of white, non-Hispanic prime-mover age men with a high school degree as a percent of the 1995 stock of those with a high school degree.

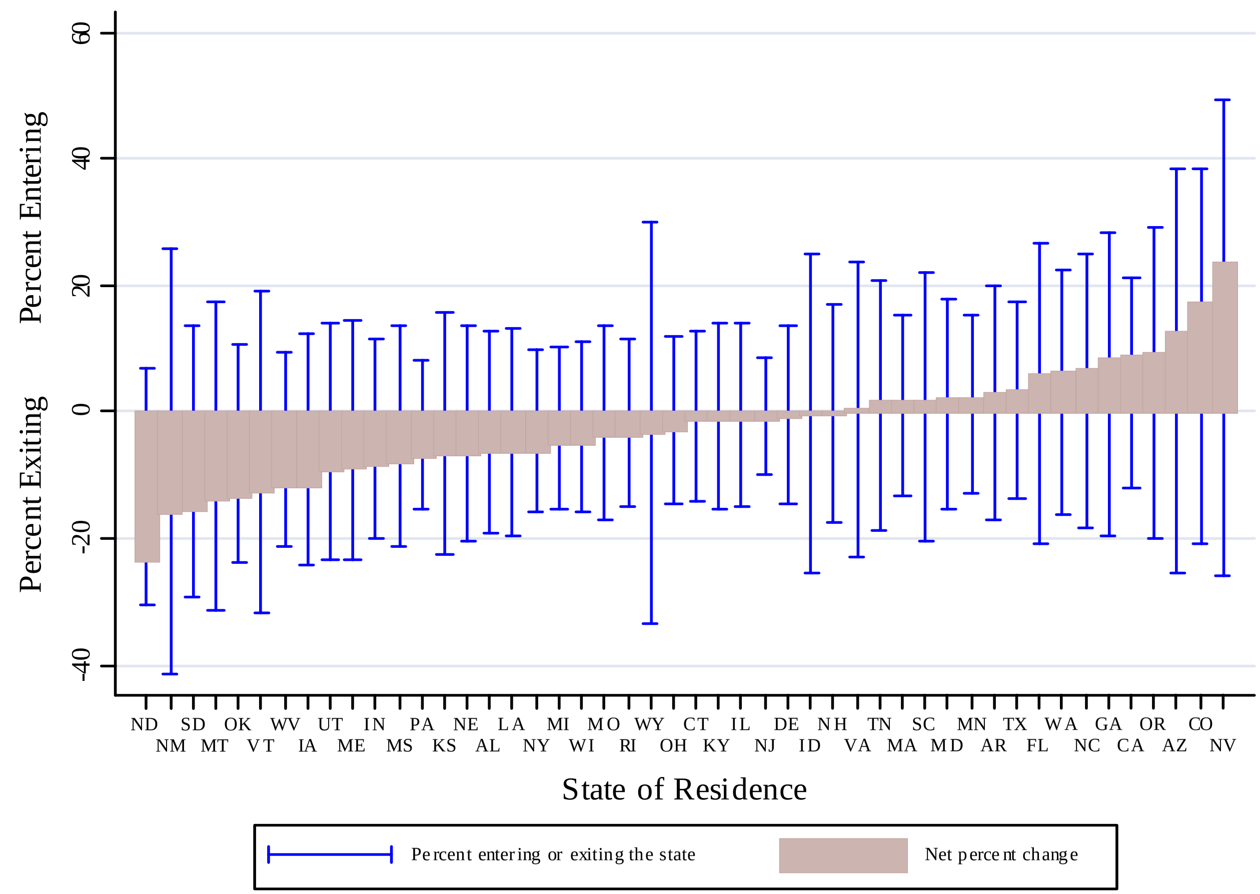


FIG. 5.4b.—Inflows, outflows and the net migration of white, non-Hispanic prime-mover age men with a college degree as a percent of the 1995 stock of college graduates.

the largest losers of high school graduates, Montana being the exception. This suggests either that gradients do not capture economic incentives (discussed below) or that amenity levels in these states may be driving lower educated workers away.

Turning attention to the states with the highest returns to education – Texas (56.0 percent), New York (57.2 percent), California (57.3 percent), Georgia (58.2 percent), and Virginia (59.9 percent) – we see slightly more promising results. In general, these states are among the largest losers of those with a high school degree and among the largest gainers of those with a college degree. The exceptions are Texas and Georgia, which attract both high school and college trained workers but relatively more college trained workers, and New York, which loses both high school and college trained workers but surprisingly relatively more college trained workers. This last result may reflect the large number of students attending school in New York before moving to another state to begin their work careers.

Countering any optimism is the large quantity of inflows and outflows relative to the net gains and loses. In California, for example, the college graduate population in the sample grew by 9.1 percent, but this is the net result of a loss of 11.9 percent of those living in California in 1995 and a gain of 21.0 percent that arrived from other states. The inflows and outflows swamp the net changes in almost every state. If migration is motivated primarily by economic reasons, this is surprising. There must be something else driving much of the observed interstate migration.

Although the gradients discussed above are convenient in that they are normalized for cost of living differences across geographic areas; it is real earnings, not returns, that provide incentives to migrate. If earnings are equal for workers in two states for all levels of schooling, then the returns to schooling must be the same. But, suppose that the returns to schooling are higher in Virginia than in California. It is possible that earnings are higher for all levels of schooling in California so that the incentives are for workers in Virginia to move to California until earnings equalize. It is also possible that earnings are higher for all levels of schooling in Virginia in which case the incentives go the other direction. Gradients don't tell us everything, and may tell us nothing, about economic incentives. Because the migration rates of college graduates are high relative to the rates for high school graduates, if earnings equalize for any group it should be those with a college degree. Suppose that the earnings of college graduates equalize, but those of high school graduates do not. Recall from Chapter II that wages are the product of education (years of schooling times quality) and the price of education. The state with the lowest school quality levels will produce high school graduates that earn the least. Since the gradient here is the college/high school differential and the earnings of college graduates are equal by assumption, the states with the lowest quality of schooling will be those with the highest gradients. This leaves a lot of doubt as to the meaning of results from analyses, such as Card and Krueger's, that rely on gradient models and interstate migration to identify school quality effects.

To further examine the notion that workers respond to differences in schooling gradients when making migration decisions, Figure 5.5 examines the correlation in migration patterns between high school and college graduates. If differences in schooling gradients capture incentives to move, then we expect higher educated workers to go in one direction, where returns are high, and low educated workers to go in another. To investigate this hypothesis, the percentages of high school and college interstate movers living in the same state in 1995 and moving to the same state in 2000 are plotted in this graph. For example, among those who moved to another state, 21 percent of high school graduates living in Nevada in 1995 moved to California by 2000 whereas 29 percent of college graduates made the same move. This point is easily identified on the graph. The line graphs the relationship if the two percentages were equal and is for reference only. The figure includes only the top ten recipients of either high school or college workers from a given state. This restriction does not exclude any origin-destination state pairing with more than five percent of the movers from the origin state.

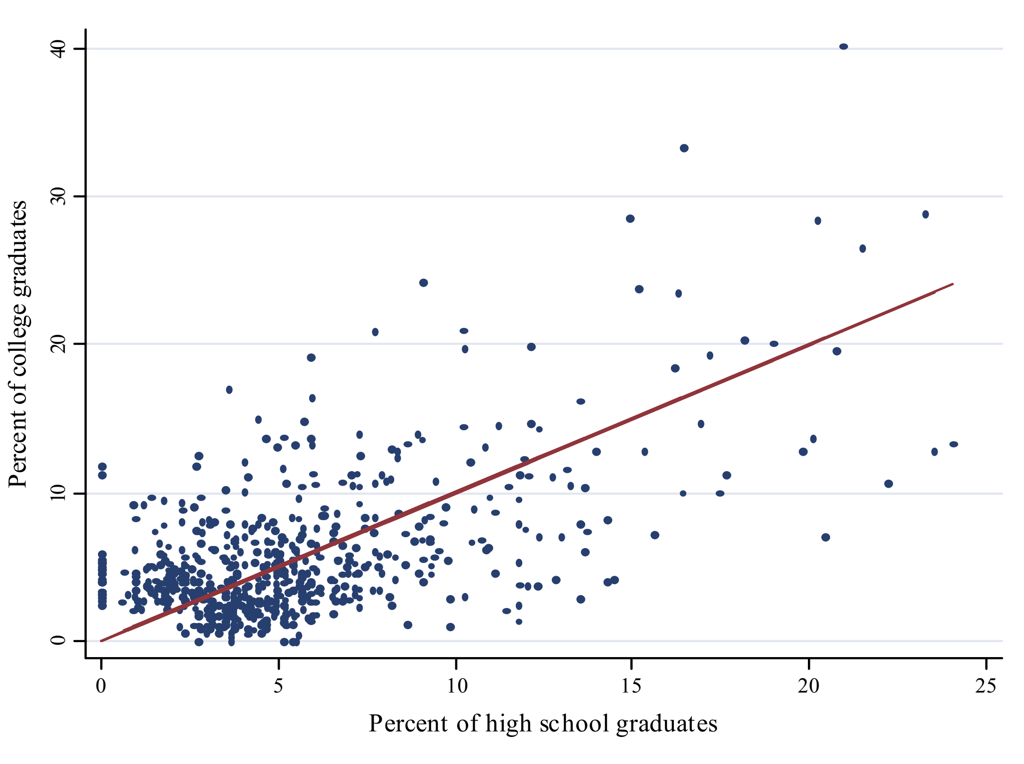


FIG. 5.5.—Do high school graduates and college graduates move to different markets? The percent of high school and college graduate interstate migrants moving from the same 1995 residence state to the same 2000 residence state.

If college graduates are attracted to states with high returns to schooling and high school graduates are repelled, then the points should lie along the axes. There are a few points along the y-axis, but not very many and even fewer along the x-axis. The correlation between the percent of high school graduates and the percent of college graduates with the same origin-destination pairing is 0.7549.[56] That high school and college graduates move together reinforces the observations made above, either gradients do not capture economic incentives or other factors (amenities, etc.) are driving migration. Either way, one should be cautious of studies that use interstate

---

[56] The correlation in computed excluding non-movers and weighted using the harmonic mean of the number of men in the 1995 residence state for the two schooling levels in each comparison. The rank correlation is slightly higher at 0.8012 and un-weighted correlations are very similar to these.

migration to identify quality effects. If it is the first explanation, then this warning is especially true for studies that measure the quality effect through schooling gradients.

Although the data does not support the notion that workers respond to gradient incentives to migrate, the data should support the idea that income is an incentive. The problem is that it is not clear how to normalize income for differences in costs of living across geographic areas. A simple way around this, for purposes of this examination, is to look at migration between two places. Does migration occur primarily in one direction so that it appears that workers are chasing economic rents? Or, does migration between locations offset so that it appears that other factors are driving interstate migration?

To explore this question, I plotted movements between large metropolitan areas (CMSA's) for college graduates. Figure 5.6 presents the number of college graduates moving between cities of five million or more inhabitants in 1995.[57] For example, between 1995 and 2000, 190 of the college graduates living in Chicago moved to D.C. and 159 of the college graduates living in D.C. moved to Chicago. The line is again for reference only and marks the points of equal migration between cities. The results in this figure are mixed. As with Figures 5.4a and 5.4b, there are large flows in both directions; however, there are also large gainers and losers. Los Angeles, for example, nets a gain of 298 college graduates from the other cities and San Francisco nets 138 college graduates. D.C., on the other hand, loses a net of 149 and Chicago 130. The large amount of movement between cities points to the importance of non-pecuniary factors, but the net gains and losses suggest that financial incentives play a role in helping to identify skill differences between workers migrating to different markets.[58]

---

[57] Linearly interpolated from 1990 and 2000 Census data. Population counts for larger metropolitan areas are obtained from the Census 2000 Redistricting Data (P.L. 94-171) Summary File, which contains counts for both 1990 and 2000.

[58] I also examined the occupations of movers between D.C. and Chicago and observed no difference in the distribution of occupations for those moving from D.C. to Chicago and the distribution of occupations for those moving from Chicago to D.C. The p-value from the Pearson Chi-Square statistic is 0.654. Occupational differences do not explain the cross-migration.

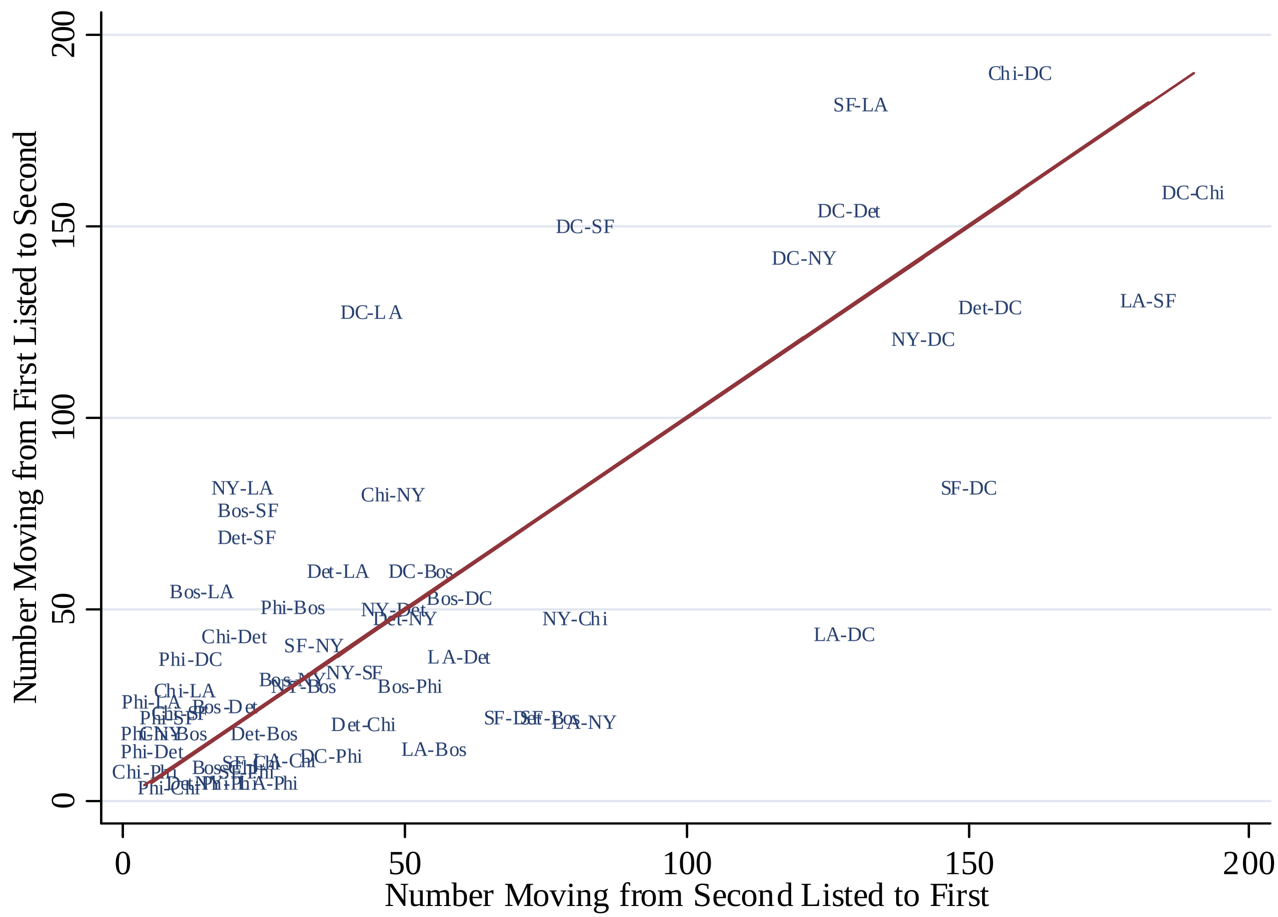


FIG. 5.6.—Movement between large cities (5,000,000 or more inhabitants) for white, non-Hispanic, prime-mover age men with a college degree, 1995 to 2000.

**Conclusions**

In this chapter, I demonstrate that birth state wage rankings within residence states, schooling levels and birth cohorts are not invariant to where they are evaluated. The number of reversals, i.e., those born in Texas do better than those from Pennsylvania in Washington but worse in Georgia, is overwhelming and completely undermines any confidence one might have had in existing school quality-earnings studies that use migration to identify school quality effects. Recall from Chapter II that studies that do not examine the quality-earnings relationship within markets do not have clear economic predictions (wages could be higher or lower for those from higher quality schools), so that these are no more reliable.

These reversals suggest that migration is selective; however, none of the studies in the literature have proposed a viable mechanism for making meaningful comparisons. The remainder of this chapter examined several factors that are important in migration decisions (age and education) and some that are not as strong (such as occupation) as well as presenting several observations regarding interstate migration that are impediments to identifying a school quality effect through earnings. These include: 1) gross migration dwarfs net migration so that it appears that non-economic factors play a major role in migration decisions, 2) the data do not support the idea of schooling-earnings gradients as incentives to migration and it's not clear what is captured by these gradients when migration is selective, 3) incomes should provide incentives to migrate but must be normalized for cost of living differences before they can be used, 4) and movements between large cities for those of the same schooling level provides mild support for the idea that workers follow economic incentives to migrate (there are net losers and gainers), but also suggests that there is much left to be explained (again, gross movements swamp net gains and losses).

Again, the presentation on migration was not meant to solve the selective migration problem, but merely to suggest that simple solutions such as the one suggested by Heckman, et al. and embraced by Card and Krueger will not untangle the comparisons in Table 5.1. It should also be noted that, even if migration could be "sorted out" so that the "other things" that are not equal and not adjusted for in these comparisons could be identified, the measurement error problems discussed in Chapters III and IV still exist and must be addressed prior to obtaining reliable estimates of the impact of the quality of schooling on earnings.

# CHAPTER VI
# CONCLUSIONS

Much of the economic literature on schooling has focused on differences in amounts earned by those of different schooling levels and found that additional years of schooling lead to higher average earnings levels. A smaller group of studies has tried to quantify the impact of measurable school characteristics on earnings. Undoubtedly, few would contend with the presumption that higher "quality" schools produce more successful students as the statement is tautological. The quality of schools is important as it affects not only individual income, but also national economic growth. In a recent piece, Hanushek (2003) presents calculations that estimate that if schools had responded to the warning issued in *A Nation at Risk* in 1983 and done something to improve math and science test scores by one standard deviation, per-capita gross domestic product (GDP) would have grown by one percent more per year and that the cumulative effect would have been to increase GDP in 2002 by $450 billion. This amount would have more than paid for all elementary and secondary schooling in that year. Unfortunately, the "something" that could have or should have been done is not identified.

This is the problem. We want to increase the performance of students by increasing the quality of schools, but the research is devoid of any policy recommendations. The school quality literature has two major branches. The first examines the impact of school quality measures on test scores and generally finds no consistent relationship between these measures and scores, but does find that there are large differences in performance between schools and within school there are large differences in performance between classes (teachers). Schools matter and teachers matter, but quantifying why or how they matter has proved more difficult. It is also unsettling that it is never clear what these tests measure and so we don't know what higher test scores identify. On the other hand, the branch that examines the impact of school quality measures on earnings generally concludes that these same school quality measures have a positive and significant effect on earnings.

This dissertation has examined this second group of studies that focus on earnings and found that the positive relationships observed come from data that are inconsistent with the implications of the underlying economic theory. The theory predicts that if schools differ in quality, then workers with the same level of schooling but educated in different schools should earn different amounts and that the differences (or rankings of average wages) should be pervasive across the markets in which education is sold. The most highly cited studies, those using the census data, have attempted to identify differences in school quality by the birth state of workers. A comparison of the average wages of workers born in different states, but working in the same market (state) shows that average wages by birth state are not consistently ranked across markets. As such, it is unclear what these studies have identified and this peculiar finding may be partly due to errors in identifying the school(s) attended, the inclusion of college trained labor coupled with the omission of any measure of college quality in the estimates, or selective migration such that workers that move to different markets possess different amounts of unobserved skill.

But even if we suppose that these school quality-earnings studies are meaningful, what would we have actually learned from them? Most conclude that increased schooling expenditures—or its major component, teacher salaries—lead to students with higher earnings after they graduate. There are no policy recommendations, however, as to how the money should be spent. There are no recommendations as to which curricula or which teaching paradigms lead to greater success or whether different programs are better suited to some students than others. What method(s) for compensating teachers is best? Is it really best to randomly place students into classes and teach them in exactly the same way placing the same emphasis on each subject for every student? Might different methods be more successful for some students than others? Might allowing for differences in emphasis by student lead to greater success after graduation? These studies are noticeably and necessarily silent on questions such as these.

But suppose that this were not the case, suppose that they did have correct and specific policy recommendations. Would they be relevant? Consider the Card and

Krueger (1992a) paper that uses 1980 Census data for male workers ages 30 to 60. These workers would have been educated between 1926 and 1968. If policy recommendations are not forthcoming until as much as 67 years after the implementation of some practice, it is doubtful that the recommendation will be of much use in formulating current policy. This suggests that testing is the preferred avenue for identifying school quality effects. But the discussion above pointed to the major problem with existing tests: it is not clear what is being identified. This suggests that new tests, that are correlated with earnings or whatever other measure of success that the policy maker wants to emphasize, need to be developed. If reading and mathematical ability are not the only skills thought to impact earnings, the tests should not be limited to knowledge in these areas. Could tests be developed to measure motivation, ambition, values, interpersonal and other skills that are valued in the market? What role do family and communities play? Studies that examine earnings controlling for age and years of schooling or performance on intelligence tests such as the AFQT leave most of the variation in individual salaries unexplained. This points to the importance of identifying these other skills and a reasonable place to start would be testing those who have already graduated and are now competing in the labor market. Development of these tests could allow for the almost immediate feedback to changes in the manner in which students are taught.

Most studies work with existing data that is limited, if not completely lacking, in its ability to identify these skills. Why are new tests not being developed and implemented? Why are new data not being collected? Cost? We are now spending three-quarters of a trillion dollars per year (and 7.5 percent of GDP) on education—limited to elementary, secondary, and accredited post-secondary schooling—with very little information as to the best way to allocate those funds and the economic literature on school quality has done little if anything to add to that knowledge. Perhaps it is time that we reevaluated the way that we are studying this important topic and developed new tools (tests) and new data that would allow for the identification of those aspects of schools and teachers that lead to greater student success, economic and otherwise, without being subject to the debilitating criticisms presented in this dissertation.

## REFERENCES

Akin, John S., and Garfinkel, Irwin. "School Expenditures and the Economic Return to Schooling." *Journal of Human Resources* 12 (Winter 1977): 460-481.

———. "The Quality of Education and Cohort Variation in Black-White Earnings Differentials: Comment." *American Economic Review* 70 (March 1980): 186-191.

Akin, John S., and Kniesner, Thomas J. "Proxies for Observations on Individuals Sampled from a Population." *Journal of Human Resources* 11 (Summer 1976): 411-413.

Altonji, Joseph G., and Dunn, Thomas A. "Using Siblings to Estimate the Effect of School Quality on Wages." *Review of Economics and Statistics* 78 (November 1996): 665-671.

Astin, Alexander W. *Who Goes Where to College?* Chicago: Science Research Associates, 1965.

Behrman, Jere R.; Rosenzweig, Mark R.; Taubman, Paul. "College Choice and Wages: Estimates Using Data on Female Twins." *Review of Economics and Statistics* 78 (November 1996): 672-685.

Betts, Julian. "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth." *Review of Economics and Statistics* 77 (May 1995): 231-250.

———. "Do School Resources Matter Only for Older Workers?" *Review of Economics and Statistics* 78 (November 1996): 638-652.

———. "The Impact of School Resources on Women's Earnings and Educational Attainment: Findings from the National Longitudinal Survey of Young Women." *Journal of Labor Economics* 19 (July 2001): 635-657.

Card, David, and Krueger, Alan B. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy* 100 (February 1992): 1-40. (a)

———. "School Quality and Black-White Relative Earnings: A Direct Assessment." *Quarterly Journal of Economics*107 (February 1992): 151-200. (b)

———. "Labor Market Effects of School Quality: Theory and Evidence." in *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success* edited by Gary Burtless. Washington, D.C.: The Brookings Institution, 1996.

Carter, Allan M. *An Assessment of Quality in Graduate Education*. Washington, D.C.: American Council on Education, 1966.

Cass, James, and Birnbaum, Max. *Comparative Guide to American Colleges*. New York: Harper and Row, 1964.

Gourman, Jack. *The Gourman Report; Ratings of American Colleges*. Phoenix: Continuing Education Institute, 1967.

Griffin, Peter, and Ganderton, Philip T. "Evidence on Omitted Variable Bias in Earnings Equations [family effects]." *Economics of Education Review* 15 (April 1996): 139-148.

Grogger, Jeff. "School Expenditures and Post-Schooling Earnings: Evidence from the High School and Beyond." *Review of Economics and Statistics* 78 (November 1996): 628-637. (a)

———. "Does School Quality Explain the Recent Black/White Wage Trend?" *Journal of Labor Economics* 14 (April 1996): 231-253. (b)

Hanushek, Eric. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* XXIV (September 1986): 1141-1177.

———. "Lost Opportunity." *Education Next* 3 (Spring 2003): 84-87.

Heckman, James; Layne-Farrar, Anne; and Todd, Petra. "Human Capital Pricing Equations with an Application to Estimating the Effect of Schooling Quality on Earnings." *Review of Economics and Statistics* 78 (November 1996): 562-610.

Johnson, George E., and Stafford, Frank P. "Social Returns to Quantity and Quality of Schooling." *Journal of Human Resources* 8 (Spring 1973): 139-155.

Johnson, William R., and Neal, Derek A. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy* 104 (October 1996): 869-895.

Jud, G. Donald, and Walker, James L. "Discrimination by Race and Class and the Impact of School Quality." *Social Science Quarterly* 57 (March 1977): 731-749.

Link, Charles R. "The Quantity and Quality of Education and Their Influence on Earnings: The Case of Chemical Engineers." *Review of Economics and Statistics* 55 (May 1973): 241-247.

———. "Graduate Education, School Quality, Experience, Student Ability, and Earnings." *Journal of Business* 48 (October 1975): 477-491.

Link, Charles R., and Ratledge, Edward C. "The Influence of the Quantity and Quality of Education on Black-White Earnings Differentials: Some New Evidence." *Review of Economics and Statistics* 57 (August 1975): 346-350. (a)

———. "Social Returns to Quantity and Quality of Education: A Further Statement." *Journal of Human Resources* 10 (Winter 1975): 78-89. (b)

———. "Proxies for Observations on Individuals Sampled from a Population: Replay." *Journal of Human Resources* 11 (Summer 1976): 413-419.

Link, Charles; Ratledge, Edward; and Lewis, Kenneth. "Black-White Differences in Returns to Schooling: Some New Evidence." *American Economic Review* 66 (March 1976): 221-223.

———. "The Quality of Education and Cohort Variation in Black-White Earnings Differentials: Reply." *American Economic Review* 70 (March 1980): 196-203.

Mincer, Jacob. *Schooling, Experience, and Earnings*. New York: Columbia University Press, 1974.

Morgan, James, and Sirageldin, Ismail. "A Note on the Quality Dimension in Education." *Journal of Political Economy* 76 (September – October 1968): 1069-1077.

Morgenstern, Richard D. "Direct and Indirect Effects on Earnings of Schooling and Socio-Economic Background." *Review of Economics and Statistics* 55 (May 1973): 225-233.

Nechyba, Thomas J. "The Southern Wage Gap, Human Capital and the Quality of Education." *Southern Economic Journal* 57 (October 1990): 308-322.

Parnes, Herbert S., and Kohen, Andrew I. "Occupational Information and Labor Market Status: The Case of Young Men." *Journal of Human Resources* 10 (Winter 1975): 44-55.

Ribich, Thomas I., and Murphy, James L. “The Economic Returns to Increased Educational Spending.” *Journal of Human Resources* 10 (Winter 1975): 56-77.

Rivkin, Steven G. “School Desegregation, Academic Attainment, and Earnings.” *Journal of Human Resources* 35 (Spring 2000):333-346.

Rizzuto, Ronald, and Wachtel, Paul “Further Evidence on the Returns to School Quality.” *Journal of Human Resources* 15 (Spring 1980): 240-254.

Snyder, T.D.; Tan, A.G.; and Hoffman, C.M. *Digest of Education Statistics 2003*. U.S. Department of Education, National Center of Education Statistics. Washington, D.C.: Government Printing Office, 2003.

Solmon, Lewis. “The Definition and Impact of College Quality.” *Explorations in Economic Research* 2 (Fall 1975): 537-587.

Speakman, Robert, and Welch, Finis. “Using Wages to Infer School Quality.” In *Handbook of the Economics of Education, Vol. 2*, edited by Eric Hanushek and Finis Welch. Amsterdam: Elsevier Science B.V., forthcoming 2007.

Strayer, Wayne. “The Returns to School Quality: College Choice and Earnings.” *Journal of Labor Economics* 20 (July 2002): 475-503.

Tremblay, Carol H. “The Impact of School and College Expenditures on the Wages of Southern and Non-Southern Workers.” *Journal of Labor Research* 7 (Spring 1986): 201-211.

U.S. Department of Education, National Center for Education Statistics. *The Condition of Education 2004*. Washington, D.C.: Government Printing Office, 2004.

Wachtel, Paul. “The Effect of School Quality on Achievement, Attainment Levels, and Lifetime Earnings.” *Explorations in Economic Research* 2 (Fall 1975): 502-536.

———. “The Effect of Earnings of School and College Investment Expenditures.” *Review of Economics and Statistics* 58 (August 1976): 326-331.

Wales, Terence J. “The Effect of College Quality on Earnings: Results from the NBER-Thorndike Data.” *Journal of Human Resources* 8 (Summer 1973): 306-317.

Welch, Finis. “Measurement of the Quality of Schooling.” *American Economic Review* 56 (May 1966): 379-392.

———. “Education in Production.” *Journal of Political Economy* 78 (January 1970): 35-59.

# APPENDIX A

# LITERATURE SUMMARY*

The following abbreviations are adopted throughout this summary: teacher/pupil (T/P) or pupil/teacher (P/T), school quality measures (Q), years of schooling or a transform thereof (S), family characteristics (F), individual characteristics (X), and average daily attendance (ADA).

## A. Decennial Census Data Matched to The Biennial Survey of Education and the Digest of Education Statistics

### 1. Welch (1966)

***Data***

Earnings Source: 1960 Census from four sources: (1) State Reports, Series D, (2) the 1 in 1000 sample, (3) the matched sample for farm operator families of the 1960 Sample of the Survey of Agriculture and the Census of Population, and (4) the subject report "Whites with Hispanic surnames".

Earnings Sample: Rural farm males, ages 25 or more with earned income in 1959, who had not attended college, and who were not residing in Alaska, Hawaii or RI. Twelve "states" are dropped in Stage (2) due to insufficient school system data. ("States" are defined separately for the segregated Southern states.)

Quality Source: *Biennial Survey*, 1955-56.

Quality Variables: Teacher/Pupil ratio (T/P ratio), teacher wages (no adjustment mentioned), enrollment, and expenditures per pupil ($/pupil; no adjustment mentioned) for rural counties.

***Model***

(1) $\hat{y}_{ij} - \hat{y}_{0j} = B_i C_j$

(2) $\hat{C}_j = Q\alpha_2 AN^{\alpha_1}(Q\sum B_i N_i)^{\alpha_2 - 1} K^{\alpha_3}$

$\hat{y}_{0j}$ is the predicted earnings of a non-educated worker living in state *j* and $\hat{y}_{ij}$ is the predicted earnings of a worker with *i* years of schooling living in state

---

* Reprinted from Speakman, Robert and Welch, Finis, "Chapter 13. Using Wages to Infer School Quality" in *Handbook of the Economics of Education, Vol. 2* edited by Eric Hanushek and Finis Welch, Copyright 2007 with permission from Elsevier.

$j$. $C_j$ is the return to one unit of schooling and $B_i$ is the number of units of schooling possessed.

In practice, estimation proceeds in two stages. In the first stage, the (log of the) return to schooling is regressed on state ($j$) and years of schooling ($i$) indicators. In the second stage, the estimates of (the log of) $C_j$ are regressed on $Q$, $K$ (non-labor inputs), $N$ (people), and education (the term in parentheses, which includes college attendees; $B_i$ is estimated in Stage 1). Also included in Stage 2 is an indicator for non-white "states" and indicators for the Pacific and Southern regions.

***Stated Findings***

The T/P ratio and teacher wages are highly correlated with $/pupil reflecting a tradeoff between the two.

The effect of teachers' wages is consistently positive across specifications.

The coefficient on T/P ratio is negatively signed and seems to capture school size as rural schools with higher T/P ratios are typically smaller and have teachers with widely diverse teaching loads.

The most important determinants of school quality are teacher salaries and the size of the secondary school.

**2. Rizzuto and Wachtel (1980)**

***Data***

Earnings Source: 1960 and 1970 Censuses.

Earnings Sample: Excludes: those under age14 or over age 65, females, those with missing state-of-birth, those with mother tongue other than English or Spanish, non-whites other than black, NILF, self-employed, and those not born in the 50 states or D.C.

Quality Source: *Biennial Survey*, 1909-10, 1919-20, …, 1959-60.

Quality Variables: $/pupil in elementary school (extrapolated from measures that include elementary and high school combined; adjusted for price levels using the index for local and state government expenditures). Also examines P/T ratio, teachers' wages (CPI adjusted), term length, and various components of expenditures.

***Model***

$$y = X\beta + \delta S + \theta Q + \varepsilon$$

Estimated separately by census year and race.

y: (log) 1959 and 1969 annual earnings.

S: years of schooling.

X: experience, experience squared, ln(weeks worked), urban residence.

Q: ln($/pupil).

***Stated Findings***

The returns to $Q$ are positive and significant for whites and blacks in both census years; they are generally higher for blacks, higher in 1960, and higher than the returns to $S$.

The marginal social rate of return to $Q$ ranges between 5 percent and 18 percent (depending on race, census year, and schooling).

Expenditure measures perform better than the other quality measures.

### 3. Nechyba (1990)

***Data***

Earnings Source: 1950-1980 Census summary tables. Data collected for whites and blacks at the state-level for the 25-34, 35-44, 45-54 and 55-64 age cohorts.

Quality Source: *Biennial Survey* as well as data extracted from published and unpublished research.

Quality Variable: Expenditures on teacher salary per pupil (expressed as a ratio of black to white expenditures).

***Model***

$$y = X\beta + \delta S + \theta Q + \varepsilon$$

$y$: median annual earnings.

$S$: median education.

$X$: percent in urban residence, fraction of cotton picked manually in the last 5 and 10 years (Cotton5, Cotton10), and the number of discrimination charges filed with the EEOC relative to the black labor force (AA).

Observations are at the state level. All variables except AA and COTTON* are ratios of the values for blacks relative to that for whites.

***Stated Findings***

$Q$ has a positive and significant effect on earnings.

Almost half of the closing gap in earnings ratios between whites and blacks is attributable to changes in the relative school quality between blacks and whites.

### 4. Card and Krueger (1992a)

***Data***

Earnings Source: 1980 Census.

Earnings Sample: White males born in the 48 states or D.C. between 1920 and 1949 and living in the 50 states or D.C. Those with imputed values for age, race, sex, education, weeks worked, or earnings are excluded as are those reporting no weeks worked. Those with annual income <$101 or with weekly wages <$36 or >$2,500 are also excluded.

Quality Source: *Biennial Survey* and *Digest of Education* (many years).

Quality Variables: P/T ratio, term length, and teacher wage (normalized by state average wages and divided by the national average in the year). Other measures examined include: percent male teachers, teacher experience, teacher education, private school attendance, and the difference in the P/T ratio between Catholic and public schools.

***Model***

(1) $y = X_1\beta_1 + S\Delta\delta_1 + S\Gamma\delta_2 + \varepsilon$

(2) $\delta_1 = X_2\beta_2 + Q\theta + \upsilon$

Stage 1 regressions are run separately for each 10-year birth cohort. Stage 2 regressions are weighted by the inverse variance of the dependent variable.

*y*: (log) 1979 weekly wages.

*S*: years of schooling truncated at the 2nd percentile for the cohort in the state where born. Δ and Γ represent state-of-birth and region-of-residence indicators.

$X_1$: experience, experience squared, marital status, SMSA status, state-of-birth and state-of-residence indicators.

$X_2$: cohort indicators and (in their preferred specification) state-of-birth indicators.

Extensions:

1) Adds median education, (log) per capita income of parents' generation, percent high school graduates and percent college graduates in the Stage 2 regressions.

2) A reduced-form model that includes *Q* and excludes *S* in Stage 1 (all cohorts combined, without cohort indicators).

***Stated Findings***

The coefficients on the quality measures are, in general, correctly signed and statistically significant.

The inclusion of family background, labor supply factors (the education distribution of the labor force), and additional quality measures do not change the conclusions from the basic model.

The state-of-birth intercepts are negatively correlated with *Q* and *S* is positively correlated with *Q*. The positive correlation with *S* more than offsets the negative correlation with the state-of-birth intercepts so that there is no evidence of a negative effect of school quality in the lower tail of the earnings distribution.

Estimates of the impact of *Q* on *Y* from the reduced-form models are 30-40 percent larger than in the two-stage models.

## 5. Card and Krueger (1992b)

***Data***

Earnings Source: 1960-1980 Censuses.

Earnings Sample: White and black males born in 18 Southern states (including D.C.), living in the metro areas of nine Northern states. See Card and Krueger (1992a) for exclusions based on imputations and wages.

Quality Source: *Biennial Survey*,1920-54 and Southern Education Reporting Services publications for the 1955-1966 data.

Quality Variables: P/T ratio, term length, and teachers' wages.

***Model***

(1) $y = X\beta + S\Delta\delta + \varepsilon$

(2) $\delta = X_2\beta_2 + \theta Q + \upsilon$

Stage (1) is run separately by race, Census year and 10-year birth cohort. The Stage (2) regressions are weighted by the inverse variance of the dependent variable. Card and Krueger (1992a) also estimate a reduced-form model that includes *Q* directly in Stage (1) and excludes *S*.

*y*: (log) weekly wages.

*S*: years of schooling, Δ represents state-of-birth indicators.

$X_1$: experience, experience squared, state-of-residence and state-of-birth indicators.

$X_2$: race by cohort indicators, race by census year indicators, and state-of-birth indicators.

***Stated Findings***

The primary model includes only the P/T ratio, which has a positive and significant effect on earnings.

When teacher salaries and term length are added, the coefficient on teacher salaries is consistently positive and significant under various specifications, but the coefficients on P/T ratio and term length change signs depending upon the specification.

School quality changes explain 50-80 percent of the change in the relative black/white return to education between the 1910 and 1940 birth cohorts.

## 6. Heckman, Layne-Farrar and Todd (1996)

***Data***

Earnings Source: 1970-1990 Censuses.

Earnings Sample: Parallels the Card and Krueger (1992a) sample restrictions with adjustments made for each Census year given real dollar amounts and the variables available.

Quality Source: *Biennial Survey*, *Digest of Education*, and state reports.

Quality Variables: P/T ratio, term length, and teachers' wages (normalized by regional average wage). Variables are adjusted given the timing of desegregation.

***Model***

(1) $y = X_1\beta_1 + S\Delta\delta_1 + S\Gamma\delta_2 + \varepsilon$

(2) $\delta_1 = X_2\beta_2 + Q\theta + \upsilon$

See Card and Krueger (1992a) for a description of the variables.

Extensions:

1) Examines the effect of $Q$ on the state-of-birth intercepts.

2) Examines sheepskin effects by adding interactions between some college and college graduate by state-of-birth indicators and by region-of-residence indicators.

3) Adds region-of-residence by region-of-birth indicators in Stage (1).

4) Adds region-of-residence by $Q$ interactions and migration distance in Stage (2).

***Stated Findings***

Correlations within region of residence (ROR) between region of birth (ROB) average wages and ROB averages of the quality measures are weak and inconsistent. However, rankings of ROB average wages within ROR show some consistency across ROR, especially for those with four or more years of college. This supports the existence of early environmental factors, but not school quality measures.

The correlation between state-of-birth (SOB) intercepts from Stage (1) and $Q$ is negative. Examining only the effect of $Q$ on $S$ will overstate the total impact of $Q$ on earnings.

Sheepskin effects are important, i.e. there are discrete jumps in the return to education for those with some college and those with 4 or more years of college.

ROB by ROR indicators (in Stage (1)) are significant and reduce the impact of $Q$ on earnings. This suggests that selective migration is an important issue. "Because of this … interaction, … no unique quality effect on returns to education can be defined independently of the market in which it is used."

When ROB by ROR interactions and sheepskin effects are included in Stage (1), "the only support for an effect of school quality on earnings is through the return to college education."

Allowing the effect of $Q$ to vary by ROR and adding migration distance in the second stage weakens (and sometimes reverses) the impact of $Q$ and reveals that the effects of $Q$ vary across regions, with the exception of teacher salary, which is positive, significant, and not statistically different across regions.

$Q$ is positively correlated with the percentage of college graduates, is weakly correlated with the percent of high school graduates, and is negatively correlated with high school dropout rates.

## B. The National Longitudinal Survey of Young Men /Young Women (NLS-YM/YW)[59]

### 7. Link and Ratledge (1975a)

***Data***

Earnings Source: NLS-YM conducted in 1967.

Earnings Sample: Males ages 15-25 out of school at least one year with positive earnings.

Quality Source: NLS-YM.

Quality Variable: $/pupil in ADA in 1968 in the district where attended high school.

***Model***

$$y = X\beta + \delta S + \theta Q + F\pi + A\varphi + \varepsilon$$

Model run separately for whites and blacks and then run with both groups combined including a race indicator.

$y$: 1967 annual earnings.

$S$: years of schooling.

$X$: experience [minimum of (age-education-6) and (age-16)], hours worked, marital status, occupational knowledge test.

$F$: Duncan index of father's occupation and whether had access to newspapers, magazines or a library card at age 14.

$A$: IQ.

[59] Link, Ratledge, and Lewis (1980) examine the NLS-YM and the PSID.

***Stated Findings***

The coefficient on *Q* is positive and significant in all models and the internal rate of return associated with additional expenditures is higher for blacks than whites.

The coefficient on "black" in the total population regression is not significant.

**8. Link and Ratledge (1975b)**

***Data***

Earnings Source: NLS-YM conducted in 1968.
Earnings Sample: Males ages 16-26 out of school at least one year.
Quality Source: NLS-YM.
Quality Variable: $/pupil in ADA in 1968 in the district where attended high school.

***Model***

$$y = X\beta + \delta S + \theta Q + F\pi + A\varphi + \varepsilon$$

Model estimated separately for whites and blacks.
*y*: (log) 1968 annual earnings.
*S*: years of schooling.
*X*: experience [minimum of (age-16) and (age-S-5)], experience squared, urban residence (integers from 1 to 6), the log of hours worked last year.
*F*: residence at age 14 (integers from 1 to 6 representing urbanness).
*A*: IQ.
*Q*: ln($/pupil).

***Stated Findings***

This study parallels Johnson and Stafford (1973) using district-level expenditure data and reports similar findings.

The coefficient on *Q* is positive and significant for both whites and blacks and the return-to-quality is higher for blacks than for whites.

The omission of *A* from the equation causes the return-to-education to be overstated by 15 percent and the return-to-quality to be overstated by 10 percent.

**9. Parnes and Kohen (1975)**

***Data***

Earnings Source: NLS-YM conducted in 1968.

Earnings Sample: Males ages 14-24 in 1966 with at least 8 years of schooling, not currently enrolled, and currently employed.

Quality Source: NLS-YM.

Quality Variables: An index constructed by the authors that includes the availability of library facilities, P/T ratio, counselors/100 students, and starting teacher's salary. (Construction not described by authors.)

***Model***

$$y = X\beta + \delta S + \theta Q + F\pi + A\varphi + \varepsilon$$

Model estimated separately for blacks and whites.

*y*: 1968 hourly earnings.

*X*: experience, health, South, SMSA, an occupational knowledge test (with three components).

*S*: years of schooling.

*F*: a constructed index based on father's education, mother's education, father's occupation, education of oldest older sibling; and availability of reading materials in the home.

*A*: IQ.

***Stated Findings***

The coefficient on *Q* is positive but not significant for blacks and is negative and not significant for whites. However, the occupational knowledge test, the primary focus of their study, has a positive and significant effect on earnings.

This study also examines the determinants of the occupational knowledge test, including *Q* (counselors per 100 pupils in the high school) which has a negative effect (higher quality leads to lower test scores) in three of the four specifications and the coefficient on *Q* is never statistically significant.

**10. Link, Ratledge and Lewis (1976)**

***Data***

Earnings Source: NLS-YM conducted in 1969.

Earnings Sample: Males ages 17-27 out of school at least one year.

Quality Source: NLS-YM.

Quality Variable: $/pupil in ADA in 1968 in the district where attended high school.

***Model***

$$y = X\beta + \eta SQ + \gamma S * EXP + \varepsilon$$

Model estimated separately for blacks and whites.

*y*: (the log of) 1969 annual earnings.
*S*: years of schooling.
*X*: ln(weeks worked), experience [minimum of (age-education-5) and (age-16)], and experience squared.
*Q*: ln($/pupil).

***Stated Findings***

This study examines the "vintage effect" of *Q* (the changing relative quality of education for blacks and whites over time) on black/white wage ratios.

The schooling by quality interaction is positive and statistically significant for both blacks and whites (evidence of a vintage effect), where as the schooling by experience interaction is not statistically significant (no evidence of a life cycle effect).

## 11. Jud and Walker (1977)

***Data***

Earnings Source: NLS-YM conducted in 1970.
Earnings Sample: Males ages 18-28 out of school at least one year with positive earnings.
Quality Source: NLS-YM.
Quality Variable: $/pupil in ADA in 1968 in the district where attended high school.

***Model***

$$(1)\ A = X_1\beta_1 \qquad Q\theta_1 + F_1\pi_1 \qquad + \varepsilon_1$$
$$(2)\ S = X_1\beta_2 \qquad + Q\theta_2 + F_1\pi_2 + A\varphi_2 + \varepsilon_2$$
$$(3)\ y = X_2\beta_3 + S\delta_3 + Q\theta_3 + F_2\pi_3 + A\varphi_3 + \varepsilon_3$$

*y*: 1970 annual earnings.
*S*: years of schooling.
$X_1$: race.
$X_2$: race, experience, experience squared, a vocational tech indicator, marital status, ability to work indicator, hours/week, South, and SMSA.
$F_1$: Duncan's index of father's occupation, size of community at age 14 (1=farm, …, 6=100,000+), father's education, mother's education.
$F_2$: Duncan index of father's occupation.
*A*: IQ.

***Stated Findings***

The coefficient on $Q$ is a statistically significant and positive determinant of ability, but not of schooling or earnings directly; this suggests that the impact of $Q$ is transmitted through a sequential process.

**12. Link, Ratledge and Lewis (1980)**

***Data***

Earnings Source 1: NLS-YM conducted in 1971 (updated from Link, Ratledge, and Lewis (1976)).

Earnings Sample 1: Males ages 19-29 who worked at least 30 weeks, at least 30 hours per week, had positive wages, and were not in school.

Earnings Source 2: PSID 1968-1972. Designed to match Akin and Garfunkel (1980), male household heads, ages 30-55 excluding self-employed and those with no earnings. Several data handling errors committed by Akin and Garfunkel are corrected.

Quality Source: NLS-YM and the *Biennial Survey* for 1930, 1940, 1950, 1960.

Quality Variables: $/pupil in 1968 in the district where attended high school (NLS-YM). $/pupil (*Biennial Survey*) was matched to the PSID based on state-of-residence at age 12 (linearly interpolated) to replicate A/G.

***Model***

$$(AG) \quad y = X\beta_1 + \delta_2 S + \gamma_1 S{*}Exp + \theta_1 Q + \eta_1 SQ + \varepsilon_1$$
$$(LRL-A) \quad y = X\beta_2 + \gamma_2 S{*}Exp + \eta_2 SQ + \varepsilon_2$$
$$(LRL-B) \quad y = X\beta_3 + \delta_3 S + \gamma_3 S{*}Exp + \eta_3 SQ + \varepsilon_3$$

Regressions are run separately for blacks and whites.

NLS-YM

$y$: (log) 1971 hourly wage race.

$X$: experience [minimum of (age-S-5) and (age-16)], experience squared.

$F$: father's education.

$A$: IQ.

$Q$: ln($/pupil).

PSID: See Akin and Garfinkel (1980).

***Stated Findings***

This study examines the vintage effect of $Q$ on earnings for blacks and whites in response to A/G and its findings run counter to those of A/G.

Using the NLS-YM, the returns to $Q$ are positive and significant in all models (except the A/G equation for blacks where they may or may not be jointly significant) and the elasticities range between .13 and .23 and are, in general, slightly higher for blacks. The returns to schooling are comparable for young blacks and whites, but unfavorable for older blacks relative to older whites.

Using the PSID, the returns to $Q$ are also positive and significant in all models (jointly in the A/G model) and the elasticity estimates range from 0.14 to 0.32. Returns to schooling are favorable for younger blacks and unfavorable for older blacks relative to whites of the same ages.

## 12. Tremblay (1986)

***Data***

Earnings Source: NLS-YM conducted in 1976.

Earnings Sample: Males ages 24-34 working full-time (35-40 hrs/week) and not enrolled in 1976. Wages are adjusted for regional (South, non-South, metro, non-metro) cost-of-living differences.

Quality Source: NLS-YM? (Source not mentioned.)

Quality Variables: $/pupil in primary and secondary school and $/pupil in 1970-71 at the most recently attended college. ($/pupil amounts are adjusted for regional cost-of-living differences and time using the CPI deflator).

***Model***

$$y = X\beta + \delta S + \theta_{PS}(Q_{PS} \times S_{PS}) + \theta_c(Q_C \times S_C) + \varepsilon$$

Regressions run separately for the South and the non-South.

$y$: (log) hourly wage rate.

$S$: years of schooling, $S_{PS}$ is years of primary and secondary education, $S_C$ is years of college.

$X$: seniority at current job, age, race, marital status, union, SMSA, industry (11 values), occupation (10 values), and whether participated in an occupational training program.

$A$: IQ.

$Q$: ln($/pupil).

***Stated Findings***

The coefficients on $Q$ (both pre-college and college) are positive in both the South and the non-South regressions, but are only significant in the South.

Estimated elasticities are higher in the South and for primary and secondary expenditures than for college expenditures.

## 13. Altonji and Dunn (1996)

### *Data*

Earnings Source: NLS-YM and NLS-YW. Utilizes all surveys through 1981 for the YM and through 1988 for the YW.

Earnings Sample: Males ages 14 to 24 in 1966 and females ages 14-24 in 1967 who were at least 24 years old at the time of the survey, not enrolled (and didn't return to school that year), with wages above $1.50/hour in 1982-84 dollars and with valid school quality data.

Quality Source: NLS-YM and NLS-YW.

Quality Variables: P/T ratio, starting wages of teachers with a B.S., $/pupil, and a composite measure constructed by the NLS. Also examines enrollment, percent disadvantaged, and percent dropping out.

### *Model*

$$y = X\beta + S\delta + SZ\gamma + Q\theta + SQ\eta + F\pi + \varepsilon$$

The model is also estimated including (*A*S*) and using differences in residence at age 14 as an instrument for differences in *Q*. All variables except *Q* are interacted with gender.

*y*: (log) hourly earnings.

*S*: years of schooling minus 12 (cubic).

X: year indicators, Experience (quartic), female, South, SMSA.

*Z*: mother's education, father's education.

*F*: two specifications: (1) family indicators or (2) # of siblings, black, and two parents in household at age 14.

*A*: IQ (model run with and without).

### *Stated Findings*

While the level effects (*Q* not interacted with *S*) are all correctly signed and significant (except P/T ratio), the coefficients on the interactions of *Q* with *S* are generally (6 of 8 estimates) perversely signed.

Adding an indicator for each family increases the coefficients on the quality variables.

Adding ability and an ability by schooling interaction has very little effect on the quality estimates.

The coefficients on *Q* are identified by movers within family in the specification with family indicators. IV estimates using city size as an instrument result in correctly signed level and interaction estimates, except for the P/T ratio by *S* interaction.

"We find that teachers' salaries, expenditures per pupil, and a composite index of school quality indicators have a substantial positive effect on the wages of high school graduates."

### 14. Betts (2001)

***Data***

Earnings Source: NLS-YW, includes all surveys between 1966 and 1991 (ages 14-24 in 1967).

Earnings Sample: Black and white women ages 18 or more and not in school.

Quality Source: NLS-YW from the 1968 survey.

Quality Variables: $/pupil, P/T ratio, starting wages of teachers with a B.S., and library books per student. (Dollar amounts are adjusted by survey administrators for cost-of-living differences between cities.)

***Model***

$$y = X\beta + Q\delta + F\gamma + \varepsilon$$

Models run separately for blacks and whites. Also examines a model with age interacted with $Q$.

$y$: (log) hourly wages.

$X$: age, age squared, marital status, # of children, year indicators, and an inverse Mills ratio (from a probit model of labor force status).

$F$: whether educated in a large city (25k+), whether educated in the South, census region where educated, father's education, mother's education, Duncan index of family head, and # of siblings in 1968.

The model does not include years of schooling.

***Stated Findings***

Elasticities are higher for black women than for white women. P/T ratios and books/pupil are significant in both regressions, although correctly signed for blacks and perversely signed for whites.

When outliers are omitted, all four measures are significant for blacks, with no meaningful changes in the results for whites.

The effects of $Q$ generally weaken with age.

## C. The National Bureau of Economic Research – Thorndike and Hagen Data (NBER-TH) Matched to the *Biennial Survey* and Various College Data Sources

### 15. Wales (1973)

***Data***

Earnings Source: NBER-TH.

Earnings Sample: No restrictions mentioned.

Quality Source: *The Gourman Report*.

Quality Variables: The Gourman ranking (averaged across departments) was divided into quintiles and matched to the undergraduate and graduate institution attended.

***Model***

$$y = X\beta + \eta SQ + F\pi + A\varphi + \varepsilon$$

*y*: 1969 monthly earnings.

*X*: whether a college teacher, whether other teacher, religion indicators, age, and marital status.

*S*: indicators for some college, bachelor's degree, some graduate training, master's degree, and Ph.d. interacted with quintiles of the undergraduate and graduate Gourman index (two separate indexes). Only significant interactions are included.

*F*: indicators for quintiles of a composite hobbies index, family income, pre-1943 schooling, math ability, father's education, and other characteristics.

*A*: indicators for quintiles of a "mathematical factor" taken from a factor analysis of 17 ability tests.

***Stated Findings***

This study examines the effect of college quality on earnings.

Quality is a positive and significant determinant of earnings at both the graduate and undergraduate levels.

This may capture increases in marginal productivity, screening by firms, or omitted ability variables.

The quality of undergraduate training is positively (but weakly) correlated with graduate training.

**16. Solmon (1975)**

***Data***

Earnings Source: NBER-TH.

Earnings Sample: Respondents with at least some college.

Quality Source: Various sources including Cass and Birnbaum's *Comparative Guide of American Colleges*, Astin's *Who Goes Where to College*, *The Gourman Report*, Carter's *An Assessment of Quality in Graduate Education*, etc. Quality variables are matched to the most recently attended college.

Quality Variables: Quality of instruction includes average faculty salary; expenditures on instruction, research and library facilities per full-time equivalent student; T/P ratio; income and expenditures per pupil; and the

Gourman index (overall and academic). Quality for students includes SAT scores, Astin's intellectualism index, and Astin's selectivity index.

***Model***

$$y = X\beta + \delta S + \theta Q + A\varphi + \varepsilon$$

*y*: (log) 1969 annual earnings.
*S*: years of schooling.
*X*: experience, experience squared.
*A*: IQ.

Extensions examine data by schooling levels, 1955 earnings, and include general occupation categories.

***Stated Findings***

This study examines the effects of college quality on earnings.

Each quality variable is significant when examined in isolation. When entered simultaneously, faculty salary and peer effects (SAT verbal score) appear to be the most important components of college quality.

The impact of *Q* increases with experience.

Extra schooling has a greater impact for those from lower quality undergraduate institutions.

The return to graduate school quality is higher for those who attended higher quality undergraduate institutions.

The coefficients on Q change only slightly when family background variables are included.

**17. Wachtel (1975)**

***Data***

Earnings Source: NBER-TH.

Earnings Sample: Excludes those who are airplane pilots, are unmarried, are in poor health, attended private or parochial schools, have earnings<$4K or >$75K (in 1958$) in 1955 or <$5K or >$75K in 1969 and those for whom school expenditure data could not be matched.

Quality Source: *Biennial Survey*, district-level elementary and high school expenditures per pupil for 1936-38. *The Gourman Report*. Office of Education data for college expenditures. National Research Council (NRC) data for the high school attended.

Quality Variables: Primarily examines $/pupil for pre-college education, but also examines P/T ratio, teacher wage, teacher wage relative to state median income, enrollment per building, term length, and expenditures on texts and other instructional expenses. College quality measures include the Gourman index and \$/student at the college attended. The NRC data (matched for a sub-sample and using data from the actual high school attended) includes size of graduating class, percent of teacher's with a graduate degree, and percent of high school graduates with Ph.d.'s.

***Model***

$$(1)\quad A = Q\theta_1 + F_1\pi_1 + \varepsilon_1$$

$$(2)\quad S = X_1\beta_2 + Q\theta_2 + F_1\pi_2 + A\varphi_2 + \varepsilon_2$$

$$(3)\quad y = X_2\beta_3 + S\delta_3 + Q\theta_3 + F_2\pi_3 + A\varphi_3 + \varepsilon_3$$

$y$: (log) annual earnings in 1969 (also examines 1955 earnings).
$S$: years of schooling.
$X_1$: age.
$X_2$: experience and ln(hours worked per week).
$F_1$: # of siblings, mother's education, and father's education.
$F_2$: father's education.
$A$: a composite of various test scores (ability).

Extensions: including median income where grew up, examining the impact of $Q$ by city size and ability, including college quality (expenditures and Gourman index), including 1955 earnings in the 1969 earnings equation, and a fuller model with occupation, religion, and additional family variables.

***Stated Findings***

This study examines the effects of both college and pre-college school quality on earnings.

The effect of \$/pupil is positive, but insignificant in the $S$ equation and positive and significant in the $A$ and $y$ equations. The rate-of-return to school spending is 12.6 percent.

All non-expenditure quality variables are significant in the earnings equation, except for the P/T ratio.

$Q$ has a larger impact in large cities and for those with lower values of $A$.

Pre-college expenditures are positively correlated with college quality and both pre-college quality and college quality have positive and significant effects on earnings.

$Q$ has a smaller impact on earnings earlier in the career. $Q$ and the rate of growth of earnings are positively correlated.

The coefficients on all three variables in the NRC data (school-level information) are positive and statistically significant.

## 18. Link and Ratledge (1976)

***Data***

Earnings Source: NBER-TH.

Earnings Sample: Excludes airline pilots, those in poor health, farm proprietors, if single, graduates of non-public high schools, and those with earnings >$75K or <$5K in 1969.

Quality Source: *Biennial Survey*, 1936-38.

Quality Variable: $/pupil at the district and state levels.

***Model***

$$y = X\beta + \delta S + \theta Q + A\varphi + \varepsilon$$

*y*: (log) 1969 annual earnings.
*S*: years of schooling.
*X*: experience.
*A*: IQ.

***Stated Findings***

This study is done in response to Akin and Kniesner's (1976) claim that it is unclear whether district-level or state-level data on *Q* contains less measurement error.

Both state-level and district-level expenditures have positive and significant effects on earnings.

The estimates using the district-level expenditures are higher with larger t-statistics than those using state-level expenditures. This is consistent with larger measurement error in the state-level data.

## 19. Wachtel (1976)

***Data***

Earnings Source: NBER-TH.

Earnings Sample: Excludes airplane pilots, those with no earnings or earnings >$75K, those in poor health, those without schooling variables, and private school attendees.

Quality Source: *Biennial Survey*, 1936-38 for pre-college measures and Office of Education data for post-secondary measures.

Quality Variables: $/pupil (elementary and high school) in the district attended (1958 $, adjusted using the deflator for state and local government purchases) and $/pupil in 1962-63 at the college attended (1958 $) for both undergraduate and graduate training.

***Model***

$$y = X\beta + \delta S + \theta Q + F\pi + A\varphi + \varepsilon$$

Estimates run by level of schooling (12,13-15,16,16+) and for all levels combined.

*y*: (log) 1969 annual earnings. (Also examines 1955 annual earnings.)

*S*: years of schooling (where levels can vary) and indicators for 12+ and 16+ years completed included in regressions with all levels combined.

*X*: experience, experience squared.

*F*: father's education.

*A*: an ability test score.

*Q*: pre-college, college, and graduate expenditures.

***Stated Findings***

This study examines the effects of pre-college, college and post-college school quality on earnings.

Pre-college, college and graduate spending all have positive and significant effects on earnings. The rates of return are 10-15 percent for pre-college expenditures, 10-19 percent for college expenditures, and 14 percent for graduate school expenditures.

## *D.* The Panel Survey of Income Dynamics (PSID) Matched to The Biennial Survey of Education

### 20. Morgan and Sirageldin (1968)

***Data***

Earnings Source: PSID. Described as a national sample of 1,525 family heads interviewed in 1965, although the source is never explicitly referenced by the authors.

Earnings Sample: Household heads with positive earnings in 1964 that attended school between 1910 and 1963 and had information for where grew up. Self-employed workers and those owning farms are excluded. This study includes both males and females and whites and minorities.

Quality Source: *Biennial Survey* (1929-30, 1939-40, and 1949-50). They also examine the Cass and Birnbaum college ranking for college quality. Not stated how college measures are mapped to workers.

Quality Variable: $/pupil in ADA averaged across the three decades above and deflated for price changes (deflator not described by authors).

***Model***

(1) $y = X\beta + \delta S + \varepsilon$

(2) $\bar{\hat{\varepsilon}} = \alpha + \theta Q + \upsilon$

The authors estimate equation (1) and then regress the average state residuals on *Q* in equation (2).

*y*: 1964 hourly earnings.

*X*: age, race, sex, and whether grew up on a farm.

*S*: years of schooling.

$\bar{\hat{\varepsilon}}$ : the average residual for each state, computed from individual's residuals in Stage 1.

***Stated Findings***

The coefficient on *Q* is positive and statistically significant with a rate-of-return of over 15 percent.

Findings could be due to omitted family variables.

There is also a positive correlation between college quality and earnings.

**21. Johnson and Stafford (1973)**

***Data***

Earnings Source: PSID. (Described as Survey Research Center data, 1968 Survey.)

Earnings Sample: Excludes African Americans, females, retirees, farmers, self-employed, those NILF, and those with no earnings in 1964.

Quality Source: *Biennial Survey*, 1957-58. Utilizes data for 1929-30, 1939-40, and 1949-50.

Quality Variables: $/pupil in ADA for elementary students (derived from the value for elementary and high school combined), deflated using the price deflator for state and local government expenditures.

***Model***

$y = X\beta + \delta S + \theta Q + F\pi + \varepsilon$

*y*: (log) 1964 hourly earnings.

*X*: experience (age-educ-7), experience squared, whether urban residence.

*S*: years of schooling.

*F*: whether grew up in an urban area.

*Q*: ln($/pupil).

***Stated Findings***

The coefficient on $Q$ is positive and statistically significant with an elasticity of 0.198 and rates-of-return between 11 percent and 21 percent.

The return to years of schooling doesn't change when $Q$ is added to the model.

The returns to school quality are higher than the returns to schooling.

Quality is also positively correlated with the amount of schooling attained.

**22. Akin and Garfinkel (1977)**

***Data***

Earnings Source: PSID conducted in 1972 (includes 1968-72 data).

Earnings Sample: Male household heads ages 30-55 in '72 that are not self-employed and that have earnings. Per capita income (PCY) by state is taken from the 1930-1960 Censuses. All monetary values ($y$, $Q$, and PCY) are deflated over time and by state.

Quality Sources: *Biennial Survey* for 1930, 1940, 1950, and 1960.

Quality Variable: \$/pupil (by race where available). Off census years are linearly interpolated and workers are assigned the value for the state where they lived when age 12.

***Model***

$$(1)\ A = Q\theta_1 + F_1\pi_1 + \varepsilon_1$$
$$(2)\ V = X_1\beta_1 + Q\theta_2 + F_1\pi_2 + A\varphi_2 + \varepsilon_2$$
$$(3)\ S = X_1\beta_3 + Q\theta_3 + F_2\pi_3 + A\varphi_3 + V\gamma_3 + \varepsilon_3$$
$$(4)\ y = X_2\beta_4 + S\delta_4 + Q\theta_4 + F_3\pi_4 + A\varphi_4 + V\gamma_4 + \tau_4 S * EXP + \varepsilon_4$$

All models are run separately for whites and blacks. Earnings models are run with and without PCY. Also examines the impact of using different deflators on earnings and adding the variables city size and area wage.

$y$: (log) the average hourly wage rate from 1968-72 (also examines average annual earnings).

$S$: years of schooling.

$X_1$: age.

$X_2$: experience EXP in (4); [age-S-6], experience squared.

$F_1$: father's income, father's education, whether middle income, whether upper income, # of siblings, whether father owned or operated a farm.

$F_2$: $F_1$ and whether grew up in a city.

$F_3$: $F_2$ excluding # of siblings.

$A$: an achievement orientation index score.

*V*: a test of verbal ability.
*Q*: ln($/pupil).

***Stated Findings***

$/pupil has a positive and significant effect in the *V*, *S*, and *y* equations and a positive, but insignificant effect in the *A* equation.

Including PCY reduces the impact of *Q*, becoming insignificant in all equations, except for the black earnings equation.

**23. Akin and Garfinkel (1980)**

***Data***

Earnings Source: PSID conducted in 1972 (includes 1968-72 data).

Earnings Sample: Male household heads ages 30 to 55 that are not self-employed and that have positive earnings. All monetary values (*y*, *Q*) are deflated over time and by state.

Quality Sources: *Biennial Survey* for 1930, 1940, 1950 and 1960.

Quality Variable: $/pupil (by race where available). Off census years are linearly interpolated and workers are assigned the value for the state where they lived when age 12.

***Model***

$$(AG) \quad y = X\beta_1 + \delta_1 S + \gamma_1 S * Exp + \theta_1 Q + \eta_1 SQ + F\pi_1 + A\varphi_1 + \varepsilon_1$$
$$(LRL) \quad y = X\beta_2 \qquad + \gamma_2 S * Exp \qquad + \eta_2 SQ + F\pi_2 + A\varphi_2 + \varepsilon_2$$

Regressions are run separately for whites and blacks.
*y*: (log) the average hourly wage rate, 1968-72.
*S*: years of schooling.
*X*: experience [minimum of (age-S-5) and (age-16)], experience squared.
*F*: father's income; grew up poor, middle class, or rich; father's education; grew up on farm.
*A*: IQ (from a sentence completion test); achievement motivational index.

***Stated Findings***

This article examines the "vintage effect" of *Q* (the change in the quality of black schools relative to white schools) on black/white wage ratios in response to Link, Ratledge, and Lewis (1976).

The coefficients on *Q* in (AG) and S*Q in (LRL) are positive and significant in both models; however, the *S*Q* interaction coefficients in (AG) are negative and insignificant. The negative interaction term is not large enough to offset the positive levels effect for all practical levels of schooling.

## E. The National Longitudinal Survey of Youth 1997 (NLSY79)[60]

### 24. Betts (1995)

***Data***

Earnings Source: NLSY79 conducted in 1990 (includes observations from 1979 through 1990).

Earnings Sample: Includes white males ages 17 or more with positive earnings, who attended public high school, living in the 50 states or D.C., and that have non-missing values for wage, race and sex. The military sub-sample is excluded. (This study also examines various sub-populations.)

Quality Source: NLSY79 (Data is for the most recently attended HS if 17 or older in 1979, matched by school to those under 17 in 1979 where possible.)

Quality Variables: T/P ratio, starting teachers' wages for those with a B.S. (relative to per capita earnings in the state), and percent of teachers with a graduate degree.

***Model***

$$y = X\beta + \delta S + \Gamma S + \theta Q + \eta SQ + \varepsilon$$

*y*: (log) weekly wages (also examined hourly and annual with similar results, not reported).

*S*: years of schooling; Γ represents nine region-of-residence indicators.

*X*: experience (quartic), SMSA, marital status, region-of-residence and year indicators.

***Stated Findings***

When school indicators are included (with *Q* omitted), they are jointly significant; schools matter.

However, the quality measures are not significant in the intercepts (not interacted with other variables) or when interacted with years of schooling.

These results are robust within various sub-samples of the population, in reduced-form estimates (without years of schooling), and when additional quality measures are included.

When state-level measures (from the *Biennial Survey*) are included, the coefficient on the T/P ratio is positive and significant and the coefficient on teachers' wages is positive and insignificant.

---

[60] Betts (1996) uses both the NLSY79 and the U.S. Census.

**25. Betts (1996)**

***Data***

Earnings Source 1: 1970 and 1980 Census.

Earnings Sample 1: Same as in Card/Krueger (1992a) except that the age restrictions vary across calculations. The 1970 Census restrictions parallel the 1980 Census restrictions with real wage adjustments calculated using the CPI.

Earnings Source 2: NLSY79.

Earnings Sample 2: Same as Betts (1995).

Quality Source: *Biennial Survey*, *Digest of Education*, and the NLSY79.

Quality Variables: T/P ratio, teachers' wages relative to teachers' wages in the census region, and percent of teachers with a graduate degree (from the NLSY79). T/P ratio (attendance based), T/P ratio (enrollment based), term length, and relative teacher's salary (from the *Biennial Survey* and the *Digest of Education*).

***Model***

(1) $y = X_1\beta_1 + \varepsilon$

(2) $\lambda = X_2\beta_2 + \theta Q + \eta QS + S\Gamma\delta + \upsilon$

Stage (1) is run for white males ages 40 to 55 using the 1980 Census. Stage (2) is run using the NLSY79 sample with the coefficients on the occupation indicators ($\lambda$'s) from Stage (1) matched to the NLSY79 respondents given their occupation. The second stage regression weights observations by the square root of the sampling variance of the occupation coefficients.

$y$: (log) 1979 weekly wages.

$S$: years of schooling. $\Gamma$ represents region of residence indicators.

$X_1$: age, age squared, marital status, SMSA, 503 occupational indicators.

$X_2$: experience (quartic), marital status, SMSA, and census region.

$Q$: T/P ratio, relative teacher salary, percent of teachers with a graduate degree.

Extensions:

1) Estimates Stage (1) for each occupation including age, marital status, and SMSA in $X_1$. The age coefficients are then substituted for the occupational indicators in Stage (2).

2) Stage (2) is estimated substituting the Duncan socioeconomic index for each respondent's occupation ($\lambda$) in Stage (2).

Betts also re-estimates the Card/Krueger model using their main specification and the 1980 Census data with the age range extended to 20 to 59 so as to overlap the ages of respondents in the NLSY79. In Stage (2) of their procedure, Betts includes age (in five-year cohort bands, 20-24, 25-29, …) by $Q$ interactions and tests whether the coefficients are significantly different for younger workers. Extensions to this model examine the 1970 Census alongside the 1980 Census and tests for differences in the return to education within cohorts between these two years.

***Stated Findings***

Most of the studies that fail to find that school spending has a significant impact on earnings examine younger workers; those that find that school spending has a significant impact on earnings examine samples that include older workers. If the impact of school spending only manifests itself later in the work career, this would explain the discrepancy in findings.

School inputs are not statistically significant predictors of mid-career earnings estimated by occupational differentials; they are not statistically significant predictors of the age-earnings profile within workers' chosen occupations; and they are not strongly related to a workers' occupational status (Duncan index).

Using Census data and the Card/Krueger model, the effects of school quality are not weaker for younger workers.

"… age dependence [of the impact of school inputs] is not the main explanation for the divergent results in the literature on the link between school resources and earnings."

**26. Griffin and Ganderton (1996)**

***Data***

Earnings Source: NLSY79 conducted in 1990.

Earnings Sample: Excluded if hours per week<20; no earnings; hourly wages<$2; not white, black or Hispanic; or missing $Q$, $F$, or $A$.

Quality Source: NLSY79.

Quality Variable: A composite is constructed which includes the number of library books, T/P ratio, counselor/pupil ratio, dropout and attendance rates, and the characteristics and salaries of teachers. (Construction not described by the authors.)

***Model***

$$y = X\beta + \delta S + \theta Q + \pi F + \varphi A + \varepsilon$$

Model estimated separately for each race and for all races combined.

$y$: (log) hourly wages.

$S$: years of schooling.

*X*: experience, experience squared, marital status, urban and regional indicators.

*F*: father's and mother's education, whether single parent family, # of siblings, reading composite (newspapers, magazines, library card).

*A*: AFQT score adjusted for age and then for differences in *Q* and *F*.

***Stated Findings***

*Q* does not have a significant, direct effect on earnings. However, in other regressions they show that "school quality is an important determinant of ability, and ability is an important determinant of earnings." Therefore, *Q* matters through its impact on *A*.

## 27. Strayer (2002)

***Data***

Earnings Source: NLSY79. (Includes data from 1979 to 1994.)

Earnings Sample: Excludes the military sub-sample, high school dropouts, and those enrolled in school.

Quality Source: NLSY79 for high school characteristics and uses IPEDS for college characteristics (only used to identify two and four year colleges).

Quality Variables: percent of teachers with a graduate degree, P/T ratio, availability of technical programs, and availability of agricultural programs.

***Model***

$$(MNL)\ \ p = f(X, Q, F, A)$$

$$(OLS)\ \ y = X\beta + \delta S + \theta Q + \varepsilon$$

(*MNL*) estimates the probability of attending no college, a 2-year college, or a 4-year college using a multinomial logit and (*OLS*) estimates separate wage equations using *OLS* for those with no college, those with 2 years of college and those with 4 years of college.

*MNL:*

*X*: sex, race, marital status, 2-year college tuition, 4-year college tuition, year indicators.

*F*: family income, mother's education, father's education, # of siblings, and whether foreign born.

*A*: AFQT percentile score.

*Q*: quality variables plus attendance rate and dropout rate of high school.

*OLS:*

*y*: (log) hourly wages.

*S*: highest grade completed.

*X*: race, sex, marital status, region (4 values), SMSA, PT/FT, union, experience (quartic), year and industry indicators, and the selection parameters from (*MNL*).

*Q*: quality variables plus the P/T ratio squared.

***Stated Findings***

*Q* has a positive and significant effect on the probability of college attendance and on the type of college attended. College choice affects post-school earnings.

The direct effect of *Q* on earnings is positive but insignificant.

"The results suggest that high school quality influences earnings by affecting college choice behavior, while the direct effect of school quality on earnings is less evident."

## F. High School and Beyond (HSB).[61]

### 28. Grogger (1996a)

***Data***

Earnings Source: HSB. Earnings in 1986 (seniors), 1988 (sophomores), and 1991 (sophomores).

Earnings Sample: The senior sample is restricted to full-time workers with positive earnings in 1986 and hourly wages between \$1 and \$100 per hour. The sophomore sample is restricted to those with monthly earnings between \$500 and \$6,000 and with at least nine months of employment. Those with values of \$/pupil that seem too low (<\$200/pupil) are also excluded.

Quality Source: HSB and NCES.

Quality Variable: \$/pupil in 1980 and 1982 in the district attended (HSB). State-level \$/pupil in 1979-80 (NCES).

***Model***

$$y = X\beta + Q\theta + F\pi + \varepsilon$$

Regressions run separately for seniors and sophomores and for each year for sophomores.

*y*: (log) hourly wages (seniors) and (log) monthly wages (sophomores).

*X*: race, census division where school located, school type (urban, suburban, or rural), whether school had an alternative-curriculum, whether school is predominantly Hispanic.

---

[61] Grogger (JOLE 1996b) uses HSB as well as the National Longitudinal Survey of the High School Class of 1972 (NLS72).

$F$: lived with father, # of siblings, father's education, family income.
$Q$: ln($/pupil).
Years of schooling and experience are excluded from the model.

***Stated Findings***

Regressions including district-level expenditures are, in general, positive but insignificant. Estimates using state-level expenditures are generally higher, but still not significant.

IV estimates (using the other year's district-level $/pupil and, separately, state-level $/pupil) are higher than the OLS estimates and are significant in some of the comparisons. This indicates that measurement error may weaken estimates that use district-level expenditures.

State-of-birth indicators in regressions that include district-level estimates are jointly significant. Omitted state effects are important and may bias state-level and state-instrumented estimates.

"School spending matters, but it matters too little."

**29. Grogger (1996b)**

***Data***

Earnings Source: HSB (senior cohort) and NLS72. Uses 1986 wages for HSB and 1979 wages for NLS72.

Earnings Sample: In the original survey and the follow-up, full-time, not enrolled, and with hourly wages between $1 and $100 (in 86$). There are no high school dropouts in these samples.

Quality Source: NLS72 and HSB (Data collected from school administrators' offices).

Quality Variables: P/T ratio, term length, whether 30 percent of teachers have a graduate degree, school size, and percent black.

***Model***

$$y = X\beta + \delta S + \theta Q + \pi F + \varphi A + \varepsilon$$

Regressions run separately for NLS72 and HSB samples.

$y$: (log) hourly earnings.
$S$: indicators for high school degree, college degree, and graduate degree.
$X$: experience, experience squared, race, region, and urban.
$F$: family income.
$A$: test scores and grades.

***Stated Findings***

The coefficients on P/T ratio and term length are insignificant whereas teacher education (NLS72 only), school size, and percent black are significant. The effects of school size are small.

School indicators are jointly significant and greatly increase the explanatory power of the model (with *Q* omitted). Including school indicators reduces the black/white wage differential by 17 percent in the 1979 data, but increase it by 38 percent in the 1986 data.

Schools matter, but school characteristics do not explain these effects.

**30. Rivkin (2000)**

***Data***

Earnings Source: HSB (sophomore cohort).

Earnings Sample: Includes black men and women from public schools in large, urban districts with at least five non-black students. Excludes those with no earnings.

Quality Source: HSB.

Quality Variables: P/T ratio and a value added measure constructed from an auxiliary regression of a 12th grade test score on the 10th grade test score, individual and family characteristics and school indicators. The school indicators are the value added measure.

***Model***

$$y = X\beta + \theta Q + F\pi + A\varphi + \varepsilon$$

Model estimated with and without *A*.

*y*: (log) 1991 monthly earnings.

*X*: gender, public in-state tuition and unemployment rate in the year of graduation.

*F*: parental education, family income, and whether the school is located in the South.

*A*: a composite of reading and math scores in the 10th grade.

*Q*: percent of students who are Hispanic, percent of students who are white, P/T ratio, value added measure.

***Stated Findings***

The P/T ratio and the school average value added measures are positive and significant in models both with *A* and without *A* included.

In regressions that examine determinants of schooling level completed, the coefficient on the value added measure was positive and significant and the coefficient on the P/T ratio was positive but insignificant.

$Q$ had a smaller impact in districts where desegregation was involuntary.

Miscellaneous Data Sources (Two or Fewer Studies Using a Particular Source): The Postcensal Survey of Professionals (1962), 1968 Urban Problems Survey, Project Talent, and Minnesota Twin Registry Data.

**31. Link (1973)**

***Data***

Earnings Source: Postcensal Survey of Professional and Technical Manpower conducted in 1962 (of persons included in the 1960 Census).

Earnings Sample: Currently employed, male chemical engineers with some college training. Professors are excluded.

Quality Source: Engineers' Council for Professional Development and Astin's selectivity index.

Quality Variables: Whether the chemical engineering program or the engineering department is accredited and Astin's selectivity index.

***Model***

$$y = X\beta + \delta S + \theta Q + F\pi + A\varphi + \varepsilon$$

$y$: 1961 annual earnings (excludes consulting earnings).

$X$: age, marital status, employer type (gov't, private, self-employed), region, whether received formal training, years in present field, and whether published.

$S$: indicators for 12-15 years, 16 years with no degree, bachelor's degree, some graduate study (omitted category), master's degree, and Ph.d.

$F$: type of residence when a youth, father's occupation, and size of high school graduating class (results not reported).

$A$: Astin's intellectualism index.

***Stated Findings***

This study examines the effects of college quality on earnings.

The coefficient on $Q$ is positive and statistically significant when $A$ is excluded from the model.

However, when Astin's intellectualism index ("a proxy for student ability and motivation") and $Q$ are simultaneously included in the model, $A$ is positive and significant whereas the coefficient on $Q$ is positive and insignificant. Higher earnings are primarily due to ability differences and not school quality differences.

**32. Link (1975)**

***Data***

Earnings Source: Postcensal Survey of Professional and Technical Manpower conducted in 1962 (of persons included in the 1960 Census).

Earnings Sample: Male electrical engineers with some graduate schooling, excluding professors.

Quality Source: Allan M. Carter's index (of faculty rankings) of the quality of graduate education in the U.S. categorized into five groups (distinguished and strong, good, adequate, marginal with no Ph.d. program, and not rated).

***Model***

$$y = X\beta + \delta S + \theta Q + F\pi + A\varphi + \varepsilon$$

*y*: 1961 annual earnings.

*X*: marital status, employer type (private, gov't, self), employment status (with job, but not at work), region of residence (4 values), whether published, whether any formal training, years in field, years with firm, age, whether published.

*S*: indicators for master's degree and Ph.d.

*F*: father's occupation and residence as a youth.

*A*: Astin's intellectualism index (3 categories).

***Stated Findings***

This study examines the effects of college quality on earnings.

The two highest quintile categories of *Q* are individually statistically significant and all quintiles of $\varphi$ are jointly significant.

The effects of higher quality diminish with age.

The typical estimates of the return-to-schooling are overstated by 25-33 percent when *Q* is omitted.

**33. Morgenstern (1973)**

***Data***

Earnings Source: 1968 Urban Problems Survey conducted by the SRC. Survey conducted in 15 largely northern cities and two suburbs.

Earnings Sample: Excludes self-employed, NILF, and non-respondents.

Quality Source: Source not mentioned, but data is matched by state and decade, suggesting the *Biennial Survey*.

Quality Variables: $/pupil (deflated by the national average for the year). In unreported results, also examines P/T ratio and teacher salaries.

***Model***

(1) $y = X\beta_1 + \delta S + \theta_1 Q + F_1\pi_1 + \varepsilon_1$ $(direct\ effect)$

(2) $S = X\beta_2 \qquad + \theta_2 Q + F_2\pi_2 + \varepsilon_2$ $(indirect\ effect)$

(3) $y = X\beta_2 \qquad + \theta_3 Q + F_2\pi_3 + \varepsilon_3$ $(reduced\ form)$

Regressions are run separately for blacks and whites.
$y$: hourly wage rates.
$X$: indicators for gender, experience, and whether living in the South.
$S$: indicators for 0-7, 8-11, 13-15, and 16+ years of schooling.
$F_1$: whether father had low education.
$F_2$: $F_1$ and Duncan index of father's occupation.

***Stated Findings***

In equation 1, the coefficient on $Q$ is significant for blacks (10 percent rate of return), but not for whites.

In equation 2, the coefficient on $Q$ is positively correlated and significant for whites and blacks and the impact is larger for blacks.

In equation 3, the coefficient on $Q$ is positive and significant and has a larger impact on whites' earnings.

"… this study … finds that, especially for blacks, the quality of education is economically important."

**34. Ribich and Murphy (1975)**

***Data***

Earnings Source: Project Talent (interview of 9th grade males in 1959 with a follow-up in 1968).

Earnings Sample: Males attending public schools that are not on active military duty and that answered the education and occupation questions. If still in school, they answered the degree plan and course of study questions. Includes several extensions on sub-populations.

Quality Source: Project Talent.

Quality Variable: \$/pupil in the system (elementary and high school) where educated in the 9th grade.

***Model***

(1) $A = X\beta_1 \qquad + \theta_1 Q + F\pi_1 \qquad + \varepsilon_1$

(2) $S = X\beta_2 \qquad + \theta_2 Q + F\pi_2 + A\varphi_2 + \varepsilon_2$

(3) $y = X\beta_3 + \delta_3 S + \theta_3 Q + F\pi_3 + A\varphi_3 + \varepsilon_3$

$y$: final adjusted lifetime earnings (predicted from 1967 annual earnings).

*X*: race, South.

*S*: years of schooling (completed or expected, depending on enrollment status).

*F*: socioeconomic index (contained in the data) and average socioeconomic index of 9th graders in the same school.

*A*: TAFQT (similar to AFQT).

***Stated Findings***

Although the coefficient on *Q* is insignificant in the *A* and *y* equations, it is positive and significant in the *S* equation. Therefore, "the chief effect of spending differences on lifetime income is found to work through this school continuation link."

The rate-of-return to increased educational spending (*Q*), however, is negative.

**35. Behrman, Rosenzweig and Taubman (1996)**

***Data***

Earnings Source: Minnesota Twin Registry survey conducted in 1994.

Earnings Sample: MZ and DZ female twins with earnings data (the authors use the last available, real wage). Data includes the names of colleges attended by the respondents.

Quality Source: CASPAR and Barron's Guides to 2- and 4-year colleges.

Quality Variables: \$/student, full-time equivalent enrolled students, students/faculty, whether grants Ph.d.'s, whether public or private, and mean salaries of full professors. Specifications also include an indicator for absent faculty salary along with all other college characteristics (results not reported by the authors).

***Model***

$$y = X\beta + S\delta + Q\theta + \varepsilon$$

Model estimated for (1) DZ twins, (2) within DZ twin pairs, (3) within MZ twin pairs, and (4) a BRT model using DZ and MZ twins. (Not clear what the BRT model is, but it has 179 variables.)

*y*: (log) full-time earnings.

*S*: years of schooling.

*X*: full-time experience.

***Stated Findings***

Examines the effect of college quality on earnings.

Attendance at higher quality colleges leads to higher earnings. "The statistically-preferred estimates suggest that Ph.d.-granting, private universities with well-paid senior faculty and smaller enrollments produce students who have significantly higher earnings later in life."

Those with higher individual and family endowments attend higher quality colleges.

## APPENDIX B
## GRADIENT DECOMPOSITION PROOF

Let $y_i = \alpha + \beta s_i + \varepsilon_i$, where i = 1, 2, … , n identifies an individual, *y* is (log) earnings, *s* is level of schooling, and $\varepsilon$ is the error term. $\alpha$ and $\beta$ are parameters, with the OLS estimate of $\beta$ being

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(s_i - \overline{s})(y_i - \overline{y})}{\sum_{i=1}^{n}(s_i - \overline{s})^2} \tag{B.1}$$

Let j = 0, 1, 2, … , J represent J+1 discrete levels of schooling. The estimate can be rewritten as,

$$\hat{\beta} = \frac{\sum_{j=0}^{J}\sum_{i\in j}(s_i - \overline{s})(y_i - \overline{y})}{\sum_{j=0}^{J}\sum_{i\in j}(s_i - \overline{s})^2} = \frac{\sum_{j=0}^{J}(s_j - \overline{s})\sum_{i\in j}(y_i - \overline{y})}{\sum_{j=0}^{J}N_j(s_j - \overline{s})^2} = \frac{\sum_{j=0}^{J}(s_j - \overline{s})N_j(\overline{y}_j - \overline{y})}{\sum_{j=0}^{J}N_j(s_j - \overline{s})^2} \tag{B.2}$$

where $N_j$ is the number of individuals and $\overline{y_j}$ is the average earnings of workers with the j[th] level of schooling. This expression can be simplified by noting that

$\overline{y}\sum_{j=0}^{J}(s_j - \overline{s})N_j = 0$ since $\overline{s} = \frac{\sum_{j=0}^{J}N_j s_j}{\sum_{j=0}^{J}N_j}$. Therefore,

$$\widehat{\beta} = \frac{\sum_{j=0}^{J}(s_j - \overline{s})N_j\overline{y}_j}{\sum_{j=0}^{J}N_j(s_j - \overline{s})^2} \tag{B.3}$$

Let,

$$\begin{aligned}
d_j &= N_j(s_j - \overline{s}) && j = 0,\ldots,J \\
s'_j &= s_j - s_{j-1} && j = 1,\ldots,J \\
\tilde{y}_j &= \frac{\overline{y}_j - \overline{y}_{j-1}}{s'_j} && j = 1,\ldots,J \\
&\textit{and } \tilde{y}_0 = \overline{y}_0
\end{aligned}$$

where $s'_j$ is the difference between adjacent schooling levels and $\tilde{y}_j$ is the increase (we hope) in average earnings given an increase in schooling by one level.

By definition,

$$\begin{aligned}
\overline{y}_j &= (\overline{y}_j - \overline{y}_{j-1}) + (\overline{y}_{j-1} - \overline{y}_{j-2}) + \cdots + (\overline{y}_1 - \overline{y}_0) + \overline{y}_0 \\
&= \frac{s'_j(\overline{y}_j - \overline{y}_{j-1})}{s'_j} + \frac{s'_{j-1}(\overline{y}_{j-1} - \overline{y}_{j-2})}{s'_{j-1}} + \cdots + \frac{s'_1(\overline{y}_1 - \overline{y}_0)}{s'_1} + \overline{y}_0 \\
&= s'_j\tilde{y}_j + s'_{j-1}\tilde{y}_{j-1} + \cdots + s'_1\tilde{y}_1 + \tilde{y}_0
\end{aligned}$$

The numerator of equation (B.3) becomes

$$\sum_{j=0}^{J}(s_j - \overline{s})N_j\overline{y}_j = \sum_{j=0}^{J}d_j(s'_j\tilde{y}_j + s'_{j-1}\tilde{y}_{j-1} + \cdots + \tilde{y}_0) \tag{B.4}$$

Expanding (B.4),

$$d_0\tilde{y}_0 + d_1(s_1'\tilde{y}_1 + \tilde{y}_0) + \cdots + d_{J-1}(s_{J-1}'\tilde{y}_{J-1} + s_{J-2}'\tilde{y}_{J-2} + \cdots + \tilde{y}_0)$$
$$+ d_J(s_J'\tilde{y}_J + s_{J-1}'\tilde{y}_{J-1} + \cdots + \tilde{y}_0)$$

and collecting terms on $y_j'$

$$s_J'\tilde{y}_J(d_J) + s_{J-1}'\tilde{y}_{J-1}(d_J + d_{J-1}) + \cdots + s_1'\tilde{y}_1(d_J + d_{J-1} + \cdots + d_1)$$
$$+ \tilde{y}_0(d_J + d_{J-1} + \cdots + d_1 + d_0)$$

The last term falls out of the equation since $\sum_{j=0}^{J} d_j = \sum_{j=0}^{J} N_j(s_j - \overline{s}) = 0$. Now,

define $\omega_i = \dfrac{\sum_{k \geq j}^{J} d_k s_i'}{\sum_{j=0}^{J} N_j(s_j - \overline{s})^2} = \dfrac{s_i' \sum_{k \geq j}^{J} d_k}{\sum_{j=0}^{J} N_j(s_j - \overline{s})^2}$ and substitute this expression into the equation for $\hat{\beta}$

$$\hat{\beta} = \sum_{j=1}^{J} \omega_j \tilde{y}_j$$

In words, the OLS estimate of $\beta$ is a weighted average of the change in earnings relative to the increase in years of schooling between adjacent schooling levels summed over all levels of schooling above the initial value.

The terms $s_j'$ and $\sum_{j=0}^{J} N_j(s_j - \overline{s})^2$ are always positive, so the sign of $\omega_j$ depends only on the sign of $\sum_{k \geq j}^{J} d_k$. Recalling that $d_j = N_j(s_j - \overline{s})$, it is clear that for all terms $s_j > \overline{s}$, $d_j > 0$ and so the weights for above average levels of schooling must be positive as they include only positive terms. Each term with $s_j < \overline{s}$ reduces the sum $\sum_{k \geq j}^{J} d_k$ ; however, only at *j* = 0 does the sum reach zero and so it must be non-negative for all value of *j*>0. Therefore, the weights, $\omega_j$ , are non-negative for all values of *j*.

# APPENDIX C
# NOTE ON THE COMPUTATION OF THE STATE-SPECIFIC COLLEGE/HIGH SCHOOL GRADUATE WAGE GRADIENTS REPORTED IN CHAPTER II (FN. 2) AND CHAPTER V

The sample is drawn from the 2000 U.S. Census of Housing and Population, PUMS 5-percent file and includes white, non-Hispanic, males ages 20 to 64, born in the United States and living in the United States in 1995, not enrolled in school, not serving in the armed forces, and not living in group quarters. Those living in Hawaii and Alaska in either 1995 or in 2000 are excluded. Those living in the District of Columbia are combined with those living in Virginia due to sample size. It is (log) weekly wage and salary earnings in 1999 that are examined in computing schooling gradients. Those with no weeks worked in 1999 and those with no wage and salary earnings are excluded from the sample. Those with more than $1,000 in self-employment income or whose self-employment income is more than 10% of all earned (wage and salary and self-employment) income are excluded. Those with allocated values for race, sex, age, Hispanic origin, state-of-birth, state-of-residence in 1995, educational attainment, marital status, military or employment status, wage and salary income, or self-employment income are excluded. The remaining sample includes 1,273,999 men.

Further, weekly wages are bottom coded at the first percentile and top coded at the ninety-ninth percentile within residence state. This procedure removes very large and very small values of weekly earnings from the sample without removing workers with high or low values from the estimates. Marital status is categorized into three groups: married and living with spouse; ever married but not currently living with spouse (includes widowed, separated, divorced, and married with spouse not present), and never married. Four birth cohorts are defined by age: 29-29, 30-39, 40-49, and 50-64. Potential experience is the minimum of age minus years of schooling minus six and age minus sixteen with values below zero set to zero. Those with less than nine years of schooling are combined with those completing the ninth grade. This was done because the sample becomes small in some states and the estimates become unreliable.

Given these sample and variable specifications, state-specific schooling gradients are estimated by regressing (log) weekly earnings on years of schooling indicators, cohort by marital status indicators, and a quartic in experience. The estimated college/high school wage premia are reported by state in Table C.1.

**Table C.1**
**State Specific Schooling Gradients, the Return to a College Degree Over a High School Degree**

| State | College Gradient | State | College Gradient | State | College Gradient | State | College Gradient |
|---|---|---|---|---|---|---|---|
| WY | 0.2365 | NE | 0.3784 | NH | 0.4494 | MD | 0.5132 |
| VT | 0.2961 | ID | 0.3809 | MI | 0.4526 | NJ | 0.5132 |
| SD | 0.2973 | IN | 0.4020 | DE | 0.4581 | TN | 0.5134 |
| MT | 0.3122 | WA | 0.4065 | SC | 0.4587 | MA | 0.5135 |
| ND | 0.3195 | OR | 0.4073 | OH | 0.4612 | IL | 0.5175 |
| LA | 0.3369 | MN | 0.4171 | MO | 0.4625 | NC | 0.5197 |
| WI | 0.3640 | NM | 0.4176 | AR | 0.4767 | CT | 0.5316 |
| IA | 0.3665 | OK | 0.4327 | AZ | 0.4976 | TX | 0.5602 |
| UT | 0.3674 | RI | 0.4361 | AL | 0.4978 | NY | 0.5722 |
| NV | 0.3754 | CO | 0.4391 | KY | 0.4981 | CA | 0.5727 |
| ME | 0.3761 | KS | 0.4464 | PA | 0.4989 | GA | 0.5820 |
| MS | 0.3762 | WV | 0.4489 | FL | 0.5070 | VA | 0.5987 |

SOURCE.-2000 Census of Population and Housing, PUMS 5-Percent Sample.
NOTE.-See the Text.

# VITA

## ROBERT B. SPEAKMAN, JR.

### BUSINESS ADDRESS

Welch Consulting
1716 Briarcrest Dr., #700
Bryan, Texas 77802
(979) 846-0303

### EDUCATION

B.A., Economics, Brigham Young University, 1990.
Ph.D., Economics, Texas A&M University, 2006.

### RESEARCH AND ACADEMIC EXPERIENCE

Economist, Welch Consulting, College Station, Texas, 1995-Present.

Associate Economist, Welch Consulting, College Station, Texas, 1993-1995.

Research Assistant, Texas A&M University, College Station, Texas, 1994-1995.

Instructor, Intermediate Microeconomic Theory, Texas A&M University, College Station, Texas, 1994.

Instructor, Principles of Microeconomics, Texas A&M University, College Station, Texas, 1993-1994.

Research Assistant, Texas A&M University, College Station, Texas, 1991-1993.

### HONORS AND AWARDS

Texas A&M University Academic Fellowship

### WORKING PAPERS

"Using Wages to Infer School Quality," with Finis Welch. In *Handbook of the Economics of Education*, edited by Eric Hanushek and Finis Welch. Elsevier Science Ltd. (Forthcoming 2007).

"Does School Quality Matter? A Reassessment," with Finis Welch, Unpublished. December 1994.