9-2009

Skill Mismatch and Wage Inequality in the U.S.

Fabian Slonimczyk
University of Massachusetts Amherst, fabian.slonimczyk@gmail.com

SKILL MISMATCH AND WAGE INEQUALITY IN THE U.S.

A Dissertation Presented

by

FABIÁN SLONIMCZYK

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2009

Department of Economics

SKILL MISMATCH AND WAGE INEQUALITY IN THE U.S.

A Dissertation Presented

by

FABIÁN SLONIMCZYK

Approved as to style and content by:

____

Peter Skott, Chair

____

Robert Pollin, Member

____

James Heintz, Member

____

Diane Flaherty, Department Chair
Department of Economics

To those that came before me

ACKNOWLEDGMENTS

Were it not customary to begin by thanking one's advisor, I would nevertheless start that way. For the last three years, Prof. Skott has not only been my mentor and critic. He has also been a good friend. I have been very fortunate to have had the chance to work (and run) with Peter. Bob Pollin, James Heintz and the rest of PERI have been supportive of my work from the beginning. I will certainly miss running into them at Gordon Hall. My family in Argentina and Uruguay has been a constant source of support. It has been hard to be away from them, as well as from all other things I had known until I moved to the United States. The strong foundation they provided me with has been invaluable in learning how to do things my own way. Special thanks to some of the people I have lived with during my stay in the Pioneer Valley: Arjun Jayadev, Daniel Esteban, Uri Strauss, Martin Rapetti and Alyssa Schneebaum.

ABSTRACT

SKILL MISMATCH AND WAGE INEQUALITY IN THE U.S.

SEPTEMBER 2009

FABIÁN SLONIMCZYK

B.Sc., UNIVERSITY OF BUENOS AIRES

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Peter Skott

This dissertation is an empirical investigation into the distributive effects of over- and under-education, defined as market outcomes such that some workers possess skills over or below those required at their jobs respectively. This type of market failure can arise in assignment and search equilibrium settings, as well as in the presence of asymmetric information regarding workers' performance on the job. The existence of permanent and sizable mismatch rates means that returns to education are depressed for over-educated workers and inflated for under-qualified workers. Thus, irreversible decisions to invest in human capital are made in a context of uncertainty regarding the exact outcomes that might arise. As in the Todaro model, where individuals decide whether to migrate to cities based on the expected values of the available alternatives, workers might decide it is worthwhile to keep investing in education even if the probability of finding appropriate employment is falling. The three chapters of the dissertation are entitled: "Skill Mismatch and Earnings: A Panel analysis of the U.S. Labor Market," "Earnings Inequality and Skill Mismatch in the U.S:

1973–2003," and "Employment and Distribution Effects of Changes in the Minimum Wage."

Skill Mismatch and Earnings: A Panel analysis of the U.S. Labor Market

This chapter examines the effect on earnings induced by a mismatch between workers' skills and the skills actually required on the job. It uses the Current Population Survey (CPS) for the period 1983–2002. The special re-interview methodology of the CPS is used to create a large panel, so that individual heterogeneity can be controlled for. Skill requirements are estimated by the median education level for each 3-digit occupation in the 1980 census occupational classification. The analysis, including the determination of skill requirements, is conducted for males and females separately. Cross-sectional analysis confirms the findings in the recent literature. Returns to required schooling are higher than the returns to attained education in standard earnings regressions. Also, for workers with similar educational attainment, over-education reduces earnings and under-education increases them. Contrary to what other studies have found, we conclude that these results are confirmed after controlling for individual fixed effects. The chapter also investigates which groups are more exposed to mismatch. I use standard probit analysis with over-education and under-education as the respective dependent variables. Women, service sector, and non-unionized workers appear to have higher probabilities of mismatch.

Earnings Inequality and Skill Mismatch

This chapter shows that skill mismatch is a significant source of inequality in real earnings in the U.S. and that a substantial fraction of the increase in wage dispersion during the period 1973–2002 was due to the increase in mismatch rates and mismatch premia. Standard human capital earnings regressions that do not decompose the education variable into required, surplus, and deficit years provide biased estimates of the relative importance of education in explaining earnings inequality. In 2000–2002 surplus and deficit qualifications taken together accounted for 4.3 and 4.6 percent of the variance in earnings, or around 15 percent of the total explained variance. The dramatic increase in over-education rates and premia accounts for around 11 and 32 percent of the increase in the coefficient of variation of log earnings during the 30 years under analysis for males and females respectively. Residual inequality is slightly diminished when the estimating equation allows the prices of surplus, required and deficit qualifications to differ but the well-studied increasing trend of within-group inequality remains otherwise unchanged. Changes in the composition of the labor force are found to be important predictors of increasing residual inequality even when skill mismatch is taken into account.

The Distributive Effects of the Minimum Wage: an Efficiency Wage Model with Skill Mismatch (co-authored with Peter Skott)
This chapter analyzes the effect of changes in the real value of the minimum wage on the wage distribution. Changes in the minimum wage and other labor market institutions affect workers in all groups and empirically appear to be good complement to standard supply and demand arguments in explaining overall inequality. We use an efficiency wage model but allow for mismatch between jobs and workers. This framework yields predictions not only on the skill premium but also on the extent of inequality within groups. To keep matters as simple as possible, we assume that high-skill workers can get two types of jobs (good and bad), whereas low-skill workers have only one type of employment opportunity (bad). As long as some matches of high-skill workers and bad jobs are sustained in equilibrium, changes in the exogenous variables will affect not only wages and employment rates but also the degree of mismatch. Thus, this paper shows that 'over-education' can be generated endogenously in efficiency wage models and that a fall in the real value of the minimum wage can (i) reduce total employment, (ii) lead to a simultaneous decline in both the relative employment and the relative wage of low-skill workers, and (iii) produce a rise in

within-group as well as between-group inequality. Evidence from the US suggests that these theoretical results are empirically relevant.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

Table **Page**

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

One of the most remarkable stylized facts about labor markets in developed economies in the last half century is the trend toward the increase in the average skill of workers. Every year an immense amount of resources are dedicated to increasing those skills, what Gary Becker and others call investments in human capital. In particular, an ever larger fraction of the American labor force has chosen to participate in college education. This dissertation delves into the question of whether a significant fraction of the resources invested in human capital formation might not be fully utilized and what the consequences are thereof.

I refer to situations in which a significant fraction of employed workers possess skills above or below those required at their jobs as over- and under-education, respectively. Such market outcomes are not what one would expect in a system with full price flexibility and perfect information and are probably undesirable in and by themselves. Over-education in particular not only implies that resources have been wasted in surplus qualifications but also that, in all probability, worker satisfaction is low. Investments in human capital surely have a pecuniary aspect. Underemployment of skills is bad in this sense because surplus qualifications are remunerated very poorly at best. Perhaps in a more fundamental sense over-education is undesirable because it implies many people's expectations and aspirations are being disappointed. In surveys of job quality, workers usually rank nonpecuniary aspects such as autonomy on the job, time self-management, and appropriate career prospects higher than monetary rewards.[1] Thus, earnings differentials between correctly matched and mismatched workers with the same level of acquired skill might underestimate the true cost of over-education.

While the increasing supply of skills brought to the market by workers is an undisputed fact, there has been a long-running debate in labor economics and sociology regarding the effect of capitalist development on the skill content of jobs. One view sees technological progress as mostly driven by employers' aim to better control the production process in order to increase their share in the distribution of income. For example, Marglin (1974) argued that neither the minute division of labor of the old putting-out system nor the centralized organization of the factory system were introduced primarily due to their technical superiority. These innovations were meant to change the strategic balance of power between capitalists and workers in favor of the former. The deskilling hypothesis of Braverman (1974) further argues that technological change under capitalism is incessantly directed at reducing the skills of the labor force. Craftsmen and artisans are replaced by assembly line workers. Deskilled labor is not only cheaper but also easier to control since workers lack direct engagement in the production process.

The opposing view holds that there is substantial technology-skill complementarity. For example, Goldin and Katz (1998) show that in the early twentieth century leading industries—those that invested more heavily in capital and used electric energy more intensely—would subsequently employ relatively more educated blue-collar workers. Technological progress does not destroy skills but on the contrary enables the creation of new positions for managers and engineers. In fact, one of the leading explanations for the increasing inequality between skill groups in the last few decades

[1] See Siebern-Thomas (2005) for example.

is that technological change has been biased toward high-skill workers (Acemoglu, 2002).

The debate over the effect of technological change on skill inspired a large number of empirical studies. In an often-cited review, Spenner (1983) argued these investigations constituted a third position—the "mixed effects or little-net-change hypothesis". According to this position there are offsetting effects of technology on skill requirements, mostly depending on the level of automation of the new processes and the specific characteristics of the organizations orchestrating the change. Also, it is necessary to study the effect of technical change both on job content and the distribution of jobs. The former effect has probably moved the economy in the direction of skill upgrading but the job distribution seems to have favored low and middle skill workers in the service sector. Recent studies all document moderate increases in skill requirements in the US (Howell and Wolff, 1991; Cappelli, 1993; Osterman, 1995). However, in light of the remarkable increases in average education attainment it is clear that the economy as a whole is not creating high-skill jobs at a fast enough pace.

"Skill" is a fuzzy concept that refers to multiple capacities and abilities that workers acquire through education and training. The debate has led researchers to emphasize some aspects of skill and ignore others. According to Spenner (1983), two key dimensions of skill are "substantive complexity" and "autonomy control". The two dimensions appear to be empirically correlated. However, while skill as substantive complexity has been emphasized in the literature on skill upgrading, autonomy control is the preferred focus of the deskilling tradition. Finally, pragmatic considerations—data availability in particular—have also played an important role in determining a metric for skill requirements.

In the essays that follow I use two different measures of skills required on the job. In chapter two I use a statistical measure involving the median education attainment for each sex-occupation group. This measure has the advantage of being widely available and should be a good proxy of true requirements as long as the occupational classification groups together jobs with similar characteristics. The main problem with a statistical measure is that it is sensitive to cohort effects. The large increases in average schooling tends to push every occupation's skill distribution to the right, so that skill requirements are artificially increased. Indeed the reason we use median education and not the arithmetic mean is that the former is less susceptible to this kind of spurious movement. The cohort effects are not neutral either, in the sense that some occupations have received an increasing proportion of new entrants. Therefore this measure cannot be used to study the evolution of mismatch in time. Because the essays in chapters three and four deal with dynamics, I use a measure based on professional assessment of skill requirements that is not sensitive to cohort effects. In this case the drawback is that data are only available for 1977 and 1991.

As mentioned above, skill mismatch can reasonably be thought to be a bad thing in and by itself. Chapter two shows that it is also bad for workers earnings. The approach in this chapter is an extension of the now traditional Mincerian wage equation. Following Duncan and Hoffman (1981), I decompose workers' schooling into required, surplus, and deficit years and estimate an earnings equation that also includes all the standard controls. The data comes from the yearly earnings extracts of the Current Population Survey for the period 1983–2002. I treat the data as a repeated cross section and estimate the equation for each year separately. As in the rest of the now extensive empirical literature on skill mismatch, I find that the returns to surplus and deficit qualifications are very small in absolute value whereas the returns to qualifications that are actually required on the job are much higher than the conventional 6% obtained in standard studies of the returns to human capital investments. Note that the standard Mincerian approach is a restricted version of Duncan and Hoffman's. Simple tests of linear restrictions can be applied to determine whether the unrestricted version has statistical support, which is what I find in every case. Thus, according to these estimates not every investor in human capital gets the same return. Workers who are employed at a job whose requirements match their acquired skills receive a substantial premium relative to over-educated workers with the same level of education. The chapter also includes a probit analysis of the determinants of mismatch. Belonging to a minority group or working for the service sector increases the probability of over-education.

The main objection that can be raised against estimating an unrestricted earnings equation involves the risk of mismeasurement caused by unobservable ability. Low-ability workers, for example, might be incorrectly classified as over-educated because—all else equal—they possess lower skill levels than high-ability workers. Note that this argument, if correct, rebuts the Duncan and Hoffman specification but also that of Mincer. The right specification should contain required qualifications only, assuming these are correctly measured. One way to address this objection is to estimate a panel version of the unrestricted equation and include individual fixed effects. Because an individual's ability level does not vary from one year to the next, the within estimator is implicitly controlling for the ability mismeasurement problem. The main contribution of this part of the dissertation is the result that controlling for fixed effects does not eliminate the statistically significant difference between the returns to required and non-required qualifications.

The second contribution of the dissertation appears in the third chapter. As Cappelli (1993, p.515) puts it: "whether the demand for skills is changing is a vitally important question for public policy, because such a change affects the distribution of income, the extent of technological unemployment, and whether there are shortages of some skills that may lead to a lack of competitiveness, especially relative to economies in which a higher proportion of the labor force possesses those skills." Strangely, despite the burgeoning interest on the increasing earnings inequality in the US, no specialized study of the impact of skill mismatch on wage inequality had been written so far. Researchers focused attention on what fraction of the increase in inequality corresponded to a widening gap between skill groups and what fraction to within-group inequality. Inequality between groups is easy to rationalize in terms of a simple supply and demand model without unemployment or skill mismatch. The fact that within-group or residual inequality is so important empirically could be seen as somewhat puzzling but could nevertheless be explained by an increase in the unobservable returns to unobservable ability. My alternative hypothesis is that the increase in skill mismatch is responsible for a significant percentage of the increase in inequality. Using a Shorrocks-type decomposition, in this chapter I show that the dramatic increase in over-education rates and premia over the period 1973–2002 accounts for around 11 and 32 percent of the increase in the coefficient of variation of log earnings for males and females respectively.

Also dedicated to the consequences of skill mismatch, the fourth chapter is different than the rest of the dissertation in two important and related ways. First, it was written in collaboration with my adviser, Peter Skott. Second, the chapter includes a formal mathematical model. Skill mismatch equilibria have been studied in the context of assignment and search theory models (Sattinger, 2006; Albrecht and Vroman, 2002). Skott (2006) shows that in an efficiency wage framework with skill and job heterogeneity there will also generally be an endogenously determined rate of over-education. The dissertation chapter extends this model to an analysis of the effect of changes in the real value of the minimum wage on employment and the wage distribution. The model is very simple and allows for only two types of job — high and low-tech — and two types of worker — high and low-skill. Furthermore, the model assumes that only high-skill workers can take high-tech positions so the feasible matches are reduced to three possibilities. We show that the binding minimum wage creates an indeterminacy in the model and analyze the results of introducing alternative closures. In what probably is the empirically more relevant case, firms have a preference for low-skill workers in low-tech jobs. Under these conditions, an increase in the minimum wage has positive employment effects for both worker types (i.e. there are monopsonistic effects). Numerical simulations also show that a falling real value of the minimum wage can lead to increasing between- and within-group inequality. Thus, the model provides an alternative explanation for the increase in inequality in the last few decades without relying on skill biased technological change.

Based on insights gained from the model, the fourth chapter includes a number of empirical applications. First, we provide the first estimates of the economy-wide elasticity of substitution between high- and low-tech jobs. Previous studies have focused attention on the degree of substitutability of high and low-skill workers (Autor et al., 2008). However, our model suggests that the appropriate inputs to the production function are jobs and not people. According to our analysis, high-tech jobs are significantly less substitutable by low-tech jobs than college workers are by high-school workers. In other words, it takes a larger proportional change in the high-tech premium to affect the job composition than it takes the college premium to affect the skill composition of the employed labor force. This result is consistent with the existence of substantial skill mismatch. Second, the model also implies that the job composition is endogenously determined. We therefore estimate reduced form equations of both the relative wage and relative employment. The estimates from this section confirm a lower elasticity of substitution between jobs than between skill types. Also, we show that the minimum wage has a negative effect on wage inequality and no significant effect on unemployment, as the model predicts.

CHAPTER 2

SKILL MISMATCH AND EARNINGS: A PANEL ANALYSIS OF THE U.S. LABOR MARKET

2.1 Introduction

The fraction of those employed in the U.S. labor force holding a college degree increased substantially during the period 1983–2002 (See table 2.1). Even more remarkable was the increase in the prevalence of those with post-graduate degrees. The spreading out of higher education to a wider spectrum of the population is not a recent phenomenon but a long run trend. In 1964, the share of all Americans who were high school dropouts was as high as 47% and the proportion of young (age 24–29) people with less than a high school diploma was 31% (Handel, 2003).

The very rapid improvement in average educational attainment naturally leads to the question whether the economy can successfully absorb the growing supply of graduates[1].

In the human capital model of the labor market workers allocate time and resources to education to maximize expected lifetime utility. Profit-maximizing firms, in turn, are willing to fully utilize the skills of their workforce—and reward workers according to their marginal product—by adopting appropriate production techniques

[1]This concern can be traced back to the 1970s. Credentialist theories within sociology argued that corporation's inflating hiring requirements induced over-investment in education (Berg, 1971). In economics, the signalling model also cast a skeptical eye toward the value of educational credentials (Spence, 1973). Freeman (1976) provided convincing evidence showing that Americans were increasingly overeducated, leading to a declining wage premium for college graduates. This was consistent with the view that the educational system's main objective consists in the socialization of students into work norms without truly increasing potential productivity (Bowles and Gintis, 1975, 1976, 2002).

Table 2.1. Education Attainment of the Employed Labor Force: 1979-2002

Year	Education level				
	LTHS	HS	Some College	College	Advanced
1979	23.0%	36.4%	21.8%	13.0%	5.8%
1980	21.6%	36.6%	22.3%	13.4%	6.1%
1981	20.7%	36.9%	22.6%	13.5%	6.3%
1982	19.3%	37.0%	23.1%	14.2%	6.5%
1983	18.2%	36.8%	23.6%	14.6%	6.8%
1984	16.2%	38.7%	22.3%	15.5%	7.3%
1985	15.4%	38.6%	22.7%	15.9%	7.5%
1986	14.7%	38.8%	23.0%	16.1%	7.4%
1987	14.4%	38.2%	23.3%	16.4%	7.6%
1988	14.1%	38.1%	23.4%	16.4%	8.0%
1989	14.1%	38.5%	23.6%	15.9%	7.9%
1990	13.7%	37.9%	24.0%	16.4%	8.1%
1991	13.0%	37.6%	24.4%	16.8%	8.2%
1992	10.8%	38.0%	26.1%	16.8%	8.2%
1993	10.3%	37.2%	27.0%	17.2%	8.3%
1994	9.9%	35.9%	28.0%	17.7%	8.5%
1995	9.7%	34.5%	28.2%	18.4%	9.1%
1996	9.5%	33.5%	28.1%	19.4%	9.5%
1997	9.5%	33.6%	27.6%	19.7%	9.5%
1998	9.4%	32.9%	27.7%	20.1%	9.9%
1999	9.0%	32.1%	27.8%	20.7%	10.5%
2000	9.1%	31.4%	27.9%	21.0%	10.6%
2001	8.6%	31.1%	28.0%	21.3%	11.0%
2002	8.1%	31.0%	27.8%	21.9%	11.2%

Source: Author's Calculations. CPS-ORG.

(Becker, 1964; Schultz, 1971). The existence of skill mismatch, individuals having acquired skills significantly different from those required by their job, appears to be ruled out as a possible outcome of the expansion in qualifications during the last few decades[2].

Workers with at least some college education have a consistently lower probability of being unemployed (figure 2.1). The gap between the unemployment series for high and low-skill workers shows a decreasing tendency during the period, both for males and for females, but have remained fairly constant. In and by itself lower unemployment probabilities, however, do not have any straightforward implications for workers' ability to find jobs where their skill levels are fully utilized. Moreover, after the already mentioned fall during the 1970s, the relative earnings of college graduates—the "college premium"—increased together with the increased supply (Levy and Mur-

[2]However, some degree of mismatch is not completely inconsistent with the neoclassical model. It is entirely possible that some workers are under- or over-educated in the short run, while firms adjust their production processes. Also, the mismatch can be rationalized within the model if workers with the same qualifications have differing unobservable "informal" human capital or job experience (or if they are heterogenous in innate ability).

nane, 1992). This last fact has led many researcher to believe that the demand for college graduates not only kept pace with supply but exceeded it, leading to a hike in the price. Others have remained more skeptical, pointing to the fact that college graduates seem to increasingly be taking jobs that do not normally require the skills acquired through college and that much of the increase in the college premium is due to the constant fall of the real wage for those with high school education only (Hecker, 1992; Shelly, 1992). In other words, while there exists strong evidence that more education does tend to improve the welfare of those who invest in human capital (Mincer, 1974; Psacharopoulos, 1981; Lemieux, 2006b), the evidence also seems to indicate that not every 'investor' gets the same return.

In the recent literature on skill mismatch, every job in the economy is characterized not only by the wage it pays but also by the qualifications it requires (Green et al., 1999; Hartog, 2000; Sloane, 2003; McGuinness, 2006). Workers filling a position are considered correctly matched if their attained education level is equal to the level required by the job. It is also possible for a worker to be mismatched, i.e. placed in a job in which she is under- or over-educated[3]. Mismatched workers have surplus or deficit qualifications for the job.

There is considerable variation in the estimated incidence of skill mismatch. Depending on the measure utilized, the country, the period, and data source, studies have found rates of over-education ranging from 10 to 42%, with an "un-weighted" average of 23.3% in the 25 studies summarized by Groot and Maassen van den Brink (2000). Their average for under-education is 14.4%[4].

Here we extend the existing literature by analyzing a consistent time series of over- and under-education for the U.S. for the period 1983–2002. We focus on the effect

[3]Other terms in the literature are over-qualified, over-schooled, over-trained, under-employed, under-utilized, etc.

[4]The standard deviations for these averages are quite high: 9.9 and 8.2 percentage points respectively.

of match status on earnings. If the human capital model is correct, the returns to required qualifications should not be statistically different from the returns to surplus and deficit qualifications. As we show below, the cross sectional evidence points in a different direction. Skill mismatch imposes substantial penalties on workers' earnings, confirming the findings in most other studies in the literature.

One possible problem with cross-sectional estimates, however, is that surplus and deficit qualifications might be the result of unobserved abilities. Lower ability individuals, for example, might take longer to obtain a given level of skill. Under this scenario it is not surprising to find that the "returns" to surplus qualifications are low. In order to address this issue we estimate the same model for a panel of individuals, which allows controlling for fixed effects. Since ability does not vary within individuals, these estimates do not suffer from the same problem as the cross sectional ones.

The paper proceeds as follows: the next section briefly reviews the existing methodologies for measuring skill mismatch and explains the benefits and possible problems with the statistical method used here. In contrast to previous applications, the estimation of skill requirements is conducted separately for females and males. The third section analyzes the distribution of skill requirements across occupations. It also presents an elementary time series analysis of the over-education series, and an analysis of mismatched rates for some relevant sub-populations. The fourth section estimates Duncan and Hoffman's ORU equation using yearly cross sections. Section five presents the panel analysis. The final section summarizes the results and concludes.

2.2 Measurement issues

There is consensus regarding the difficulty of measuring educational requirements. Researchers have used three main approaches, all of which have advantages and drawbacks. First, the subjective or worker self-assessment method utilizes information given by workers themselves, typically the answer to a question such as "(w)hat was the minimum formal qualification required for (entering) this job?" (Dolton and Vignoles, 2000, p. 182, cited in Chevalier, 2003). The advantage of this approach is the relative specificity of the information regarding the particular job. All other methodologies assign the same educational requirement to all jobs within a pre-determined group or category. However, the measure probably leads to biases arising from respondents' confusion among the qualifications required for entering, keeping, and performing the job. Workers might also inflate their answers as a form of self-praise or simply regurgitate whatever the standard requirement is supposed to be according to custom. At the most basic level workers' assessment of the qualifications required at their jobs are based on a limited (and probably rather small) number of individual experiences regarding educational levels and jobs.

The job-analysis or "objective" measure relies on systematic evaluation by professional job analysts who specify the required level of skills for the job titles in an occupational classification. The best example of such analysis is the United States Dictionary of Occupational Titles (DOT: U.S. Department of Labor, 1977, 1991). A problem with the DOT is that it provides a variety of alternative measures of job-skill requirements. Cognitive, interactive and motor skill indices are linked to consistent employment matrices (267 occupations and 64 industries). The most often used measure of workplace skills is called "General Educational Development" (GED). On a scale of one to six, GED measures mathematical, language and reasoning skills for each job title[5].

The DOT has clear definitions and detailed measurement instructions that all analysts are supposed to follow. Unfortunately, carrying out such detailed analysis is

[5] A good analysis of the trends in the GED and other DOT measures of required skills can be found in Wolff (2000).

very expensive, so the DOT is published only at very wide time intervals. Moreover, later editions mostly repeat the description and analysis of occupations already contained in previous editions, the new research mostly focusing on new categories[6]. An implication is that longitudinal studies require strong assumptions about the behavior of the measures between the years for which there is data. For example, Vaisey (2006) uses a database compiled by Autor, Levy, and Murnane (2003) to allocate a GED score to each occupation in the 3-digit 1970 and 1980 occupational classification for the years 1971 and 1991. He is then forced to apply a linear interpolation to allocate a GED value to occupations in other years. A final problem involves translating the GED score into a "years of education required for the job" measure, which usually requires some extra assumptions. For example, Vaisey uses a regression imputation approach using a separate dataset that contains both the 3-digit occupational codes and self-reported (subjective) education requirements.

As the discussion so far indicates, it is very difficult to accurately identify those jobs that require college degree or some other level of skills. First, standards differ among workers, employers, and experts. Second, whatever standards prevail at one point in time are subject to change due to technological improvements and capital accumulation. The third and final measure of skill mismatch uses a statistical approach to try to overcome these problems. It involves looking at the actual distribution of education for a given occupation and establishing cutoff points beyond which an individual is designated as under- or over-educated. In most studies (Clogg and Shockey, 1984; Verdugo and Verdugo, 1989; Groot, 1993; Cohn and Khan, 1995; Bauer, 2002) the cutoff is one standard deviation below and above the mean, although other measures of central tendency (the median) can be used to attenuate the influence of extreme values. This measure of mismatch is always available and consistent

[6]Spenner (1985) reviews the quality of this type of skill requirement assessment.

for all occupations. It has been criticized because of the arbitrary nature of the one-standard-deviation criterion and because it might be subject to cohort effects when large numbers of workers with the same education level move into an occupation.

Studies that compare the three approaches to measuring mismatch find low correlation among them. Also, the worker self-assessment strategy seems to result in higher estimates of the incidence of over-education (See for example, McGuinness, 2006). In this study we use a statistical approach to measure the incidence of skill mismatch.

2.2.1 The Data

We use the NBER extract of the CPS earnings files (merged outgoing rotation groups) for the years 1983–2002[7]. For the estimation of skill requirements and the calculation of the rates of mismatch, we restrict the sample to employed wage and salary workers who were not students at the time of interview. Other than these restrictions, every individual 16 years of age or older is included. Table 2.2 shows the cumulative effects of these restrictions on sample size.

Because non-response rates are high for the earnings module, the BLS allocates earnings to non-respondents by means of a hot-deck imputation method. While the system arguably increases efficiency for some calculations, it has been shown to produce significant biases in estimates of earnings equations. Also, because the hot-deck involves duplicating the frequency of donors' earnings, it systematically reduces estimates of overall inequality. Thus, for calculations that involve earnings, we exclude observations with allocated earnings whenever the corresponding allocation flag is available. To correct for possible non-random selection into non-response, the sample weight is adjusted by using a probit estimate of the probability of response. The

[7]Details on many issues discussed in this subsection are available in a separate data appendix.

Table 2.2. Sample Restrictions

Year	CPS-ORG Full Sample Size	Employed (%)	Non-student (%)	Wage & Salary (%)	Earnings not allocated (%)
1983	348,521	57.88	57.87	50.84	43.29
1984	343,665	59.53	56.36	49.29	41.45
1985	343,591	60.13	57.19	50.25	42.69
1986	338,051	60.69	57.73	50.80	45.00
1987	337,000	61.52	58.31	51.32	43.85
1988	320,821	62.27	58.88	51.73	43.63
1989	322,883	62.95	59.59	52.29	50.64
1990	339,342	62.80	59.46	52.30	50.56
1991	335,832	61.66	58.48	51.32	49.53
1992	330,588	61.46	58.33	51.17	49.59
1993	326,517	61.72	58.30	51.10	49.46
1994	317,743	62.53	58.60	51.35	50.64
1995	312,973	62.90	59.10	51.97	47.40
1996	276,749	63.17	59.50	52.42	39.73
1997	279,569	63.78	59.94	52.84	39.82
1998	279,221	64.06	60.43	53.50	39.63
1999	281,677	64.25	60.57	53.86	37.87
2000	282,249	64.40	60.97	54.20	36.94
2001	301,952	63.66	60.62	53.93	35.94
2002	328,675	62.73	59.79	53.24	35.42

Note: Restrictions applied sequentially from left to right. The columns give the fraction of the full sample remaining after the corresponding restriction is applied.

earnings weight is also adjusted by multiplying by usual weekly hours, so as to make the sample of hourly earnings representative of the total hours worked in the economy.

The earnings variable we use is constructed to represent real hourly earnings including overtime, tips and commissions. The standard Pareto distribution adjustment is applied to correct for topcoding. No exclusions of observations were made due to implausible or “extreme” wage values. Hourly earnings are weekly earnings including overtime, tips and commissions divided by usual weekly hours, except in the case when a separate (and higher) hourly rate is provided. Earnings are deflated using the CPI-U-X1 series.

Our period of analysis is 1983–2002, i.e. the period during which the 1980 Census occupational classification was used in the CPS. Minor changes in the 3-digit classification were introduced in 1991, so we adjust the occupation variable in the years prior to the change to retain continuity. The other important variable used in this study is educational attainment. We follow the imputation procedure developed by Jaeger (1997a, 2003) to obtain a consistent measure of the highest grade completed[8]. Tables 2.3 and 2.4 contain descriptive statistics for the most important variables used.

[8]The exception is for individuals with at least some college in the years 1992-7. Details in the appendix.

Figure 2.1. Unemployment

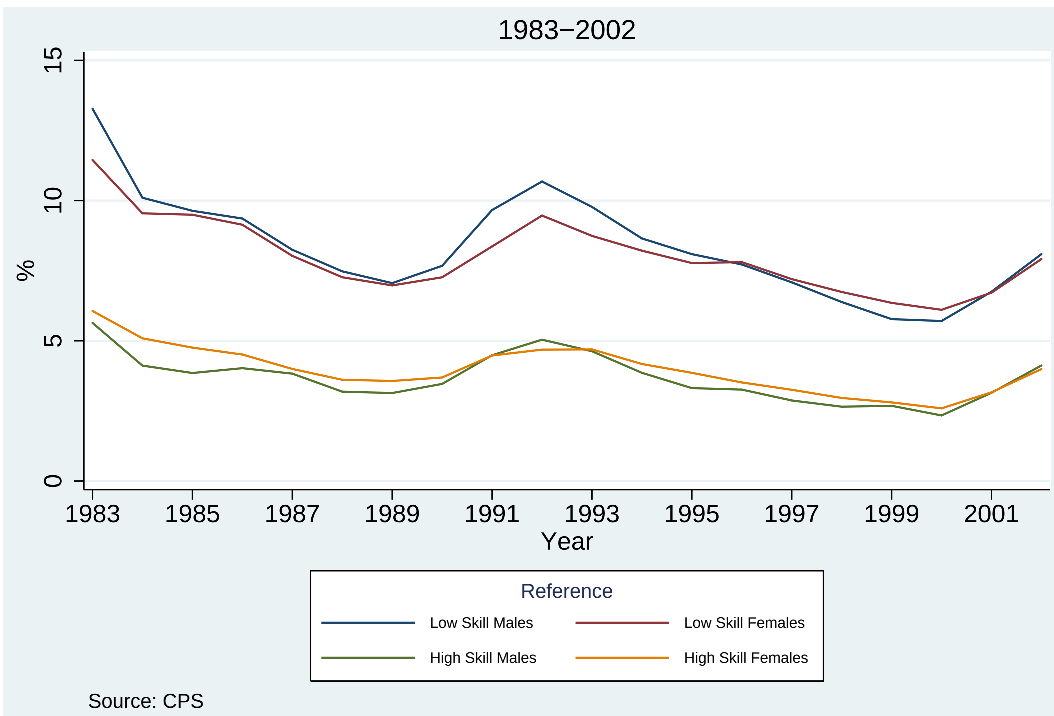


Table 2.3. Descriptive Statistics for Male Sample

Year	Age	Imputed Highest Grade Completed	Nonwhite	Married	Union	Agric	Manuf	Serv	Public Sector	Part time Econ reasons	Metro	Obs
1983	36.65	12.75	17.8%	67.1%	27.7%	2.8%	28.4%	68.8%	15.9%	4.9%	72.1%	93336
1984	37.45	12.82	18.3%	69.7%	26.6%	2.6%	29.8%	67.6%	15.7%	4.2%	71.8%	89540
1985	37.45	12.83	19.8%	69.1%	25.3%	2.5%	29.2%	68.3%	15.3%	4.2%	76.4%	90935
1986	37.42	12.88	20.4%	68.7%	24.6%	2.4%	28.9%	68.6%	15.3%	4.1%	80.2%	89239
1987	37.54	12.91	21.0%	68.5%	23.8%	2.4%	28.0%	69.6%	15.4%	3.8%	80.4%	89251
1988	37.60	12.94	21.5%	67.5%	23.3%	2.3%	27.7%	70.0%	15.2%	3.7%	80.5%	85534
1989	37.70	12.95	21.8%	64.7%	22.6%	2.3%	27.5%	70.2%	15.0%	3.4%	80.2%	86837
1990	37.61	12.95	23.5%	64.3%	22.0%	2.4%	27.1%	70.5%	15.0%	3.7%	80.4%	90902
1991	37.83	13.00	23.9%	64.5%	21.8%	2.5%	26.4%	71.1%	15.4%	4.5%	80.5%	87770
1992	38.02	13.16	24.0%	64.1%	21.1%	2.5%	25.5%	72.0%	15.4%	4.7%	80.1%	85969
1993	38.07	13.22	24.2%	64.1%	20.7%	2.4%	25.0%	72.7%	15.5%	4.4%	79.9%	84588
1994	38.39	13.29	24.1%	64.4%	20.4%	2.2%	25.2%	72.6%	15.1%	3.3%	80.1%	82614
1995	38.52	13.28	24.2%	64.3%	19.6%	2.2%	25.1%	72.7%	14.7%	3.2%	80.2%	82718
1996	38.81	13.33	25.3%	64.1%	19.1%	2.3%	24.6%	73.1%	14.2%	3.0%	81.7%	73615
1997	38.94	13.29	26.5%	63.2%	18.4%	2.3%	24.4%	73.3%	13.7%	2.7%	82.2%	74863
1998	39.14	13.29	27.0%	62.9%	18.2%	2.2%	23.9%	73.8%	13.7%	2.4%	82.3%	75786
1999	39.37	13.35	26.9%	62.9%	18.1%	2.1%	23.0%	74.9%	13.7%	2.2%	82.5%	77132
2000	39.53	13.33	29.2%	62.5%	17.1%	2.4%	22.2%	75.5%	13.4%	2.0%	82.9%	78027
2001	39.82	13.36	29.3%	62.8%	16.9%	2.1%	21.4%	76.5%	13.5%	2.6%	83.2%	83019
2002	40.10	13.39	29.6%	62.8%	16.5%	2.3%	20.5%	77.2%	13.7%	2.8%	82.8%	88454

Table 2.4. Descriptive Statistics for Female Sample

Year	Age	Imputed Highest Grade Completed	Nonwhite	Married	Union	Agric	Manuf	Serv	Public Sector	Part time econ reasons	Metro	Obs
1983	36.26	12.77	18.5%	58.2%	18.0%	1.1%	15.9%	83.0%	19.4%	7.6%	72.7%	81955
1984	37.36	12.84	19.5%	61.7%	17.6%	1.1%	16.9%	82.1%	19.4%	7.0%	72.2%	78113
1985	37.33	12.89	20.1%	61.0%	16.6%	1.0%	16.1%	82.9%	19.3%	6.7%	76.7%	80294
1986	37.39	12.94	20.4%	61.1%	16.2%	1.0%	15.7%	83.3%	19.2%	6.4%	80.1%	80864
1987	37.49	12.98	21.0%	61.3%	15.6%	1.0%	15.5%	83.5%	19.3%	5.9%	80.4%	81956
1988	37.66	13.02	21.4%	61.2%	15.7%	0.9%	15.2%	83.9%	19.4%	5.5%	80.5%	78735
1989	37.97	13.06	21.6%	57.6%	15.5%	0.9%	15.0%	84.1%	19.5%	5.1%	80.0%	80483
1990	38.03	13.10	22.8%	57.2%	15.6%	0.9%	14.4%	84.7%	19.7%	5.2%	80.2%	84925
1991	38.30	13.15	22.9%	57.6%	15.4%	0.9%	14.1%	85.0%	20.0%	6.1%	80.2%	83151
1992	38.48	13.33	23.1%	57.7%	15.5%	0.8%	13.7%	85.5%	20.1%	6.2%	80.0%	81945
1993	38.70	13.40	23.2%	57.5%	15.7%	0.9%	13.0%	86.1%	20.7%	6.3%	79.7%	81236
1994	38.97	13.45	23.5%	57.9%	15.8%	0.8%	12.9%	86.3%	20.1%	4.3%	79.8%	79227
1995	39.12	13.44	23.6%	58.0%	15.0%	0.7%	12.6%	86.7%	19.8%	4.1%	80.0%	78719
1996	39.30	13.54	24.9%	57.9%	14.5%	0.8%	12.5%	86.6%	19.5%	4.0%	82.0%	70465
1997	39.48	13.52	25.8%	57.2%	14.0%	0.8%	12.5%	86.7%	19.1%	3.6%	82.0%	71698
1998	39.70	13.51	26.5%	56.6%	13.7%	0.8%	12.1%	87.1%	18.9%	3.2%	82.3%	72301
1999	39.89	13.55	27.4%	56.5%	13.7%	0.8%	11.6%	87.6%	19.3%	2.8%	82.3%	73283
2000	40.13	13.56	28.4%	56.4%	13.8%	0.8%	11.6%	87.5%	19.3%	2.6%	82.3%	73813
2001	40.43	13.61	28.7%	56.5%	13.8%	0.9%	10.7%	88.4%	19.4%	2.9%	82.7%	78799
2002	40.79	13.66	28.8%	56.5%	13.7%	0.9%	9.7%	89.4%	19.6%	3.3%	82.7%	85313

2.3 Skill Requirements

Surveys asking workers what level of education is *required* at their current jobs indicate that most jobs in retail sales; administrative support (including clerical); service; farm; precision production, craft, and repair; and operator, fabricator, and laborer occupational groups do not require a college degree for entry, nor do they offer job duties attractive to most college graduates. In contrast, jobs in managerial, professional, and/or technical occupations require a degree, in the sense that the skills generally learnt in college are necessary in order to successfully accomplish most tasks involved by the job. Thus, a first approach to measuring the proportion of college graduates and post-graduates whose skills are underutilized at their jobs involves focusing on skilled workers that are employed in occupations within retail sales and the other major occupational groups identified as most often containing non-college jobs. This strategy, originally developed in Hecker (1992), can be seen as a preliminary version of the statistical approach utilized here.

The statistical method relies on the 3-digit occupational classification of the 1980 census. The classification comprises 501 occupations. Figure 2.2 exemplifies the methodology for the case of female apparel sales workers (coded 264). Workers' education attainment in this occupation if clearly concentrated around the median of 12 years of formal schooling. The imputed years of schooling variable takes only discrete values. The points in the scatter plot have been added some jitter, so that the relative frequency of each value is represented by the density of the cloud of points around the true value[9]. For each year we also create a one standard deviation interval below and above the median. The resulting cutoffs are smoothed using a Hodrick-Prescott filter. Workers whose formal schooling is above or below the cutoffs are considered mismatched .

[9]Thus, the cloud of points around the cutoff lines represent workers who are *not* considered mismatched.

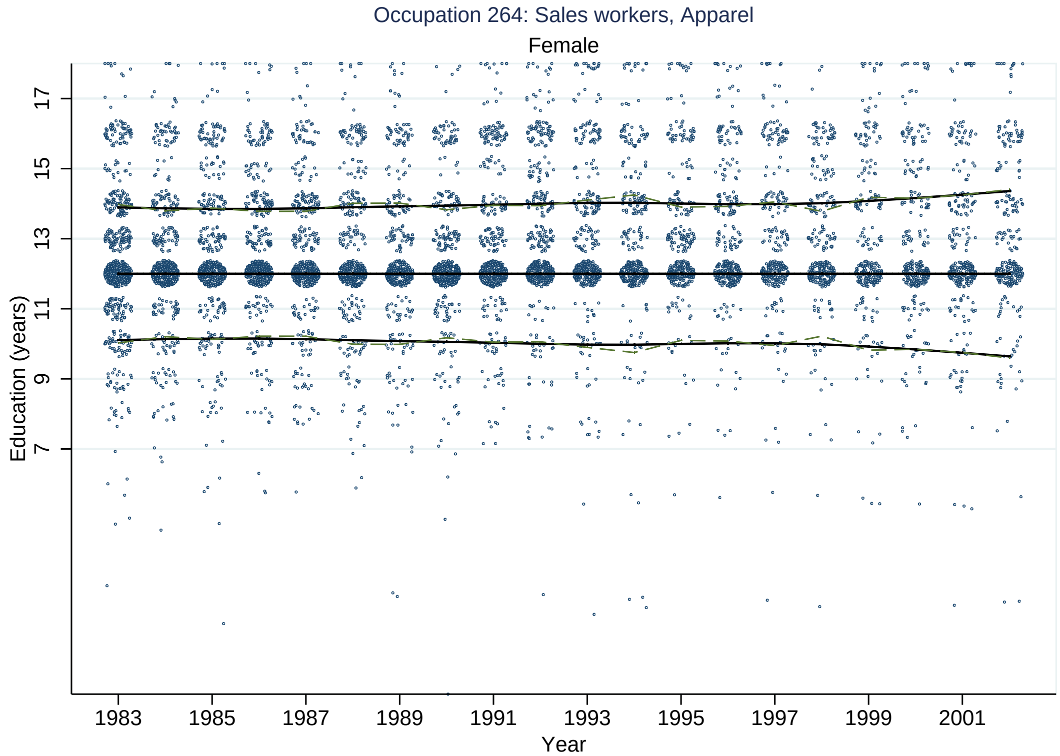


Figure 2.2. Example of Skill Requirement Estimation

Tables 2.5 and 2.6 present statistics for skill requirements and mismatch rates aggregated at the 2-digit occupational classification. Skill requirements range from less than high-school equivalent in farming, forestry and fishing occupations to 18 years (the topcode) in occupations that require attending graduate schools like law and medicine. During the 20 year period requirements increased by around half a year on average, both for males and females. The increase should be compared to the 0.64 and 0.89 year increase in workers' education attainment, again for males and females respectively. As the education distribution moved to the right, median education in many occupations has increased shifting the estimate for required qualifications. Note also that the range within which an individual is classified as correctly matched is normally quite large: 4 years of schooling or more in almost all cases.

For males, the occupational groups with highest rates of over-education (education beyond the upper cutoff) are computer equipment operators, salesmen, and protective service employees. In all these cases the over-education rates range between 20 and 25%. Women suffer from markedly higher rates, with secretaries and financial records processing occupations rising above 30%. Overall, over-education is clearly more prevalent among women than men. With few exceptions, under-educated workers are high-school dropouts. Because the dropout category is more common among males, this type of mismatch also is. The occupations in which it is most prevalent include technicians and managers. For females, however, under-education is more common in the sciences and in some health treatment occupations.

Table 2.5. The Evolution of Skill Requirements: Males

	Occupation (2-digit)	Educ Requirement Mean	Educ Requirement Average Yearly Change	Lower cutoff	Upper cutoff	Overeduc rate Mean	Overeduc rate Average Yearly Change	Undereduc rate Mean	Undereduc rate Average Yearly Change
1	Officials & administrators, pub admin	15.74	0.00	13.5	18.0	3.1%	0.06%	20.0%	-0.30%
2	Other executive, admin & managerial	15.59	0.03	13.4	17.9	9.1%	-0.34%	27.1%	-0.11%
3	Management related occupations	15.49	0.02	13.6	17.4	16.8%	0.25%	16.6%	-0.14%
4	Engineers	16.03	0.00	14.2	17.8	21.9%	0.40%	21.8%	-0.16%
5	Mathematical and computer scientists	16.05	0.00	14.2	17.9	19.1%	0.71%	22.8%	-0.23%
6	Natural Scientists	17.40	0.03	16.0	18.9	3.5%	-0.79%	23.2%	0.45%
7	Health diagnosing occs	18.00	0.00	17.2	18.8	0.0%	0.00%	5.3%	-0.23%
8	Health assessment and treatment occs	16.05	0.02	14.5	17.6	16.1%	0.07%	15.9%	-0.17%
9	Teachers, college and university	17.99	0.00	17.1	18.9	0.7%	0.21%	14.2%	-0.27%
10	Teachers, except college and university	17.14	0.04	15.7	18.6	3.2%	-0.10%	19.6%	0.89%
11	Lawyers and judges	18.00	0.00	17.2	18.8	0.0%	0.00%	7.0%	-0.41%
12	Other professional specialty occs	16.20	0.01	14.2	18.2	9.1%	-0.09%	20.4%	0.05%
13	Health technologists and technicians	14.55	-0.01	12.6	16.5	19.6%	0.86%	21.2%	0.46%
14	Engineering and science technicians	13.65	0.04	11.8	15.5	21.5%	0.12%	14.6%	1.05%
15	Technicians, exc. Health, Engineering & Science	15.65	0.02	13.7	17.6	17.2%	0.22%	29.5%	-0.44%
16	Supervisors and proprietors, sales occs	13.50	0.05	11.4	15.7	21.5%	-0.87%	4.7%	0.04%
17	Sales reps, finance and business serv	15.49	0.05	13.5	17.6	10.6%	-0.38%	25.9%	0.00%
18	Sales reps, commodities, exc. retail	14.34	0.05	12.2	16.5	9.8%	0.07%	24.2%	1.03%
19	Sales workers, retail & personal serv	12.57	0.02	10.5	14.7	21.5%	0.23%	6.3%	-0.11%
20	Sales related occs	12.93	-0.02	10.5	15.3	22.7%	0.23%	9.1%	-0.15%
21	Supervisors, admin support	13.54	0.00	11.4	15.6	20.8%	-0.25%	10.7%	0.03%
22	Computer equipment operators	13.64	0.02	11.7	15.6	24.5%	0.16%	3.4%	-0.05%
23	Secretaries, stenographers, and typists	13.68	0.05	11.4	15.8	21.3%	-0.78%	5.5%	0.11%
24	Financial records processing	13.68	0.06	11.6	15.7	20.9%	-1.10%	5.3%	0.02%
25	Mail and message distributing	12.33	0.03	10.5	14.1	23.6%	0.11%	4.0%	-0.11%
26	Other admin support, inc. clerical	12.93	0.03	10.8	15.1	18.1%	-0.01%	8.4%	-0.07%
27	Private household service occs	11.65	0.04	8.7	14.7	11.5%	1.20%	18.8%	0.05%
28	Protective service	12.87	0.04	10.9	14.9	22.0%	0.16%	12.8%	0.39%
29	Food service	11.96	0.01	9.1	14.8	9.8%	0.12%	15.2%	0.11%
30	Health service	12.23	0.04	10.1	14.4	17.4%	-0.04%	6.6%	-0.11%
31	Cleaning and building service	12.00	0.00	9.2	14.8	6.4%	0.04%	20.3%	-0.35%
32	Personal service	12.33	-0.01	10.1	14.6	18.5%	0.04%	9.0%	-0.42%
33	Mechanics and repairers	12.14	0.01	10.2	14.1	15.2%	0.35%	10.8%	-0.29%
34	Construction trades	11.99	0.00	9.8	14.2	11.3%	0.12%	13.5%	-0.07%
35	Other precision prod, craft, & repair	12.01	0.00	9.9	14.2	15.4%	0.30%	9.8%	-0.18%
36	Machine opertrs and tenders,exc precis	11.96	0.00	9.5	14.4	6.8%	0.16%	15.7%	-0.27%
37	Fabricatrs,assemblrs,inspectrs,samplrs	11.99	0.00	9.6	14.4	7.3%	0.09%	12.6%	-0.26%
38	Motor vehicle operators	12.00	0.00	9.9	14.1	8.9%	0.23%	12.3%	-0.53%
39	Other transp & material moving occs	12.01	0.00	9.8	14.2	6.7%	0.23%	16.0%	-0.58%
40	Construction laborers	12.00	0.00	9.1	14.9	4.8%	-0.03%	20.9%	-0.38%
41	Freight, stock & materials handlers	11.99	0.00	9.7	14.3	7.4%	-0.09%	12.9%	-0.65%
42	Other handlers,equip. cleaners, helpers, laborers	11.99	0.00	9.4	14.6	5.3%	0.08%	16.4%	-0.26%
43	Farm operators and managers	12.06	0.01	9.1	15.0	20.9%	0.19%	12.9%	-0.42%
44	Farm workers and related occupations	11.15	0.01	7.5	14.8	5.3%	0.10%	24.1%	0.22%
45	Forestry and fishing occs	11.76	0.02	9.0	14.5	7.5%	0.19%	18.5%	-0.23%
	Total	13.38	0.02	11.2	15.6	12.1%	0.05%	15.7%	-0.07%

Note: Source CPS-ORG 1983-2002

Table 2.6. The Evolution of Skill Requirements: Females

	Occupation (2-digit)	Educ Requirement Mean	Educ Requirement Mean Yearly Change	Lower cutoff	Upper cutoff	Overeduc rate Mean	Overeduc rate Mean Yearly Change	Undereduc rate Mean	Undereduc rate Mean Yearly Change
1	Officials & administrators, pub admin	14.0	0.06	11.8	16.3	22.4%	-0.29%	6.7%	1.08%
2	Other executive, admin & managerial	14.1	0.06	12.0	16.4	20.3%	-0.51%	10.4%	0.31%
3	Management related occupations	14.6	0.06	12.6	16.7	16.7%	-0.03%	17.8%	0.63%
4	Engineers	16.0	0.05	14.3	17.8	17.7%	0.35%	18.1%	0.64%
5	Mathematical and computer scientists	16.0	0.00	14.0	17.9	16.2%	0.84%	28.9%	0.13%
6	Natural Scientists	17.0	0.06	15.7	18.6	8.3%	-0.49%	16.6%	1.06%
7	Health diagnosing occs	17.9	0.01	16.6	19.3	0.5%	0.17%	13.7%	-0.54%
8	Health assessment and treatment occs	15.8	0.06	14.3	17.5	11.9%	0.16%	25.6%	0.38%
9	Teachers, college and university	17.8	0.02	16.8	19.0	1.0%	0.14%	17.9%	-0.36%
10	Teachers, except college and university	16.9	0.06	15.4	18.6	5.8%	-0.59%	20.5%	1.07%
11	Lawyers and judges	17.9	0.02	16.7	19.1	0.4%	-0.13%	10.3%	-0.35%
12	Other professional specialty occs	15.9	0.01	13.9	18.0	8.1%	-0.03%	20.2%	0.07%
13	Health technologists and technicians	13.6	0.03	12.0	15.2	17.1%	-0.16%	13.2%	-0.44%
14	Engineering and science technicians	13.5	0.04	11.4	15.5	23.3%	-0.24%	8.7%	0.31%
15	Technicians, exc. Health, Engineering & Science	14.9	0.00	12.9	16.9	17.4%	-0.32%	22.7%	-0.23%
16	Supervisors and proprietors, sales occs	12.3	0.01	10.4	14.4	27.8%	-0.35%	4.5%	-0.19%
17	Sales reps, finance and business serv	13.9	0.03	11.9	16.0	23.0%	-0.31%	9.0%	0.08%
18	Sales reps, commodities, exc. retail	14.0	0.05	11.8	16.2	20.1%	-1.11%	6.8%	1.21%
19	Sales workers, retail & personal serv	12.0	0.00	10.2	13.9	18.2%	0.09%	9.9%	-0.13%
20	Sales related occs	12.3	0.00	10.3	14.2	18.8%	-0.84%	7.9%	0.10%
21	Supervisors, admin support	12.6	0.08	10.8	14.5	26.4%	-0.08%	1.9%	0.05%
22	Computer equipment operators	12.2	0.04	10.5	13.8	27.5%	-0.41%	2.2%	0.16%
23	Secretaries, stenographers, and typists	12.1	0.01	10.5	13.6	30.6%	0.43%	1.8%	0.06%
24	Financial records processing	12.1	0.00	10.4	13.7	30.1%	0.69%	2.2%	-0.07%
25	Mail and message distributing	12.1	0.00	10.4	13.8	22.2%	0.29%	4.9%	-0.21%
26	Other admin support, inc. clerical	12.2	0.02	10.5	14.0	25.7%	0.22%	3.5%	-0.01%
27	Private household service occs	11.6	0.07	8.5	14.7	6.6%	0.15%	21.4%	0.04%
28	Protective service	12.6	0.02	10.6	14.5	20.2%	0.06%	10.7%	-0.15%
29	Food service	12.0	0.00	9.9	14.1	11.0%	0.05%	15.2%	-0.28%
30	Health service	12.0	0.00	10.1	14.0	16.6%	-0.16%	11.7%	-0.45%
31	Cleaning and building service	11.9	0.01	9.1	14.7	3.2%	-0.05%	23.7%	-0.38%
32	Personal service	12.2	0.00	10.2	14.1	15.5%	0.44%	8.5%	-0.27%
33	Mechanics and repairers	12.4	0.03	10.5	14.4	20.2%	0.14%	8.9%	-0.25%
34	Construction trades	12.1	0.00	10.2	14.1	16.1%	0.62%	12.7%	-0.59%
35	Other precision prod, craft, & repair	12.0	0.01	9.7	14.3	9.0%	0.16%	13.2%	-0.25%
36	Machine opertrs and tenders,exc precis	11.9	0.03	9.3	14.4	4.1%	0.09%	21.0%	-0.27%
37	Fabricatrs,assemblrs,inspectrs,samplrs	11.9	0.00	9.5	14.4	5.2%	0.16%	16.2%	-0.37%
38	Motor vehicle operators	12.0	0.00	10.4	13.7	14.5%	0.44%	11.6%	-0.11%
39	Other transp & material moving occs	12.0	0.00	10.3	13.7	10.5%	0.41%	14.3%	-1.01%
40	Construction laborers	12.0	0.00	9.9	14.1	8.1%	-0.18%	15.2%	0.02%
41	Freight, stock & materials handlers	12.0	0.00	10.1	13.9	10.4%	0.00%	13.8%	-0.50%
42	Other handlers,equip. cleaners, helpers, laborers	12.0	0.00	9.3	14.7	4.5%	0.07%	17.8%	-0.09%
43	Farm operators and managers	12.5	0.01	10.3	14.7	17.7%	0.75%	14.4%	-0.94%
44	Farm workers and related occupations	11.5	-0.05	8.4	14.7	10.7%	0.35%	18.3%	-0.21%
45	Forestry and fishing occs	11.8	0.08	9.1	14.2	19.2%	-2.07%	19.6%	-1.20%
	Total	13.2	0.02	11.3	15.2	17.1%	-0.02%	12.0%	-0.05%

Note: Source CPS-ORG 1983-2002.

Figures 2.3 and 2.4 show the evolution of skill mismatch during the period. For the entire employed labor force, over-education rates increased slightly for males and remained fairly constant for females. Over-education fell markedly for high-skill workers, mostly because of the increasing rates of participation in college and graduate programs. As discussed above, the incorporation to the labor force of cohorts of highly educated workers may inflate the estimate of required qualifications. As a result, the cutoff points that determine who is classified as mismatched have shifted upwards. Thus, the statistical measure of skill requirements does not produce reliable estimates of the evolution in time of over- and under-education rates. In this study we focus on inference at the cross section level. The exception is the panel study, which links individuals in two consecutive survey years. Because the panel is very short, cohort effects do not significantly affect the results.

The over-education series display a slight hump in the early 90s, more or less in synchrony with the higher unemployment rates of the previous years. The visual impression is not confirmed by further analysis, however. A regression of the over-education rate on a time trend and two lags of the unemployment rates for high- and low-skill workers does not render any significant coefficient for males or females.

The risk of mismatch is not evenly distributed. According to table 2.7, over-education is particularly important for workers 25–34 years old. One possible interpretation is that mismatch of this kind is chosen rationally as a career path. Also white workers have higher risk of being over-educated and lower risk of under-education. It might seem like over-education is not a problem related to minority status. The simple descriptives also indicate higher over-education in the service sector and higher under-education in manufacturing. Finally, participation in a union contract reduces both forms of skill mismatch. These results are nonetheless affected by the different levels of education attainment within the sub-populations. In table 2.8 we present results from probit analysis conducted for the degrees of mismatch. Once education

Figure 2.3. The Incidence of Over-education

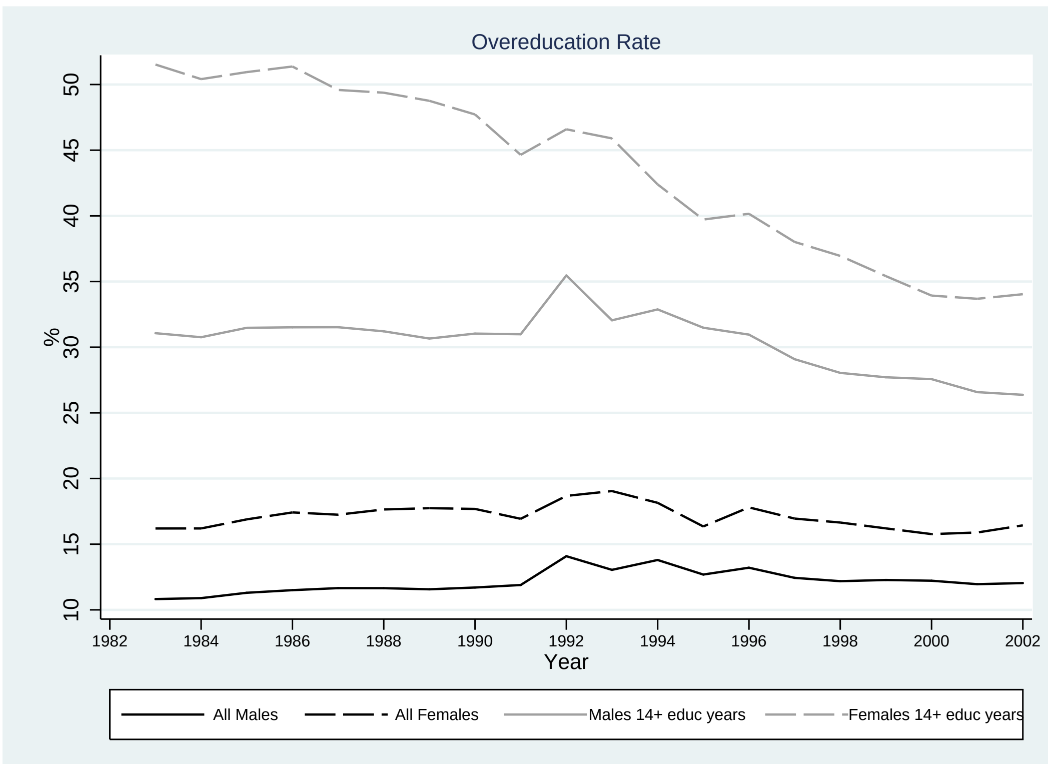


Figure 2.4. The Incidence of Under-education

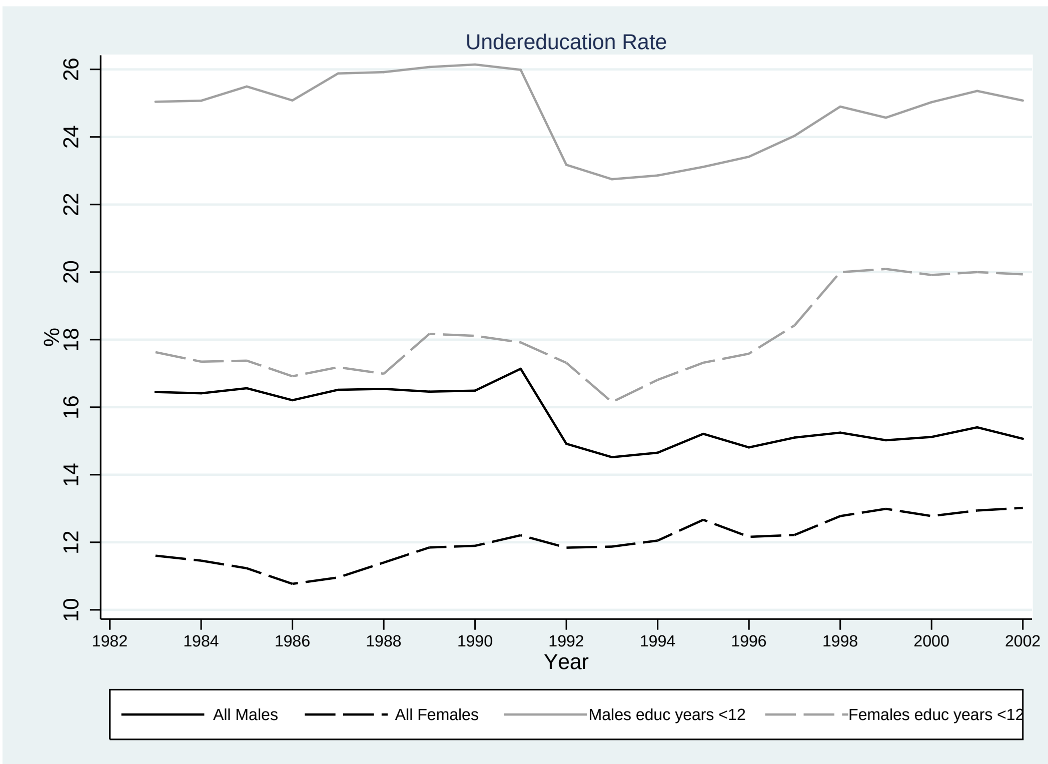


Table 2.7. Mismatch Rates for Sub-populations

Subpopulation	Overeduc Rate (%)		Undereduc Rate (%)	
	Male	Female	Male	Female
16–24 years old	6.0	15.1	13.4	10.5
25–34 years old	13.2	20.9	12.8	9.6
35–54 years old	13.8	16.5	15.7	12.5
55 and older	10.5	12.6	24.8	17.8
White	12.7	17.4	13.4	10.6
Non-white	10.5	16.1	22.6	16.8
Manufacturing	11.2	12.2	16.0	14.4
Services	12.7	17.9	15.2	11.7
Union	11.8	13.9	12.2	11.7
Non-union	12.3	17.6	16.6	12.2

Note: Source CPS-ORG.

and other controls are introduced, the effect of age on the probability of mismatch is almost negligible. The marginal effect of minority status is actually positive for over-education and negative for under-education, as one might have expected. Unions decrease over-education among females and under-education for males, but the effects are reverted for the other cases. The stylized fact is that higher education attainment increases the likelihood of over-education and reduces under-education, and that differences in the degree of mismatch for different groups depend on differences in how educated their members are.

2.4 Mismatch and Earnings

Skill mismatch has been found to affect workers' earnings. In one interpretation, labor productivity is determined by the characteristics of the job including skill requirements and not by workers' characteristics (Thurow, 1975). Employers may give preference to workers with higher educational attainment because this and other characteristics are taken to signal low training costs. Once hired, workers are taught the skills actually required on the job. On one hand, jobs requiring more skills will tend to pay higher wages because labor productivity is higher at those jobs (see Bewley, 1999, for an argument to the contrary). Workers have an incentive to invest in acquiring skills in order to get the higher earnings. On the other hand, if workers

Table 2.8. Probit Analysis of Skill Mismatch

Mg. Effects	(1) Overed Females	(2) Overed Males	(3) Undered Females	(4) Undered Males
Non-white	0.009***	0.002***	-0.003***	-0.000
	[0.000]	[0.000]	[0.001]	[0.001]
Manufacturing	-0.008***	-0.001***	0.005*	0.051***
	[0.001]	[0.000]	[0.003]	[0.002]
Services	-0.013***	-0.003***	0.030***	0.059***
	[0.001]	[0.000]	[0.002]	[0.001]
Age 25-34	-0.002***	0.000***	0.033***	0.056***
	[0.000]	[0.000]	[0.001]	[0.001]
Age 35-54	-0.005***	0.000	0.046***	0.086***
	[0.000]	[0.000]	[0.001]	[0.001]
Age 55-	-0.002***	-0.000***	0.057***	0.146***
	[0.000]	[0.000]	[0.001]	[0.002]
Union	-0.019***	0.002***	0.021***	-0.047***
	[0.000]	[0.000]	[0.001]	[0.000]
Metro	0.005***	-0.000***	0.003***	0.017***
	[0.000]	[0.000]	[0.001]	[0.001]
Married	-0.005***	-0.001***	0.013***	0.024***
	[0.000]	[0.000]	[0.000]	[0.001]
High School Grad	0.019***	0.147***	-0.215***	-0.199***
	[0.005]	[0.016]	[0.001]	[0.001]
Some College	0.641***	0.944***	-0.164***	-0.168***
	[0.017]	[0.003]	[0.001]	[0.001]
College Grad	0.862***	0.994***	-0.140***	-0.179***
	[0.011]	[0.000]	[0.000]	[0.000]
Advanced Degree	0.913***	0.999***	-0.116***	-0.154***
	[0.009]	[0.000]	[0.000]	[0.000]
Observations	1502856	1614665	1502964	1614680

Note: For dummy variables dF/dx is discrete change from 0 to 1. Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Regional and year dummies not reported.

are able to get the same job with less prior formal schooling, then the return to their human capital will appear to be higher. Conversely, surplus education should not receive compensation of any sort.

Tables 2.9 and 2.10 provide evidence in favor of the first point. Male occupations that require only a high-school degree such as construction laborers tend to pay lower hourly earnings than engineering and science related occupations. Similarly, the remarkable transition of women out of secretarial and clerical jobs and into managerial occupations has resulted in higher earnings. Regarding the second point, workers with surplus (deficit) education earn slightly more (less) than those that are correctly matched. In other words, the returns to non-required years appear not to be zero. This raises the question whether our distinction between required and non-required years is relevant.

Table 2.9. Occupational distribution and earnings: Males

	Occupation (2-digit)	Distribution (%) 1983	1992	2002	Earnings Correct Match Mean Log ($)	Average growth (%)	Earnings Overeduc Mean Log ($)	Average growth (%)	Earnings Undereduc Mean Log ($)	Average growth (%)
1	Officials & administrators, pub admin	0.7	0.7	0.7	2.23	0.7	2.33	-2.4	1.90	1.0
2	Other executive, admin & managerial	7.7	8.7	10.8	2.24	0.7	2.43	-0.5	1.93	0.7
3	Management related occupations	3.1	3.1	3.1	2.10	0.5	2.29	0.6	1.92	0.0
4	Engineers	3.0	2.8	2.8	2.33	0.2	2.46	0.0	2.14	0.4
5	Mathematical and computer scientists	0.6	1.1	2.1	2.30	0.5	2.44	0.8	2.19	-0.8
6	Natural Scientists	0.6	0.6	0.5	2.20	1.1	2.28	-5.6	2.03	-0.4
7	Health diagnosing occs	0.4	0.5	0.8	2.29	3.4	2.00	-	1.87	1.7
8	Health assessment and treatment occs	0.5	0.6	0.7	2.11	2.0	2.24	2.3	1.87	1.7
9	Teachers, college and university	0.8	0.8	0.9	2.20	0.5	1.98	-	1.84	1.2
10	Teachers, except college and university	2.0	1.9	2.2	1.96	0.8	2.03	-0.8	1.73	-0.5
11	Lawyers and judges	0.6	0.6	0.7	2.55	1.5	-	1.2	2.06	1.5
12	Other professional specialty occs	2.8	2.8	3.1	1.85	0.7	2.11	1.0	1.63	0.5
13	Health technologists and technicians	0.4	0.5	0.5	1.75	1.4	1.94	0.5	1.65	-0.2
14	Engineering and science technicians	1.7	1.6	1.4	1.89	0.4	1.95	1.8	1.83	0.8
15	Technicians, exc. Health, Engineering & Science	1.2	1.8	1.2	2.17	0.6	2.28	2.1	2.03	0.8
16	Supervisors and proprietors, sales occs	2.6	3.0	3.4	1.78	1.0	2.06	2.5	1.52	0.1
17	Sales reps, finance and business serv	1.6	1.7	2.0	2.12	1.3	2.38	1.3	1.83	0.6
18	Sales reps, commodities, exc. retail	2.1	2.0	1.8	2.04	1.1	2.28	1.3	1.77	0.5
19	Sales workers, retail & personal serv	3.0	3.1	3.2	1.37	0.6	1.61	1.2	1.17	0.9
20	Sales related occs	0.0	0.0	0.0	1.51	-1.2	1.67	-0.5	1.16	-3.1
21	Supervisors, admin support	0.7	0.6	0.4	1.96	-0.5	2.15	0.6	1.88	-0.7
22	Computer equipment operators	0.4	0.4	0.3	1.76	0.5	1.91	1.9	1.58	-0.2
23	Secretaries, stenographers, and typists	0.2	0.1	0.1	1.55	-0.1	1.85	0.4	1.52	-3.3
24	Financial records processing	0.5	0.4	0.3	1.62	0.1	1.77	-0.1	1.54	1.1
25	Mail and message distributing	1.2	1.0	0.9	1.80	0.4	1.86	0.4	1.56	1.6
26	Other admin support, inc. clerical	3.9	4.2	4.0	1.61	0.1	1.74	-3.5	1.51	0.4
27	Private household service occs	0.1	0.1	0.0	0.80	1.3	0.89	1.0	0.68	-3.7
28	Protective service	3.0	3.2	3.3	1.66	0.5	1.90	1.0	1.69	1.7
29	Food service	3.6	3.4	3.7	1.12	0.6	1.28	0.3	1.04	-0.1
30	Health service	0.4	0.4	0.5	1.32	0.3	1.49	1.3	1.11	-0.5
31	Cleaning and building service	3.4	3.1	2.6	1.35	0.3	1.44	1.1	1.20	0.0
32	Personal service	0.6	0.6	0.7	1.31	1.5	1.47	0.5	1.09	1.6
33	Mechanics and repairers	7.4	7.2	6.6	1.76	0.3	1.91	0.3	1.56	-0.2
34	Construction trades	6.7	6.7	7.7	1.77	-0.1	1.94	0.2	1.53	-1.0
35	Other precision prod, craft, & repair	6.1	5.4	4.3	1.81	-0.3	1.99	0.5	1.55	-1.2
36	Machine opertrs and tenders,exc precis	6.1	5.4	4.2	1.59	0.0	1.70	0.0	1.36	-0.8
37	Fabricatrs,assemblrs,inspectrs,samplrs	3.0	3.0	2.4	1.64	-0.2	1.76	0.5	1.36	-1.4
38	Motor vehicle operators	5.0	5.6	5.7	1.57	0.2	1.62	0.2	1.43	-0.1
39	Other transp & material moving occs	2.3	2.1	2.0	1.72	-0.2	1.94	2.4	1.57	-0.4
40	Construction laborers	1.1	1.1	1.4	1.51	-0.1	1.63	0.1	1.36	-0.4
41	Freight, stock & materials handlers	2.6	2.2	2.0	1.38	0.2	1.48	0.7	1.18	-0.1
42	Other handlers,equip. cleaners, helpers, laborers	3.2	3.0	2.6	1.35	0.1	1.47	2.6	1.15	-0.6
43	Farm operators and managers	0.1	0.1	0.1	1.29	0.6	1.58	1.7	1.04	1.5
44	Farm workers and related occupations	3.0	2.8	2.4	1.11	1.2	1.35	-0.6	1.00	0.9
45	Forestry and fishing occs	0.2	0.2	0.1	1.47	0.0	1.71		1.12	-0.2
	Total	100	100	100	1.77	0.49	1.96	0.44	1.64	0.27

Note: Source CPS-ORG.

Table 2.10. Occupational distribution and earnings: Females

Occupation (2-digit)		Distribution (%) 1983	1992	2002	Earnings Correct Match Mean Log ($)	Average growth (%)	Earnings Overeduc Mean Log ($)	Average growth (%)	Earnings Undereduc Mean Log ($)	Average growth (%)
1	Officials & administrators, pub admin	0.4	0.5	0.7	1.86	1.8	2.20	1.0	1.76	4.5
2	Other executive, admin & managerial	4.4	6.8	9.5	1.79	1.5	2.05	2.0	1.58	1.3
3	Management related occupations	2.8	4.1	5.1	1.82	1.1	1.96	1.1	1.70	0.8
4	Engineers	0.2	0.3	0.4	2.20	0.9	2.33	1.1	1.98	-0.9
5	Mathematical and computer scientists	0.3	0.7	1.1	2.15	0.7	2.32	3.7	2.02	1.4
6	Natural Scientists	0.2	0.3	0.3	2.00	1.1	2.13	3.3	1.84	1.6
7	Health diagnosing occs	0.2	0.2	0.5	2.12	2.5	2.06	-32.0	1.68	0.6
8	Health assessment and treatment occs	3.9	4.5	5.0	2.04	1.7	2.14	1.3	1.92	1.7
9	Teachers, college and university	0.6	0.6	0.7	2.00	0.7	1.93	31.7	1.76	2.4
10	Teachers, except college and university	5.5	6.4	7.2	1.83	1.1	1.93	1.3	1.48	1.5
11	Lawyers and judges	0.2	0.3	0.4	2.37	2.3	2.54	-3.1	1.78	4.9
12	Other professional specialty occs	2.9	3.5	4.2	1.78	1.1	1.85	1.1	1.49	0.5
13	Health technologists and technicians	2.3	2.5	2.6	1.64	0.8	1.75	0.7	1.57	1.0
14	Engineering and science technicians	0.5	0.5	0.5	1.66	0.6	1.76	0.4	1.60	1.5
15	Technicians, exc. Health, Engineering & Science	0.8	1.2	1.0	1.90	1.1	1.92	0.6	1.72	0.3
16	Supervisors and proprietors, sales occs	1.4	2.0	2.6	1.45	1.0	1.73	1.9	1.24	-0.2
17	Sales reps, finance and business serv	1.3	1.5	1.9	1.70	1.1	1.93	2.1	1.58	0.1
18	Sales reps, commodities, exc. retail	0.5	0.7	0.7	1.75	2.5	2.01	2.7	1.50	2.2
19	Sales workers, retail & personal serv	8.7	6.6	5.7	1.08	0.5	1.25	1.1	0.97	0.1
20	Sales related occs	0.1	0.1	0.1	1.16	-2.0	1.50	0.6	1.09	2.0
21	Supervisors, admin support	0.9	0.9	0.8	1.71	0.5	1.88	0.7	1.56	0.4
22	Computer equipment operators	0.9	0.9	0.3	1.50	0.9	1.59	1.8	1.37	-1.0
23	Secretaries, stenographers, and typists	11.5	8.4	4.9	1.48	0.6	1.54	0.7	1.33	0.9
24	Financial records processing	4.9	4.0	3.2	1.47	0.7	1.53	0.8	1.33	0.6
25	Mail and message distributing	0.6	0.7	0.6	1.69	0.8	1.76	0.8	1.42	1.5
26	Other admin support, inc. clerical	12.4	14.4	13.9	1.41	0.5	1.51	0.8	1.24	0.5
27	Private household service occs	2.1	1.3	1.0	0.67	2.4	0.87	4.8	0.60	2.4
28	Protective service	0.5	0.7	0.9	1.49	0.6	1.71	1.7	1.39	2.5
29	Food service	7.4	5.9	5.3	0.98	0.8	1.07	1.9	0.91	0.8
30	Health service	3.8	3.7	4.3	1.23	0.7	1.40	0.5	1.06	0.4
31	Cleaning and building service	2.5	2.4	2.4	1.12	0.3	1.23	0.6	1.05	0.0
32	Personal service	2.1	2.5	2.9	1.14	1.0	1.30	0.5	0.94	-0.7
33	Mechanics and repairers	0.3	0.3	0.4	1.73	-0.1	1.91	0.4	1.48	0.5
34	Construction trades	0.1	0.1	0.2	1.55	0.3	1.76	1.4	1.40	-0.2
35	Other precision prod, craft, & repair	1.7	1.7	1.4	1.41	0.1	1.67	0.7	1.24	0.0
36	Machine opertrs and tenders,exc precis	5.5	4.0	2.5	1.24	0.6	1.38	1.6	1.08	-0.3
37	Fabricatrs,assemblrs,inspectrs,samplrs	2.4	2.2	1.5	1.34	0.0	1.46	0.6	1.11	-0.8
38	Motor vehicle operators	0.6	0.7	0.9	1.38	0.7	1.48	1.1	1.28	1.1
39	Other transp & material moving occs	0.1	0.1	0.1	1.58	0.5	1.72	0.0	1.43	-0.9
40	Construction laborers	0.0	0.0	0.1	1.42	0.4	1.62	2.0	1.26	-1.8
41	Freight, stock & materials handlers	0.6	0.6	0.8	1.20	0.3	1.30	-0.2	1.07	-1.1
42	Other handlers,equip. cleaners, helpers, laborers	1.1	1.0	1.1	1.24	-0.4	1.26	0.4	1.01	-0.3
43	Farm operators and managers	0.0	0.0	0.1	1.16	3.3	1.42	0.7	1.10	0.3
44	Farm workers and related occupations	0.9	0.6	0.6	1.00	0.5	1.18	-0.1	0.93	-0.1
45	Forestry and fishing occs	0.0	0.0	0.0	1.32	1.5	1.39	4.2	1.00	-7.9
	Total	100	100	100	1.49	1.13	1.63	1.21	1.37	1.54

Note: Source CPS-ORG.

2.4.1 The ORU approach

The now standard approach to test whether non-required years of schooling have different returns involves estimating a modified Mincerian earnings equation[10]:

$$\ln W_i = X_i \cdot \gamma + \begin{bmatrix} Q_i^r & Q_i^s & Q_i^u \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \epsilon_i \tag{2.1}$$

W_i represents an individual's earnings, which are assumed to have a log-normal distribution (ϵ_i is the random part) conditional on a vector of personal characteristics X_i (including a constant) and qualifications Q_i. The vectors of parameters to be estimated are γ and β. The qualifications variable has three components: required (r), surplus (s), and deficit (u) qualifications[11] Each of these qualifications variables are measured in years of formal education.

We first estimate equation 2.1 treating each year of data as a separate cross section. Figures 2.5 and 2.6 present the estimates for the returns to required, deficit, and surplus schooling for males and females respectively. For comparison purposes we include estimates from a traditional Mincerian equation, where the qualifications variable is not decomposed. Returns to education increased during the 80s and then fell toward the end of the decade, recovering only slowly during the 90s. The pattern is very similar when the education variable is decomposed. The striking result, however, is how much higher the returns are for required years of education with respect to surplus or (the absolute value of) deficit years. In fact, the usual Mincerian returns can be seen as a weighted average between required, surplus and deficit years. The usual approach would lead us to conclude that an extra year of education increases

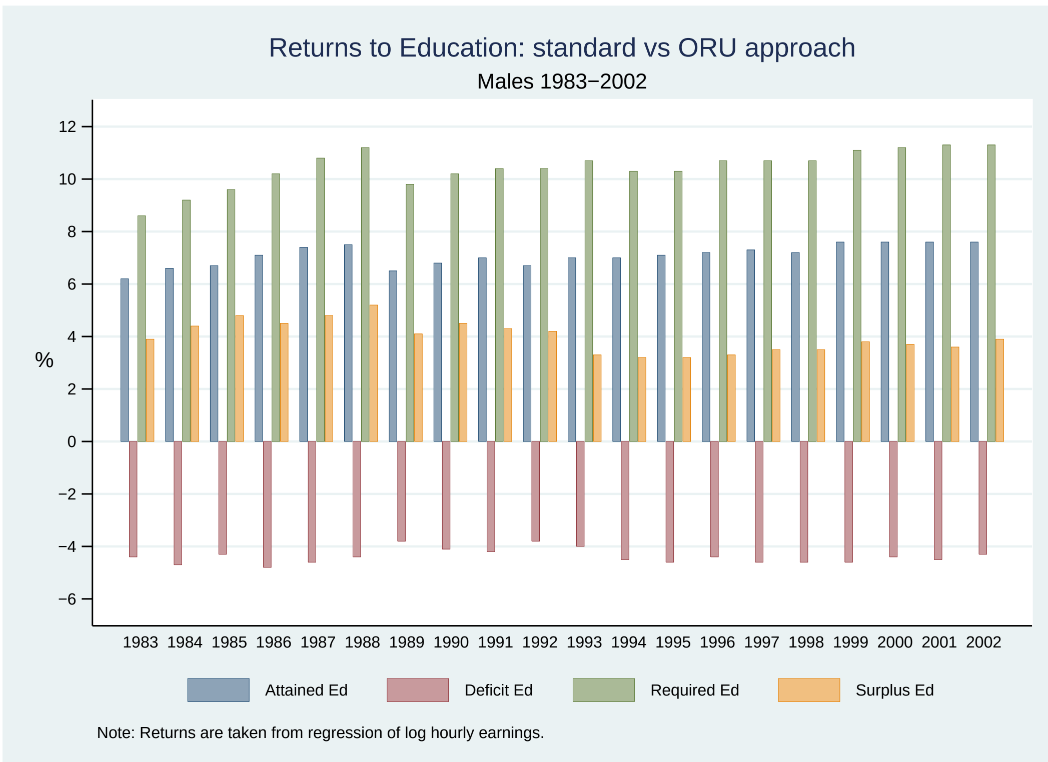
---

[10]This approach was first developed in Duncan and Hoffman (1981).

[11]The standard Mincerian approach would correspond to the particular case where $\beta_1 = \beta_2 = \beta_3$, so that required, surplus, and deficit education all receive the same return. The other particular case of note corresponds to Thurow's (1975) job competition model, where $\beta_2 = \beta_3 = 0$.

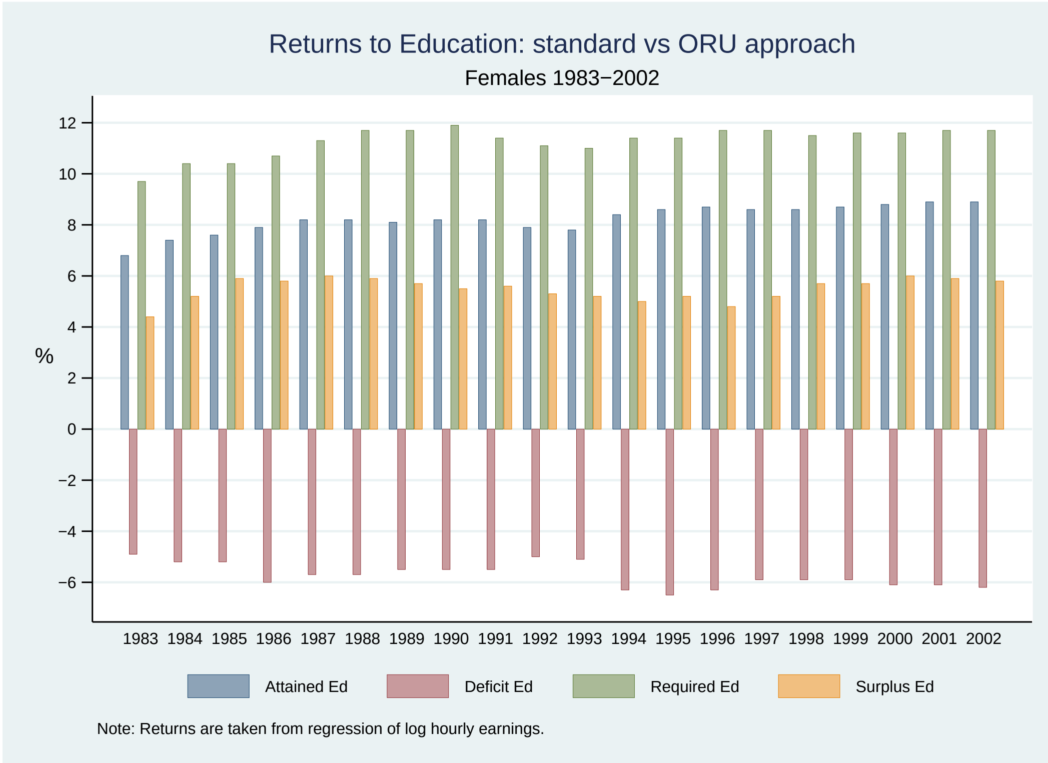
earnings by around 7 or 8 percent for males and females respectively. The Duncan-Hoffman approach, however, would make such judgement conditional on whether the extra year is required on the job or not. If the extra year is required, the increase in earnings will be between 10 and 12 percent. For surplus years, however, the return is only 4 or 6 percent, again for males and females respectively. Similarly, adding one year of education has low returns for under-educated workers.

Figure 2.5.



Tables 2.11 through 2.14 have detailed regression results for the Duncan-Hoffman equation. The controls include a quartic on age, minority status, part-time status, an indicator for married individuals, union contract, as well as geographic, manufacturing and services dummies. Tables 2.15 and 2.16, in turn, explore the evolution of returns for different sub-populations. Estimates of equation 2.1 confirm the usual finding regarding the concave form of the age-earnings profile. The impact of education on earnings is felt more strongly among workers age 35 and older, especially for males.

Figure 2.6.



Regression results include the usual negative estimate for the non-white dummy. Non-whites also have higher estimates for the traditional returns to attained education in the sub-group regressions. The result disappears, however, when the education variable is decomposed. Returns to required education are *lower* for non-whites (the exception are the 1992 estimates for males, which are practically identical for both groups). The under-education discount for non-whites is around 5 percent for males and between 5 and 7 percent for females. Considering the high rates of under-education among minority workers, further investments in human capital would pay off even if mismatch remained at the same levels.

During the period 1983–2002 the shift of employment from manufacturing to services has deepened (see again tables 2.3 and 2.4). For male workers the returns to education are higher in manufacturing, especially for correctly matched workers. In the case of women, manufacturing has higher returns only in the Duncan-Hoffman

Table 2.11. ORU Equation Estimation for Males: 1983–1992

	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Age	0.083	0.043	0.084	0.088	0.046	0.041	0.092	0.071	0.048	0.051
	[0.015]**	[0.019]*	[0.014]**	[0.012]**	[0.015]**	[0.014]**	[0.015]**	[0.011]**	[0.013]**	[0.012]**
$Age^2/100$	-0.129	-0.009	-0.151	-0.163	-0.011	-0.008	-0.187	-0.116	-0.028	-0.036
	[0.057]*	[0.073]	[0.055]**	[0.046]**	[0.057]	[0.052]	[0.057]**	[0.042]**	[0.051]	[0.046]
Age^3	0	0	0	0	0	0	0	0	0	0
	[0.000]	[0.000]	[0.000]	[0.000]*	[0.000]	[0.000]	[0.000]*	[0.000]	[0.000]	[0.000]
Age^4	0	0	0	0	0	0	0	0	0	0
	[0.000]	[0.000]	[0.000]	[0.000]*	[0.000]	[0.000]	[0.000]*	[0.000]	[0.000]	[0.000]
Deficit Qualif	-0.044	-0.047	-0.043	-0.048	-0.046	-0.044	-0.038	-0.041	-0.042	-0.038
	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**
Required Qualif	0.086	0.092	0.096	0.102	0.108	0.112	0.098	0.102	0.104	0.104
	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**
Surplus Qualif	0.039	0.044	0.048	0.045	0.048	0.052	0.041	0.045	0.043	0.042
	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**
Nonwhite	-0.165	-0.177	-0.191	-0.173	-0.173	-0.173	-0.158	-0.158	-0.16	-0.161
	[0.005]**	[0.005]**	[0.005]**	[0.005]**	[0.005]**	[0.006]**	[0.005]**	[0.005]**	[0.005]**	[0.005]**
PT, economic reasons	-0.128	-0.186	-0.178	-0.198	-0.203	-0.173	-0.176	-0.175	-0.156	-0.184
	[0.009]**	[0.011]**	[0.010]**	[0.011]**	[0.011]**	[0.013]**	[0.011]**	[0.011]**	[0.009]**	[0.010]**
Union	0.185	0.184	0.185	0.184	0.177	0.176	0.169	0.157	0.156	0.162
	[0.004]**	[0.004]**	[0.004]**	[0.004]**	[0.004]**	[0.005]**	[0.005]**	[0.004]**	[0.004]**	[0.005]**
Married	0.148	0.144	0.142	0.141	0.148	0.158	0.123	0.124	0.125	0.128
	[0.005]**	[0.006]**	[0.005]**	[0.005]**	[0.005]**	[0.005]**	[0.005]**	[0.004]**	[0.004]**	[0.004]**
Constant	-1.441	-1.009	-1.497	-1.637	-1.246	-1.208	-1.48	-1.298	-1.151	-1.215
	[0.131]**	[0.173]**	[0.132]**	[0.114]**	[0.136]**	[0.129]**	[0.137]**	[0.106]**	[0.125]**	[0.118]**
Observations	66867	63488	70492	77369	74775	70819	83118	86837	83617	82177
R-squared	0.42	0.41	0.42	0.42	0.43	0.43	0.4	0.41	0.42	0.41

Notes: Robust standard errors in brackets. * significant at 5%, ** significant at 1%. Geographic and econ sector controls not reported.

Table 2.12. ORU Equation Estimation for Males: 1993–2002

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Age	0.058	0.007	0.009	0.033	0.062	0.047	0.032	0.025	0.039	0.036
	[0.012]**	[0.012]	[0.017]	[0.015]*	[0.016]**	[0.015]**	[0.017]	[0.014]	[0.015]**	[0.017]*
$Age^2/100$	-0.044	0.147	0.12	0.026	-0.083	-0.035	0.004	0.032	-0.007	0.003
	[0.045]	[0.042]**	[0.064]	[0.058]	[0.060]	[0.055]	[0.063]	[0.053]	[0.054]	[0.065]
Age^3	0	0	0	0	0	0	0	0	0	0
	[0.000]	[0.000]**	[0.000]**	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Age^4	0	0	0	0	0	0	0	0	0	0
	[0.000]	[0.000]**	[0.000]**	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Deficit Qualif	-0.04	-0.045	-0.046	-0.044	-0.046	-0.046	-0.046	-0.044	-0.045	-0.043
	[0.001]**	[0.001]**	[0.002]**	[0.002]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**
Required Qualif	0.107	0.103	0.103	0.107	0.107	0.107	0.111	0.112	0.113	0.113
	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**
Surplus Qualif	0.033	0.032	0.032	0.033	0.035	0.035	0.038	0.037	0.036	0.039
	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**
Nonwhite	-0.149	-0.149	-0.172	-0.182	-0.176	-0.173	-0.172	-0.158	-0.162	-0.158
	[0.005]**	[0.005]**	[0.006]**	[0.006]**	[0.006]**	[0.006]**	[0.006]**	[0.006]**	[0.006]**	[0.006]**
PT, economic reasons	-0.197	-0.18	-0.181	-0.195	-0.204	-0.174	-0.155	-0.168	-0.159	-0.214
	[0.010]**	[0.013]**	[0.015]**	[0.014]**	[0.015]**	[0.016]**	[0.016]**	[0.017]**	[0.014]**	[0.017]**
Union	0.169	0.192	0.19	0.216	0.217	0.212	0.201	0.2	0.19	0.196
	[0.005]**	[0.005]**	[0.006]**	[0.006]**	[0.006]**	[0.006]**	[0.006]**	[0.006]**	[0.006]**	[0.006]**
Married	0.124	0.121	0.134	0.141	0.14	0.139	0.135	0.132	0.138	0.143
	[0.004]**	[0.005]**	[0.006]**	[0.005]**	[0.005]**	[0.005]**	[0.005]**	[0.006]**	[0.006]**	[0.006]**
Constant	-1.371	-0.785	-0.773	-1.049	-1.293	-1.065	-0.904	-0.851	-0.986	-0.913
	[0.115]**	[0.120]**	[0.169]**	[0.147]**	[0.155]**	[0.145]**	[0.163]**	[0.139]**	[0.145]**	[0.164]**
Observations	80475	80577	53129	55652	56446	56108	53991	53060	55518	59745
R-squared	0.41	0.37	0.37	0.4	0.39	0.39	0.4	0.39	0.38	0.37

Notes: Robust standard errors in brackets. * significant at 5%, ** significant at 1%. Geographic and econ sector controls not included.

Table 2.13. ORU Equation Estimation for Females: 1983–1992

	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Age	0.244	0.246	0.217	0.227	0.192	0.185	0.165	0.164	0.163	0.172
	[0.017]**	[0.013]**	[0.016]**	[0.014]**	[0.015]**	[0.012]**	[0.012]**	[0.012]**	[0.011]**	[0.013]**
$Age^2/100$	-0.752	-0.761	-0.65	-0.683	-0.547	-0.521	-0.456	-0.457	-0.443	-0.466
	[0.066]**	[0.051]**	[0.063]**	[0.054]**	[0.057]**	[0.048]**	[0.045]**	[0.045]**	[0.042]**	[0.050]**
Age^3	0	0	0	0	0	0	0	0	0	0
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**
Age^4	0	0	0	0	0	0	0	0	0	0
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**
Deficit Qualif	-0.049	-0.052	-0.052	-0.06	-0.057	-0.057	-0.055	-0.055	-0.055	-0.05
	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**
Required Qualif	0.097	0.104	0.104	0.107	0.113	0.117	0.117	0.119	0.114	0.111
	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**
Surplus Qualif	0.044	0.052	0.059	0.058	0.06	0.059	0.057	0.055	0.056	0.053
	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**
Nonwhite	-0.072	-0.083	-0.07	-0.076	-0.084	-0.071	-0.069	-0.081	-0.075	-0.08
	[0.005]**	[0.005]**	[0.005]**	[0.005]**	[0.005]**	[0.005]**	[0.005]**	[0.004]**	[0.004]**	[0.005]**
PT, economic reasons	-0.206	-0.219	-0.232	-0.234	-0.229	-0.26	-0.206	-0.211	-0.22	-0.241
	[0.006]**	[0.008]**	[0.008]**	[0.007]**	[0.008]**	[0.009]**	[0.008]**	[0.008]**	[0.007]**	[0.008]**
Union	0.165	0.165	0.154	0.155	0.155	0.145	0.125	0.128	0.11	0.105
	[0.005]**	[0.005]**	[0.005]**	[0.005]**	[0.005]**	[0.006]**	[0.005]**	[0.005]**	[0.005]**	[0.005]**
Married	-0.001	-0.006	-0.002	-0.004	0	-0.001	0.009	0.016	0.01	0.012
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.005]	[0.004]*	[0.004]**	[0.004]*	[0.004]**
Constant	-2.897	-2.967	-2.809	-2.894	-2.676	-2.659	-2.397	-2.386	-2.367	-2.502
	[0.150]**	[0.125]**	[0.149]**	[0.132]**	[0.137]**	[0.119]**	[0.114]**	[0.114]**	[0.113]**	[0.128]**
Observations	59767	56496	63452	70964	69715	66236	76695	80895	79036	78068
R-squared	0.37	0.36	0.36	0.36	0.37	0.38	0.37	0.37	0.36	0.37

Notes: Robust standard errors in brackets. * significant at 5%, ** significant at 1%. Geographic and econ sector controls not included.

Table 2.14. ORU Equation Estimation for Females: 1993–2002

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Age	0.157	0.108	0.101	0.13	0.075	0.035	0.066	0.055	0.066	0.131
	[0.011]**	[0.011]**	[0.019]**	[0.018]**	[0.013]**	[0.013]**	[0.016]**	[0.014]**	[0.017]**	[0.017]**
$Age^2/100$	-0.399	-0.22	-0.215	-0.32	-0.12	0.014	-0.119	-0.075	-0.106	-0.348
	[0.040]**	[0.038]**	[0.072]**	[0.070]**	[0.046]**	[0.045]	[0.058]*	[0.052]	[0.063]	[0.061]**
Age^3	0	0	0	0	0	0	0	0	0	0
	[0.000]**	[0.000]**	[0.000]	[0.000]**	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]**
Age^4	0	0	0	0	0	0	0	0	0	0
	[0.000]**	[0.000]	[0.000]	[0.000]**	[0.000]	[0.000]*	[0.000]	[0.000]	[0.000]	[0.000]**
Deficit Qualif	-0.051	-0.063	-0.065	-0.063	-0.059	-0.059	-0.059	-0.061	-0.061	-0.062
	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**
Required Qualif	0.11	0.114	0.114	0.117	0.117	0.115	0.116	0.116	0.117	0.117
	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**
Surplus Qualif	0.052	0.05	0.052	0.048	0.052	0.057	0.057	0.06	0.059	0.058
	[0.002]**	[0.002]**	[0.002]**	[0.003]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**
Nonwhite	-0.079	-0.088	-0.087	-0.099	-0.091	-0.087	-0.084	-0.083	-0.093	-0.096
	[0.005]**	[0.005]**	[0.006]**	[0.006]**	[0.005]**	[0.005]**	[0.006]**	[0.006]**	[0.006]**	[0.006]**
PT, economic reasons	-0.224	-0.198	-0.227	-0.212	-0.209	-0.205	-0.198	-0.184	-0.223	-0.21
	[0.007]**	[0.010]**	[0.012]**	[0.010]**	[0.010]**	[0.011]**	[0.012]**	[0.013]**	[0.013]**	[0.012]**
Union	0.107	0.119	0.143	0.151	0.136	0.122	0.135	0.102	0.106	0.098
	[0.005]**	[0.006]**	[0.007]**	[0.007]**	[0.007]**	[0.007]**	[0.007]**	[0.007]**	[0.007]**	[0.007]**
Married	0.015	0.013	0.013	0.009	0.015	0.015	0.011	0.007	0.017	0.019
	[0.004]**	[0.004]**	[0.005]*	[0.005]	[0.005]**	[0.005]**	[0.005]*	[0.005]	[0.005]**	[0.005]**
Constant	-2.363	-1.856	-1.774	-2.118	-1.574	-1.102	-1.39	-1.181	-1.472	-2.019
	[0.115]**	[0.112]**	[0.188]**	[0.176]**	[0.125]**	[0.124]**	[0.164]**	[0.139]**	[0.168]**	[0.163]**
Observations	77320	77178	51289	54234	55259	54698	52864	51736	54134	58942
R-squared	0.36	0.33	0.34	0.36	0.37	0.36	0.36	0.36	0.35	0.36

Notes: Robust standard errors in brackets. * significant at 5%, ** significant at 1%. Geographic and econ sector controls not included.

Table 2.15. ORU Estimates for Sub-groups: Males

	1983				1992				2002			
	Attained Ed	Deficit Ed	Required Ed	Surplus Ed	Attained Ed	Deficit Ed	Required Ed	Surplus Ed	Attained Ed	Deficit Ed	Required Ed	Surplus Ed
Age												
16–24	3.21	-2.85	5.16	-0.46	3.05	-1.89	5.37	2.00	3.91	-2.47	6.83	1.41
25–34	5.25	-4.01	6.93	2.97	6.05	-3.49	9.02	3.43	6.78	-4.14	9.53	4.46
35–54	7.06	-4.90	9.72	5.34	7.52	-4.29	11.14	4.97	8.38	-4.55	12.17	4.37
55–	5.72	-3.34	9.62	4.72	6.65	-3.66	11.77	3.66	7.73	-4.02	12.19	2.58
Non-white	6.51	-4.90	8.42	3.99	7.60	-4.53	10.38	4.25	8.48	-5.04	10.84	3.85
White	5.04	-3.48	9.27	3.42	5.26	-3.22	10.31	3.82	6.58	-3.57	12.43	4.13
Manufacturing	7.60	-5.01	12.22	4.82	7.66	-4.35	12.84	5.13	8.55	-4.64	13.26	5.36
Services	5.61	-4.36	7.31	3.35	6.59	-3.91	9.66	3.75	7.61	-4.43	10.84	3.48
Non-union	6.87	-4.62	9.81	4.58	7.15	-3.81	11.36	4.50	8.07	-4.24	12.35	4.05
Union	3.31	-3.45	3.89	1.28	3.97	-3.16	5.25	2.27	3.88	-3.13	4.63	2.63

Table 2.16. ORU Estimates for Sub-groups: Females

	1983				1992				2002			
	Attained Ed	Deficit Ed	Required Ed	Surplus Ed	Attained Ed	Deficit Ed	Required Ed	Surplus Ed	Attained Ed	Deficit Ed	Required Ed	Surplus Ed
Age												
16–24	4.83	-5.46	8.58	0.81	4.98	-3.89	8.69	2.70	5.53	-4.66	7.78	2.38
25–34	7.08	-5.17	9.38	5.05	8.32	-5.38	11.43	5.96	9.00	-6.29	11.54	5.64
35–54	7.21	-4.85	10.11	5.14	8.29	-5.30	11.26	5.71	9.37	-6.37	12.27	6.48
55–	5.95	-4.17	9.86	3.70	6.67	-4.32	11.14	2.92	8.09	-5.50	11.23	4.52
Non-white	7.04	-5.34	9.64	4.17	8.41	-5.46	11.10	5.32	9.61	-7.24	11.58	6.06
White	6.12	-4.08	10.11	5.23	6.82	-4.46	11.70	5.17	7.91	-5.29	12.49	5.16
Manufacturing	6.41	-4.11	11.38	6.27	7.79	-4.62	14.05	7.26	8.56	-4.83	14.24	7.46
Services	6.83	-5.23	9.44	3.97	7.90	-5.14	10.85	4.98	9.05	-6.50	11.58	5.65
Non-union	7.07	-5.15	10.76	4.52	8.15	-5.27	12.31	5.35	9.17	-6.28	12.53	6.01
Union	5.83	-3.96	7.34	3.82	6.40	-4.16	7.55	4.15	7.40	-5.33	8.52	3.76

approach. Thus, the shift to services might offer at least a partial explanation for why the returns to education have not grown faster during the period. Finally, it is interesting to note how low the returns are in the union sector. Moreover, it is in this sector where the returns to attained and required education come closer to each other. A plausible conclusion is that the Mincerian model is more relevant in relatively more unionized economies, with lower mismatch rates and low returns to surplus and deficit education.

2.5 Panel Estimation

A remarkable feature of previous empirical studies that examine the effects on earnings of educational mismatch is the robustness of their findings, which seem to hold across different time periods and different countries. A potential problem of many of the existing studies however is that they employ only cross-section data. It is thus possible that the results of these studies, as well as the results presented here thus far, are biased due to unobserved heterogeneity of individuals. Suppose the real population level equation is given by:

$$\ln W_{i,t} = X_{i,t} \cdot \gamma + \begin{bmatrix} Q^r_{i,t} & Q^s_{i,t} & Q^u_{i,t} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + c_i + \epsilon_{i,t} \tag{2.2}$$

where, as before, W, X and Q represent earnings, individual characteristics and qualifications and where we have added c—an unobservable individual-level variable (typically thought of as individual ability level). Because c is usually correlated with the other regressors, omitting it from the estimating equation—as in equation 2.1—introduces bias in the results. In our particular case, individuals with lower innate ability might need more education to successfully perform at a job for which they are formally over-educated. In this case, the returns to surplus and deficit education in equation 2.1 will be underestimated in absolute value. In other words, if we controlled for individual heterogeneity, we should expect the return to deficit, required and surplus schooling to become closer to each other (and potentially the same if the human capital model is true).

There are several possible solutions for the omitted variable problem, including the use of proxy and instrumental variables. In addition, if we can observe the same cross-section units at different points in time, then it is possible consistently to estimate the γ and β without having to further deal with the unobservable effect (Wooldridge, 2002, ch. 10). In this subsection, we estimate equation 2.2 using pooled OLS, random and fixed effects panel estimators.

We exploit the particular time structure of the CPS interview system to construct a weakly balanced panel for the years 1983–2002. CPS respondents are interviewed for 4 consecutive months, then kept out of the sample for 8 months, and finally reinserted for 4 months before leaving the sample permanently. Attrition problems aside, this design should lead to a 75% overlap of respondents across consecutive months, and a 50% between the same month of consecutive years. However, because only the outgoing rotation groups (month in sample 4 and 8) are asked earnings-related questions, each year the maximum possible panel size is one eighth of the total CPS sample. This maximum is substantially above the actual fraction of individual respondents that can be matched across survey years. There are several reasons for this. On the one hand, there is attrition in the sample due to non-response, mortality, and migration. On the other hand, there is the unfortunately important issue of recording errors.

Table 2.17. Merge rates and Sample Restrictions in CPS-ORG Panel

Year	Full	Merged	Matched	Employed	Non-student	Wage-Salary	Unedited
1983	174,141	76	71	37	.	31	24
1984	170,053	37	35	19	18	15	12
1985	171,145	18	17	10	9	8	6
1986	168,835	74	69	38	36	31	25
1987	168,763	70	63	36	34	29	22
1988	160,780	73	68	39	37	31	26
1989	163,171	75	70	40	38	32	31
1990	169,257	75	70	40	38	32	30
1991	166,151	75	56	31	30	26	24
1992	164,138	75	60	32	31	26	25
1993	162,699	74	56	31	29	25	24
1994	157,540	29	23	12	12	10	10
1995	158,307	24	19	10	9	8	2
1996	139,473	79	63	34	33	27	18
1997	140,702	78	71	42	40	34	22
1998	140,416	78	65	38	36	31	19
1999	141,527	79	65	38	36	31	18
2000	138,808	79	65	38	37	31	17
2001	149,939	79	61	37	36	31	17

Columns 3–8 expressed as fraction of full sample.

Beginning in 1980, individuals at a point in time are uniquely identified in the CPS by two variables: a household identified and an individual line number within the household (for details see Madrian and Lefgren, 1999). In theory, these two identifiers should remain constant over time. This is not true, however, for the same individual *across time* because the same identifier might be given to a different person in case the original respondent moves away from the housing unit. A third variable is supposed to register the cases when a new household is interviewed. But recording errors in the latter variable are very common. Because of these recording errors, matching individuals across CPS years results both in “false positives”—matches that do not represent the same individual across time—and “false negatives”—matches that are not made even when they should have.

Table 2.17 presents statistics on the matching process for the ORG files 1983–2002. Column 2 has the starting sample size. The other columns present the percentage of the initial sample that remains after all previous restrictions apply. The "naïve merge rate" in column 3 represents all the merges that occur by matching observations from contiguous years solely on the basis of the three identifiers. These merge rates are within the range of success of previous studies, with the exception of years 1994-95 (CPS overhaul) and 1984-85. To eliminate the false positives[12], we discard from the sample naïve matches for which the sex or race is different across time, or for which education and age increase more than one or two years respectively. Eliminating this estimated false positives leaves us with the match rate in column 4. The other sample restrictions mirror the ones in the previous section.

Table 2.18. Panel descriptive stats

	Males				Females			
	Mean	SD overall	SD between	SD within	Mean	SD overall	SD between	SD within
Log Earnings	1.83	0.58	0.53	0.23	1.52	0.53	0.49	0.21
Education Attainment	13.25	2.86	2.86	0.13	13.31	2.57	2.57	0.14
Required Education	13.49	2.13	2.05	0.55	13.25	2.02	1.94	0.57
Surplus Ed (if overed)	3.05	1.08	1.02	0.32	3.05	1.11	1.07	0.30
Deficit Ed (if undered)	3.94	1.96	1.85	0.28	3.62	1.89	1.76	0.26

The resulting panel is weakly balanced. It is balanced because each individual is observed in exactly two periods. It is only weakly balanced because the periods are not the same for every individual. In table 2.18 we present descriptive statistics for the key variables. Of particular importance is the standard deviation of the education variables within panels, since this is the source of identification for the fixed effect estimator.

Table 2.19 further explores this aspect of the data. 88% of all individuals do not change their match status over the two periods. As expected, the most frequent case

[12]Not much can be done about the false negatives.

is that of workers who are correctly matched in both periods. The data do not tell us whether individuals have changed jobs from one year to the next. We take a change in occupation or industry as an (admittedly imperfect) indicator of job transition. Not surprisingly the fraction of those whose match status changes over time is higher in the latter case. Among the possible match status changes, transitions in and out of correctly matched are also the most frequent. Other reassuring results involve the high under-education and the near-zero over-education rate for high-schoolers and dropouts, and the opposite result for workers with advanced degrees. Finally, note the very high rate of permanence in union jobs.

Table 2.19. Transitions in Match Status

	No Change			Transitions						
	Matched (%)	Overeduc (%)	Undereduc (%)	M-0 (%)	U-M (%)	O-M (%)	M-U (%)	O-U (%)	U-O (%)	Obs
All Transitions										
All workers	67.2	10.02	10.39	3.4	3.07	2.78	2.88	0.13	0.12	758,775
Male	67.44	8.55	12	3.11	3.2	2.54	3.02	0.07	0.07	401,515
Female	66.93	11.69	8.59	3.73	2.92	3.06	2.72	0.19	0.18	357,260
Transitions with occ or ind change										
All workers	61.69	9.02	8.57	5.35	4.99	4.86	5.05	0.23	0.23	400,698
Male	62.43	7.46	10.24	4.83	5.22	4.35	5.22	0.12	0.13	216,953
Female	60.82	10.86	6.6	5.97	4.73	5.46	4.85	0.36	0.35	183,745
Transitions by Age group										
16-24	75.07	6.41	6.51	2.86	3.63	2.25	3.01	0.16	0.12	63,309
25-44	67.2	11.24	8.49	3.81	2.96	3.17	2.85	0.14	0.13	427,022
45-66	65.66	8.99	13.79	2.93	3.13	2.35	2.93	0.1	0.12	251,517
65-	60.58	8.21	22.31	2.32	2.72	1.5	2.28	0.05	0.03	16,927
Transitions by Education Level										
LTHS	38.69	0	52.91	0.01	4.88	0.01	3.51	0	0	86,429
HS	85.95	0	6.58	0.01	3.51	0.01	3.94	0	0	305,443
Some	61.97	13.01	6.06	6.74	4.18	3.96	3.39	0.33	0.37	166,317
College	54.63	25.75	2.26	7.64	1.02	7.36	0.83	0.3	0.23	128,216
Advanced	56.38	29.58	0.07	6.62	0.13	7.02	0.11	0.06	0.02	72,369
Transitions by Minority status										
White	68.32	10.16	8.88	3.48	3.1	2.91	2.9	0.13	0.12	596,276
Non-white	63.08	9.53	15.93	3.11	2.96	2.33	2.79	0.14	0.13	162,499
Transitions by Union status										
Non-union	65.53	10.4	10.85	3.64	3.28	3	3.06	0.12	0.12	565,082
Union in 1 period only	67.78	9.02	9.89	3.46	3.29	2.93	3.18	0.25	0.2	70,125
Union in both periods	74.49	8.89	8.57	2.3	1.97	1.7	1.88	0.09	0.1	123,568

What is the effect of these transitions on earnings? In table 2.20 we see that workers who did not change their match status—whether they changed their jobs or not—enjoyed average raises of between 1.4 and 3.8 percent yearly. The picture is completely different for match status transitions. Moving into (out of) over-education is generally accompanied with a penalty (prize) on earnings. The opposite is true about under-education. These results generally confirm our findings from the cross-section regressions.

Table 2.20. Changes in log earnings by transition type

	No Change			Transitions					
	Matched	Overed	Undered	M-0	U-M	O-M	M-U	O-U	U-O
All Transitions									
All	0.027	0.032	0.019	-0.013	-0.021	0.092	0.080	0.162	-0.086
Male	0.026	0.027	0.017	-0.013	-0.029	0.086	0.076	0.174	-0.105
Female	0.030	0.036	0.022	-0.013	-0.011	0.097	0.085	0.158	-0.078
Transitions between different industries or occupations									
All	0.030	0.035	0.016	-0.019	-0.028	0.098	0.082	0.171	-0.087
Male	0.028	0.032	0.014	-0.019	-0.036	0.090	0.077	0.183	-0.105
Female	0.033	0.038	0.020	-0.020	-0.018	0.106	0.089	0.166	-0.079

Next we estimate equation 2.2. Table 2.21 presents results for pooled OLS, random effects and fixed effects estimations. The first two estimators yield results that are very much in line with the cross-section OLS estimates. However, the standard Hausman test decisively rejects the null hypothesis that the random effects estimator is consistent[13].

Table 2.21. Panel Regression Results

	Females				Males			
	(1) Pooled OLS	(2) RE	(3) FE	(4) FE restrict	(1) Pooled OLS	(2) RE	(3) FE	(4) FE restrict
Attained Ed	0.088***	0.088***	0.012***	0.018***	0.075***	0.075***	0.006**	0.010***
	[0.000]	[0.000]	[0.002]	[0.003]	[0.000]	[0.000]	[0.003]	[0.004]
Age	0.044***	0.043***	0.053***	0.064***	0.058***	0.063***	0.062***	0.068***
	[0.000]	[0.000]	[0.004]	[0.006]	[0.000]	[0.000]	[0.004]	[0.006]
Age2/100	-0.047***	-0.046***	-0.058***	-0.072***	-0.061***	-0.066***	-0.074***	-0.088***
	[0.000]	[0.000]	[0.003]	[0.004]	[0.000]	[0.000]	[0.003]	[0.004]
Deficit Ed	-0.059***	-0.068***	-0.012***	-0.021***	-0.045***	-0.057***	-0.003	-0.008**
	[0.000]	[0.000]	[0.003]	[0.004]	[0.000]	[0.000]	[0.003]	[0.004]
Required Ed	0.114***	0.115***	0.025***	0.027***	0.106***	0.107***	0.017***	0.017***
	[0.000]	[0.000]	[0.003]	[0.003]	[0.000]	[0.000]	[0.003]	[0.004]
Surplus Ed	0.062***	0.069***	0.004	0.010***	0.046***	0.056***	-0.001	0.004
	[0.000]	[0.001]	[0.003]	[0.003]	[0.001]	[0.001]	[0.003]	[0.004]
Observations	647632	684266	684266	294606	694398	734608	734608	329538
R-squared	0.351	.	0.012	0.022	0.370	.	0.010	0.017
Number of id		342810	342810	147643		367697	367697	164930

Notes: Standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Regressions include controls for married, part time, public sector, manufacturing, services, union, and year dummies. Pooled OLS also includes nonwhite, metropolitan area and region dummies.

The fixed effects estimator confirms the basic result found in the recent literature on skill mismatch. The returns to surplus and deficit qualifications are smaller in absolute value than the returns to required qualifications. Indeed, deficit education for females and surplus education in general are not statistically different from zero,

[13] The test-statistic is 11,891 for males and 12,890 for females. It is distributed under the null with a chi-square with 11 degrees of freedom (the test was conducted without year dummies to avoid a not positive-definite covariance matrix).

which is exactly the result predicted by Thurow's (1975) model. One problem with the fixed effects results involves the very low returns to required education. The reason for the low estimate is the insufficient within-panel variation of the education variables (see again table 2.18). The problem is even more pronounced for a standard returns to attained education because there is a lack of variation due to changes in requirements (we include estimates from the standard Mincer equation in the table for comparison purposes). The fourth and eighth columns contain estimates for the occupation or industry change restricted sample. Estimates increase in size somewhat and the surplus schooling estimate is now significant for females. But the overall result is the same.

Bauer (2002) and Tsai (2007) conduct similar analysis using panel data[14]. In both cases, the fixed effects estimator results in returns to surplus and deficit schooling that are very close to the returns to required schooling. The conclusion that is drawn is that, once individual heterogeneity is controlled for, the result common in the literature on skill mismatch disappears and the human capital model is revalidated. In the Bauer study the result is based on a modal measure of mismatch, according to which all individuals below or above the modal education level in a 2-digit occupation are considered mismatched. Of course, the resulting rates of mismatch are extremely high: 51% for males and 67% for females. If a majority of workers is mismatched, it is not surprising that returns to surplus and deficit education approach the returns to modal education[15]. The Tsai study is specially interesting for comparison since it is based on a U.S. survey for a very similar period. Her fixed effects estimates for the returns to surplus and deficit schooling are very close to the returns to required

[14]Bauer uses a German panel dataset for the period 1984–1998. Tsai uses the U.S. Panel Study of Income Dynamics for the period 1979–2005.

[15]Bauer also estimates the model with a mean-plus-one-standard-deviation measure. In this case the fixed effects model yields an estimate for the returns to required schooling that is not significantly different from zero.

schooling: 0.005 below and 0.01 above respectively. The estimate for the returns to required education is quite low at 2%[16]. It is clear that all fixed effects estimations of earnings equations face the challenge of low within variation of the education variable.

2.6 Conclusion

The rapidly growing proportion of college and post-college educated workers in the labor force implies a challenge to the labor market's ability to adapt. The literature on skill mismatch suggests that (1) an important proportion of workers possess formal training substantially different from the one required on the job and (2) the average returns to deficit and surplus years of schooling are lower in absolute value than the returns to schooling that is required on the job. In this paper we estimate skill requirements with a one standard deviation range around the median years of education for each 3-digit occupation in the 1980 census classification. The analysis is conducted for males and females separately.

Our analysis indicates that over-education rates during the period were around 12% and 17% of the employed labor force for males and females respectively. Under-education rates were 16% and 12%. Regression estimates at the cross section level yield results that are quite close to those found in the literature. The returns to required schooling are substantially higher than the standard returns to attained education. More importantly, the returns to surplus and deficit schooling are very low in absolute value and represent only around 45% of the returns to required schooling. Thus, both the human capital model—which would predict equal returns to adequate, over- and under-education—and the job competition model—that would predict zero returns to surplus and deficit years—can be rejected.

---

[16] Tsai estimates the equation with a modal measure, obtaining practically the same results as with the mean measure.

If unobservable individual ability levels are correlated with observable qualifications, then OLS estimation is inconsistent. The estimated returns from equation 2.1 are thus suspect of bias due to an omitted variable problem. We work around this issue by merging consecutive CPS surveys into a panel. Pooled OLS and random effects estimators are in line with the cross section analysis. However, the Hausman specification test rejects the null of consistency of these estimators. While the fixed effects estimator is always consistent, we must face the problem of low within variation in individual qualification levels. The fixed effects estimates confirm the most important qualitative result in the literature, namely that surplus and deficit schooling have significantly lower returns. Nevertheless, the estimated returns are probably lower than the true values. All existing attempt to estimate ORU equations with a fixed effects estimator have to face the problem of low variation of the qualifications variables within panels. Because larger or longer panels are unlikely to offer substantially more variation, future research that tries to tackle the individual heterogeneity issue by means other than panel analysis (e.g. natural experiments) is warranted.

CHAPTER 3

EARNINGS INEQUALITY AND SKILL MISMATCH IN THE U.S: 1973–2002

3.1 Introduction

Over the last three decades wage inequality in the U.S. has increased. Several studies have focused on the rise of wage disparities among groups of workers defined by education attainment and experience in the labor force. Around a third of the variation in earnings at a point in time can be explained in this way (Levy and Murnane, 1992; Bound and Johnson, 1992; Katz and Autor, 1999). Also, *changes* in the wage distribution can be explained by the same factors to a significant extent (Fields, 2003). However, residual or within-group wage inequality—i.e wage dispersion among workers with the same education and experience—is generally believed to account for most of the increase in overall inequality. One possible explanation for the increase in residual inequality involves unobservable differences in human capital. If individuals differ in ability levels, then an increase in either the dispersion of those abilities or the rewards that accrue to them could account for the rise of inequality within groups. An alternative story not explored in the literature on wage inequality relies on the dispersion of outcomes within education groups because of the existence of skill mismatch.

According to assignment and other models, equilibrium in the labor market might be such that not all workers are allocated to jobs in which their skills are required[1].

[1]Assignment models are reviewed in Sattinger (1993). A skill mismatch equilibrium is also present in the search model in Albrecht and Vroman (2002) and the efficiency wage model in Skott (2006).

Some workers will be over-educated for their jobs, meaning that the skills they possess are above those required on the job. Similarly, some workers might have less qualifications than those required. Normally, over(under)-educated workers will have lower(higher) returns to human capital than correctly matched workers with the same levels of skill. An increase in these match differentials or in the overall rates of mismatch would certainly inflate the residual dispersion in traditional human capital regressions that restrict the returns to surplus and deficit qualifications to be the same as the returns to qualifications that are actually required on the job. The present study analyzes to what extent the *levels* of overall and within-group inequality can be explained when this type of skill mismatch is taken into consideration. The paper also considers how *changes* in skill mismatch and mismatch premia affect the wage distribution.

Using the method developed in Fields (2003), I show that the explanatory power of education in accounting for levels of earnings inequality is greater than what it would appear when skill mismatch is ignored. The differences are in the order of 5 percent, or almost 20 percent of the total explained variation in earnings. Surplus and deficit qualifications are roughly equally important in explaining inequality in the male wage distribution at a point in time, while surplus qualifications are more important for females. The paper also shows that the increase in over-education rates and premia in the last 30 years produced a very sharp increase in the relative importance of surplus qualifications in explaining wage dispersion. Indeed, when looking at changes in the wage distribution surplus qualifications are very important. Around 10 and 26 percent of the changes in the Gini coefficient for males and females respectively can be explained by increases in this factor alone. The contribution of deficit qualifications, however, is almost negligible (these are the main findings of the paper, which can be found in tables (3.8–3.9) in section 4.

Starting with Duncan and Hoffman (1981), the empirical literature on skill mismatch has been centered around the estimation of an equation of the form:

$$
\begin{aligned}
\ln W_{i,t} &= X_{i,t} \cdot \gamma_t + \begin{bmatrix} Q^r_{i,t} & Q^s_{i,t} & Q^d_{i,t} \end{bmatrix} \cdot \begin{bmatrix} \beta^r_t \\ \beta^s_t \\ \beta^d_t \end{bmatrix} + \epsilon_{i,t} = Z_{i,t} \cdot \alpha_t \qquad (3.1) \\
\alpha'_t &= \begin{bmatrix} \gamma'_t & \beta^r_t & \beta^s_t & \beta^d_{i,t} & 1 \end{bmatrix} \\
Z_{i,t} &= \begin{bmatrix} X_{i,t} & Q^r_{i,t} & Q^s_{i,t} & Q^d_{i,t} & \epsilon_{i,t} \end{bmatrix}
\end{aligned}
$$

where i and t index individuals and time respectively. $W_{i,t}$ represents earnings, which are assumed to have a log-normal distribution ($\epsilon_{i,t}$ is the random part) conditional on a vector of personal characteristics $X_{i,t}$ (including a constant and some function of age or experience) and qualifications $Q_{i,t}$. The vectors of parameters to be estimated are γ_t and β_t. The novelty of the approach involves splitting the qualifications variable into three parts: required (r), surplus (s), and deficit (d) qualifications.[2] For convenience, I also introduce here a more succinct notation—using one matrix (Z) and one vector of parameters (α) only—that will become useful later. Conditional on choosing and obtaining data for every job's educational requirements, equation (3.1) can be estimated using standard multivariate analysis.[3]

[2]Details on these qualifications variables are provided below. All on-the-job training is assumed to be required on the job so no decomposition applies in this case. A similar model has been estimated (Verdugo and Verdugo, 1989) that uses attained education and indicator variables for over- and under-educated workers in the right-hand-side instead of the required, deficit and surplus schooling variables. The latter model has been criticized because the returns to surplus and deficit schooling cannot be clearly identified (Cohn, 1992).

[3]The standard Mincerian approach would correspond to the particular case where $\beta^r_t = \beta^s_t = -\beta^d_t$, so that required, surplus, and deficit education all receive the same return. The other particular case of note corresponds to Thurow's (1975) job competition model, where $\beta^s_t = \beta^d_t = 0$.

There are several extensive surveys of studies that use this approach (Green et al., 1999; Hartog, 2000; Sloane, 2003; McGuinness, 2006). As a general rule, all studies tend to confirm Sicherman's (1991) stylized facts relating to the earnings of over- and under-educated workers:

1. The earnings of over-educated workers are less than the earnings of those who have the same level of education but are in jobs where those qualifications are required (e.g. a college graduate working at a grocery store earns less on average than a college graduate who is an investment banker).

2. Over-educated workers' earnings are however generally above the earnings of workers in their same occupation or job type, who are perfectly matched qualifications-wise (i.e., the college graduate in the grocery store tends to earn more than a high-school graduate occupying a similar position).

3. The earnings of under-educated workers are more than the earnings of those with the same level of education but who are perfectly matched (e.g. a high-school graduate who becomes a manager generally earns more than the average high-school graduate).

4. The co-workers of under-educated workers who have the appropriate formal training tend to earn more than them.

There is considerable variation in the estimates of the incidence of skill mismatch. Depending on the measure utilized, the country, the period, and data source, studies have found rates of over-education ranging from 10 to 42%, with an "un-weighted" average of 23.3% in the 25 studies summarized by Groot and Maassen van den Brink (2000). Their average for under-education is 14.4%.[4] Rubb (2003) provides a consistent meta-analysis of 85 estimates of the β parameters. The return to required

[4]The standard deviations are quite high: 9.9 and 8.2 percentage points respectively.

education is 9.6% on average. Each year of surplus schooling yields 5.2%. Finally, deficit qualifications take away 4.8% from the required education returns.

These figures seem significant enough to motivate the suspicions that (i) skill mismatch accounts for a significant part of earnings inequality, and (ii) changes in skill mismatch and match premia might have contributed to the observed changes in the wage distribution. However, the link between skill mismatch and wage inequality has not been researched so far. In the methodology developed in Fields (2003), these two points correspond to the "levels" and the "differences" questions respectively. The first question can be answered through a decomposition of earnings inequality into *relative factor inequality weights*, each of which measure the importance of the factor in explaining earnings inequality at a point in time. The important levels question in this paper is: how large are the factor inequality weights of surplus and deficit qualifications? A related question is whether the exclusion of these variables from the analysis, as is usually done in studies of earnings inequality, matters at all (I show that it does). The second question is similarly addressed with the use of *differential* factor inequality weights. I show that the evolution of these weights in the case of surplus and deficit qualifications is tightly linked to changes in over- and under-education rates and depth, as well as to the returns that accrue to surplus and deficit schooling.

In the next section, I start by discussing the data and my methodology to measure the mismatch variables. Section three introduces the decomposition of overall wage inequality into relative factor inequality weights. This decomposition has the advantage of being generalizable to an important class of inequality indexes.[5] Here and throughout I conduct the analysis for males and females separately. The fourth section addresses the differences question. Because different inequality measures behave

[5]The main conditions are that the index be continuous and symmetric. Detailed conditions can be found in Fields (2003) and Shorrocks (1982).

differently (including cases when some measures increase and others fall) the differential weights differ across inequality indexes. I analyze five different measures: the gini coefficient, the coefficient of variation, and the 90–10, 90–50 and 50–10 wage gaps. The objective of the section is not only to determine which factors have contributed most to changes in each of the inequality measures but also whether the contribution proceeded mostly from changes in regression coefficients or from changes in the factors' variance. In particular, we are interested in trying to tell apart the effect of increases in the prevalence and the depth of mismatch from the effect of changes in the returns to surplus and deficit qualifications.

The paper also looks at residual inequality. Section five first asks whether the introduction of skill mismatch in the specification of the earnings equation changes the observed patterns of residual dispersion. I find that while within-group inequality is slightly diminished, the well-known upward trend of within-group inequality is still present. Lemieux (2006a) finds that much of the increase in residual inequality is due to changes in the composition of the labor force. Within group inequality is higher among more educated and older workers, whose share in the labor force has increased. I then investigate whether the composition effects still remain when the residuals come from equation (3.1) rather than the standard Mincerian version. The concluding section summarizes the findings.

3.2 Measurement issues

In this section, I describe how the qualifications variables are constructed and briefly describe the data sources utilized. I also present a descriptive analysis of the prevalence of over- and under-education.

3.2.1 Skill Requirements Measure

There is consensus regarding the difficulty of measuring skill requirements. Researchers have used three main approaches, all of which have advantages and drawbacks.[6] In the present study skill requirements are measured using the job-analysis or "objective" method. This measure relies on systematic evaluation by professional job analysts who specify the required level of skills for the job titles in an occupational classification. In the United States this information is available in the Dictionary of Occupational Titles (DOT, U.S. Department of Labor, 1977, 1991). One problem with the DOT is that it provides a variety of alternative measures of job-skill requirements. Cognitive, interactive and motor skill indices are linked to consistent employment matrices (267 occupations and 64 industries). The most often used measure of workplace skills is called "General Educational Development" (GED). On a scale of one to six, GED measures mathematical, language and reasoning skills for each job title.[7] The DOT has clear definitions and detailed measurement instructions that all analysts are supposed to follow. Unfortunately, carrying out such detailed analysis is very expensive, so the DOT is published only at very wide time intervals (1977 and 1991 are the last two years for which there is data). Moreover, later editions do not completely renovate the data. Rather, the new research mostly focuses on new categories leaving the description and analysis of occupations already contained in previous editions almost intact.[8] An implication is that longitudinal studies require extra assumptions about the behavior of the measures for the years for which there is no data.

---

[6]Slonimczyk (2008a) has a brief review of the three methods and their comparative advantages and disadvantages. More extensive discussions can be found in Green et al. (1999) and Chevalier (2003).

[7]An analysis of the trends in the GED and other DOT measures of required skills can be found in Wolff (2000).

[8]Spenner (1985) reviews the quality of this type of skill requirement assessment.

Given the estimate for Q^r the other two qualifications variables are defined as follows:

$$Q^s_{i,t} = \mathbf{1}\big(E_{i,t} - Q^r_{i,t} > l\big) \cdot \big(E_{i,t} - Q^r_{i,t}\big)$$

$$Q^d_{i,t} = \mathbf{1}\big(Q^r_{i,t} - E_{i,t} > l\big) \cdot \big(Q^r_{i,t} - E_{i,t}\big)$$

where $\mathbf{1}(\text{“}x\text{”})$ is an indicator function equal to 1 if the statement “x” is true and 0 otherwise, and $E_{i,t}$ is education measured in years of formal schooling. The parameter l is a positive number representing a chosen level of tolerance to mismatch that might or might not depend on individual characteristics.[9] I set $l = 1$.[10]

3.2.2 Data

With the exception of the skill requirements measure, the data come from the NBER extracts of the CPS earnings files for the period 1973–2002. During 1973–78 earnings related questions were asked to the full CPS sample only in May. Starting in 1979, earnings questions have been asked every month to around a fourth of the sample (the outgoing rotation groups (ORG) in CPS jargon). Details on the treatment of the CPS data are discussed in separate appendix. Here I only briefly discuss how the May and ORG earning supplements are processed. As in most other studies of earnings inequality, the sample is restricted to employed wage and salary workers. Only individuals between 16 and 64 years of age with positive potential experience are kept. In trying to cope with the high non-response rates for the earnings module, starting in 1979 the BLS has allocated earnings to non-respondents by means of a hot-deck imputation method. Because earnings were not allocated to non-respondents during 1973-78, observations with imputed earnings have to be ignored to keep the

[9] If l depends on individual characteristics such as occupation or industry then it is more appropriate to speak of a tolerance function. Also note that if $l > 0$, correctly matched individuals will have Q^r in the range $[E_{i,t} - l, E_{i,t} + l]$.

[10] Within a reasonable range the results reported here are robust to different choices for this parameter.

series consistent over the whole period. I also drop observations for 1994 and the first eight months of 1995, a period during which allocation flags are not available.

The earnings variable we use is constructed to represent real hourly earnings including overtime, tips and commissions. A known advantage of the May/ORG CPS earnings data is that it provides a point-in-time measure of earnings. Hourly earnings are weekly earnings including overtime, tips and commissions divided by usual weekly hours, except in the case when a separate (and higher) hourly rate is provided. Earnings are deflated using the CPI-U-X1 series. As in most of the literature on earnings inequality, I multiply the sampling weights by usual weekly hours so as to make the sample of hourly earnings representative of the total hours worked in the economy. I also adjust—"winsorize"—topcoded earnings, multiplying them by 1.4. After the 1994 CPS overhaul respondents with variable hours are allowed to answer that their weekly "hours vary". I use a method developed by Schmitt (2003) to allocate weekly hours to these workers.

The educational attainment variable is also of great importance in this study. In 1992 the education item in the CPS questionnaire was modified. Previously individuals had been asked for the highest completed grade of schooling (in years). The new item asks for the highest degree obtained. In 1998 a new battery of questions was added that permit determining the highest grade completed in most cases. I follow the imputation procedure developed by Jaeger (1997a, 2003) to obtain a consistent measure of the highest grade completed over the whole period.[11]

During the period 1973–82 the CPS used the industrial and occupational classification of the 1970 census. The 1980 census classifications are available during for the rest of the period under analysis. Minor changes were introduced in the classifications

[11]The exception is for individuals with at least some college in the years 1992-7. Details in the appendix.

in 1991, so we adjust the occupation variable in the years prior to the change to retain continuity.

As in Vaisey (2006), I use the database compiled by Autor, Levy, and Murnane (2003) as a source for the required qualifications variable.[12] Thanks to work done by the U.S. Census Bureau personnel, DOT job title codes and some estimates of required qualifications were added to a CPS file. For each occupation, Autor et al. calculated weighted sample means of the GED scores. Independent measures for males and females are available, so the problem generated by the heterogeneity of jobs and requirements within occupations is at least partially taken care of.

I use the 3-digit 1970 and 1980 occupational classification to merge the GED scores to the CPS data for the years 1977 and 1991 respectively. Only the highest of the three GED scores is binding, so I drop the other two. GED values in years other than 1977 and 1991 are obtained through linear interpolation.[13] A final problem involves converting the GED score into the "years of education" unit of measurement. Vaisey (2006) solves the problem using a separate dataset containing both the 3-digit occupational codes and self-reported (subjective) education requirements measured in years of education. The functional form that best maps GED scores into the education requirements variable is a cubic polynomial, which can then be used to convert GED scores for other years. I follow the same approach.

3.2.3 Mismatch rates

Figure (3.1) shows the joint distribution of required qualifications and education at the beginning and the end of the period. To make both years of data comparable, I use a random sub-sample of 2002 workers so that both scatter plots have roughly

[12]Prof. Autor, Levy, and Murnane generously shared these data with Prof. Vaisey, who kindly let me use it too.

[13]All the findings reported in this study remain qualitatively identical if the dataset is restricted to the years 1977 and 1991.

the same number of dots. It is clear that workers with higher qualifications tend to be allocated to jobs with higher requirements. If workers also tended to be correctly matched, the observations would be aligned along the 45 degree lines. However, the slopes from the simple OLS regressions of required qualifications on education are around 0.6.

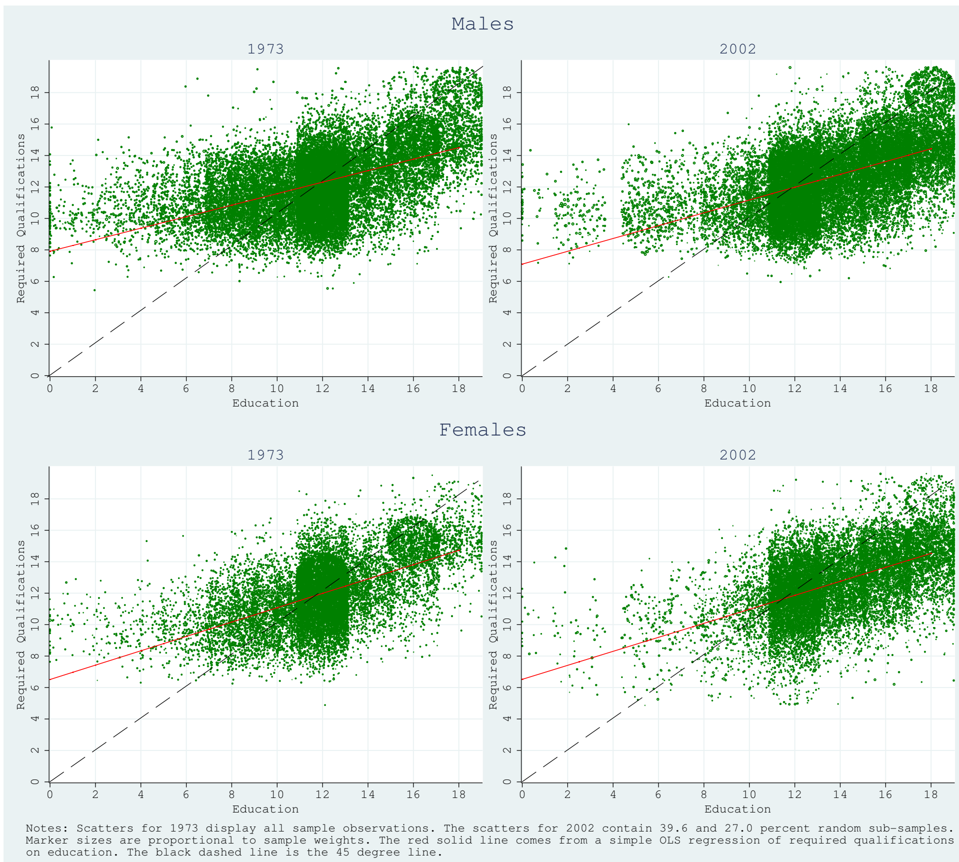


Figure 3.1. Required Qualifications and Education

Both for females and for males it is possible to discern two trends. First, the labor force has become more educated. Second, a much higher proportion of workers have fallen below the 45 degree line, leading to higher over-education rates. The latter point is confirmed by figure (3.2), which shows the evolution of mismatch rates during 1973–2002. Over-education rates for males and females follow a remarkably similar path, starting in 1973 at around 15% and increasing constantly throughout the period to reach levels of around 35% of the employed labor force. Under-education, on the contrary, follows a downward trend.

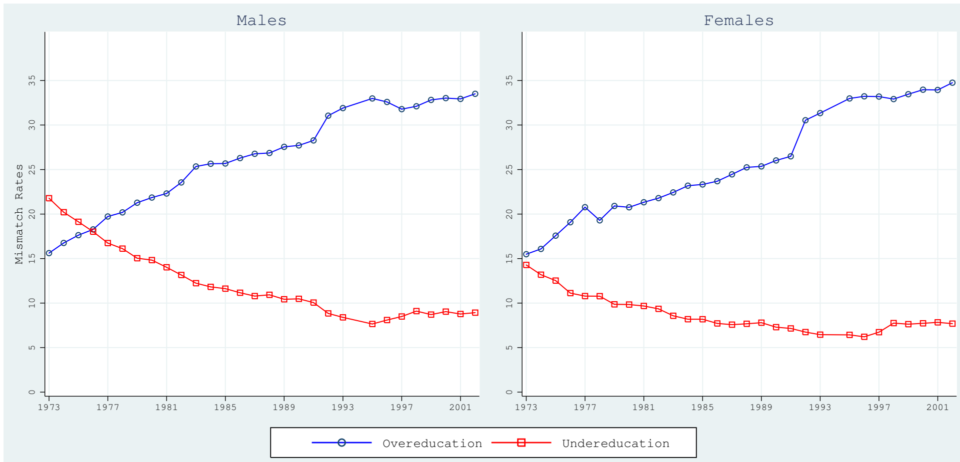


Figure 3.2. Mismatch Rates 1973–2002

Tables (3.1) and (3.2) present descriptive statistics for the most important variables used in the analysis below. Earnings decreased on average for males in the sample and increased for females. As mentioned above, education attainment grew significantly throughout the period. Skill requirements grew as well but at a much slower pace. Interestingly, there are no strong differences between skill requirements for males and for females. The changes in mismatch rates are reflected on average surplus and deficit qualifications, with the former increasing constantly and the latter decreasing in almost every subperiod.

3.3 The Levels Question

Wage inequality in the U.S. has increased significantly in the last three decades. As shown in figure (3.3), measures of overall inequality in log earnings for males like the Gini coefficient, the coefficient of variation, and the 90–10 percentile gap increased

Table 3.1. Descriptive Statistics: Males

	1973	1983	1992	2002
$\ln W$	1.91	1.79	1.77	1.88
	0.49	*0.52*	*0.56*	*0.57*
E	11.90	12.85	13.23	13.50
	3.14	*2.94*	*2.94*	*3.02*
Overeducated	15.6%	25.3%	31.1%	33.5%
Undereducated	21.8%	12.2%	8.8%	8.9%
Q^r	12.27	12.44	12.45	12.60
	2.04	*2.03*	*2.04*	*2.11*
Q^s	0.49	0.82	1.03	1.14
	1.21	*1.51*	*1.66*	*1.73*
Q^d	0.89	0.49	0.37	0.37
	1.91	*1.48*	*1.37*	*1.34*
Age	36.93	36.17	36.77	38.51
	12.73	*12.02*	*11.14*	*11.47*
Married	79.7%	70.5%	64.6%	63.0%
Non-white	14.6%	17.4%	22.6%	28.2%
Part-time	1.6%	3.5%	3.3%	2.3%
Public Sector	15.6%	16.0%	15.0%	13.4%
Manufacturing	34.5%	29.6%	26.1%	20.9%
Services	62.8%	67.5%	71.3%	76.7%
Sample Size	23,078	76,746	72,192	59,765

Notes: Standard deviations for continuous variables are in italics under the sample means.

by around 20% during the period.[14] The Gini, for example, increased from 0.144 in 1973 to 0.174 in 2002. This is a very significant change for earnings inequality, which usually moves slowly. The timing of the change is also interesting. Inequality remained practically constant during the 70s and then had an explosive period of growth during the first half of the 80s. The increase in inequality then slowed down until the early 90s, and finally remained constant or slightly decreased during the remaining years. A quite different story can be told if one looks at inequality in the upper and the lower-tiers of the distribution separately. After the calm 70s, the 90–50 percentile gap increased sharply like the other measures. However, with the exception of a brief decline around 1987 the growth in inequality in the upper tier continued at the same pace into the 90s. The series is practically flat during 1992–97

[14]Growth rates are calculated as log differences. For the percentile gaps, the growth rates correspond to the difference between the rates of growth of the corresponding percentile wages.

Table 3.2. Descriptive Statistics: Females

	1973	1983	1992	2002
$\ln W$	1.51	1.47	1.55	1.69
	0.44	*0.45*	*0.50*	*0.53*
E	12.01	12.88	13.42	13.79
	2.65	*2.50*	*2.59*	*2.73*
Overeducated	15.5%	22.4%	30.5%	34.8%
Undereducated	14.3%	8.6%	6.7%	7.7%
Q^r	11.99	12.32	12.49	12.66
	1.95	*1.89*	*1.93*	*2.06*
Q^s	0.46	0.73	1.03	1.20
	1.13	*1.44*	*1.66*	*1.78*
Q^d	0.51	0.29	0.23	0.26
	1.39	*1.05*	*0.99*	*1.00*
Age	36.73	35.64	36.96	38.98
	13.32	*12.11*	*11.19*	*11.71*
Married	65.2%	59.3%	56.2%	54.6%
Non-white	17.0%	19.2%	22.6%	27.9%
Part-time	2.7%	5.8%	4.4%	2.7%
Public Sector	22.2%	20.5%	20.5%	20.2%
Manufacturing	23.4%	18.4%	15.0%	10.6%
Services	75.9%	80.9%	84.2%	88.6%
Sample Size	15,929	67,979	69,516	59,724

Notes: Standard deviations for continuous variables are in italics under the sample means.

but then continues growing at a fast pace. In contrast, the 90–10 gap started growing earlier and faster but then decreased sharply after 1987. By 2002, inequality in the left half of the earnings distribution was only slightly higher than in 1973. For males, increasing inequality in the right half of the wage distribution explains almost all of the growth in the 90–10 percentile gap. Indeed, the wage distribution for males was slightly left-skewed at the beginning of the period but significantly right-skewed at the end.

Earnings inequality among women behaved quite differently, as can be seen in figure (3.4). After falling during the 70s, the Gini and the coefficient of variation increased during the early 80s but then stagnated and eventually decreased slightly toward the end of the period. The overall increase was only around half that experienced by the same measures for males (the female Gini went from 0.161 to 0.178). In contrast, the 90–10 gap increased much more—by around 30%—and actually sur-

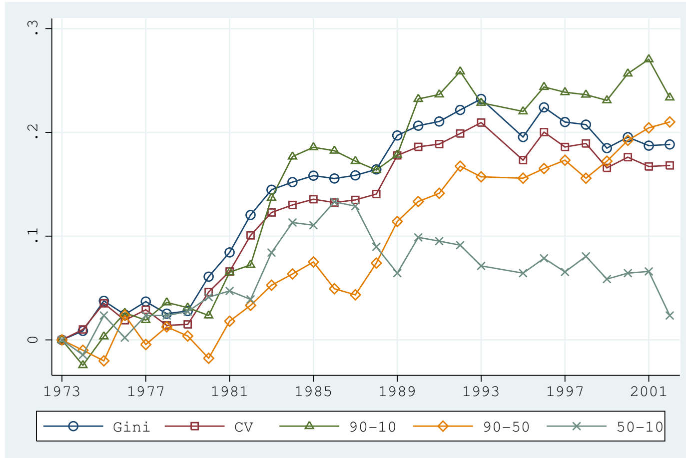


Figure 3.3. The Evolution of Earnings Inequality: Males

(1973=0)

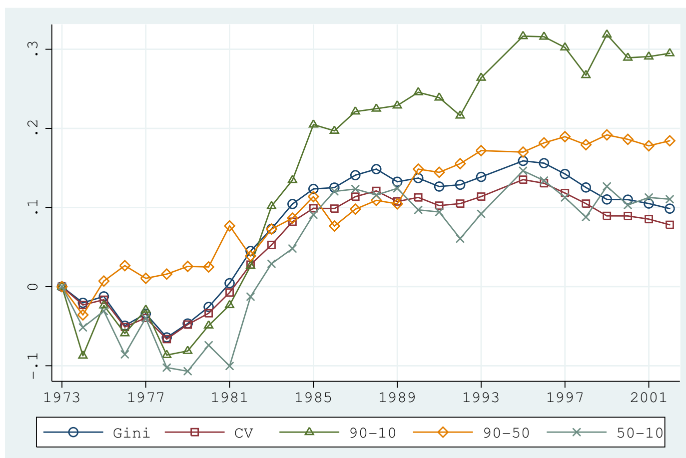


Figure 3.4. The Evolution of Earnings Inequality: Females

(1973=0)

passed the increase experienced by males in the same measure. The difference comes entirely from the lower tail of the female wage distribution, which literally collapsed during the early 80s and never recovered.

3.3.1 Relative Factor Inequality Weights

A necessary step before we can focus on explaining these changes involves looking into the factors that cause inequality at each point in time. Assuming equation

(3.1) is the true income generating function, the variance of log earnings—calculated over individuals at a point in time—can be decomposed into a sum of covariances as follows:

$$\sigma^2(\ln W) = \sigma^2(\sum_{j=1}^{J+2} a_j \cdot z_j) = \sum_{j=1}^{J+2} \text{cov}(z_j \cdot a_j, \ln W) \tag{3.2}$$

where a_j is the j^{th} element of vector a, the OLS estimate of α defined above. There are $J+2$ columns (z_j) in the matrix Z, corresponding to J variables or factors, a column of ones for the constant, and the residuals (the OLS "estimates" of the error term ϵ). Using the definition of the correlation coefficient and a little more algebra yields:

$$\begin{aligned} S_j &= \frac{a_j \cdot \sigma(z_j) \cdot \text{cor}(z_j, \ln W)}{\sigma(\ln W)} \\ \sum_{j=1}^{J+2} S_j &= 100\% \end{aligned} \tag{3.3}$$

where S_j—the *relative factor inequality weight* associated with factor j in matrix Z—represents the fraction of earnings variance that can be attributed to that factor. Each S_j has two building blocks. To see this point more clearly it is useful to look at the case where all factors are orthogonal[15]:

$$S_j = \frac{{a_j}^2 \cdot \sigma^2(z_j)}{\sigma^2(\ln W)} \tag{3.4}$$

On one hand, the factor's potential to explain the variance in earnings depends on the degree of variation in the factor itself. This aspect is represented by the standard deviation of the factor ($\sigma(z_j)$). On the other hand, the effect of the variation

[15] When all regressors are orthogonal $a_j = \frac{\text{cov}(z_j, \ln W)}{\sigma^2(z_j)}$.

in the factor on earnings inequality is limited by the extent to which the two are statistically associated. The latter aspect is represented by the regression coefficient a_j for the factor. The relative factor inequality weights provide information not given by classical regression analysis. A regression provides a measure of statistical association of the dependent variable with each of the regressors but no information on the extent to which the regressor can explain variability in the dependent variable. For example, many studies have shown that the returns to required qualifications (around 9 to 10 percent) estimated using earnings regressions are higher than those of surplus qualifications (2 to 5 percent). The difference between their respective factor inequality weights, however, is much higher as I show below. The reason is that the variability in requirements is greater than in surplus qualifications. Also note that, because by definition the residual is not correlated with any of the regressors, the sum of the first $J+1$ relative factor inequality weights is the R^2 from the regression.

I estimate relative factor inequality weights for three versions of equation (3.1). In the baseline version I include a full set of age dummies[16] but I do not allow for non-linearities in the qualifications variables. The second specification splits the qualifications variables into dummies too.[17] Finally, the third specification allows for non-linearities and also includes a number of extra controls: non-white, married, industry (3 sectors), part-time, and public sector indicators, and 9 region dummies. For comparison purposes, I also estimate the same equations using the standard

[16]The rationale for including dummies rather than a polynomial in age is that the right functional form appears to have changed in time. A quadratic function seems to fit well the beginning half of the series but a quartic in age seems more appropriate for later years (these changes are analyzed in detail in Lemieux, 2006b). If a factor enters into the equation as a string of dummies, then the relative factor inequality weight associated with it is just the sum of the inequality weights calculated for each of the dummies.

[17]The categories are 0–8, 9, 10, 11, 12, 13–15, 16, and 17–18 years of required schooling; 0, 2, 3, 4, 5, and 6 or more years for surplus education; and 0, 2, 3, 4, 5, 6, 7, and 8 or more years for deficit education.

human capital specification (with actual qualifications instead of required, surplus, and deficit qualifications as regressor).[18]

Tables (3.3) and (3.4) present my estimates for the relative inequality weights under the six variations, for males and females respectively. In answering the levels question I focus on the most recent data. The main findings are very similar for males and females. In the standard earnings equation, education appears as the factor with the greatest explanatory power. At least a fifth of earnings inequality for males and a fourth for females can be explained with this single factor. The age factor—a proxy for experience in the labor market—is a far second, explaining 10–14 or 7–8 percent of inequality for males and females respectively. All the control factors taken together explain less earnings inequality than the age factor. These results are roughly in line with those found in Fields (2003).

3.3.2 Restricted and Unrestricted Estimates

Are these estimates trustworthy? In order to answer this question it is useful to think of the estimate for the regression coefficient of the attained education factor as a restricted estimate of the coefficient of the required qualifications factor. As mentioned above, the restriction being imposed is that $\beta^r = \beta^s = -\beta^d$. A standard result in econometrics is that the restricted OLS estimators are unbiased and efficient if the restrictions are true but biased otherwise.[19] Conditional on the estimates of equation (3.1) being consistent[20], the validity of the restriction can be assessed with an F-test. Such a test unequivocally rejects the null hypothesis of true restrictions

---

[18]The categories for the standard education variable are 0–4, 5–8, 9, 10, 11, 12, 13–15, 16, and 17–18.

[19]See, for example, Johnston and DiNardo (1997, ch.3)

[20]The estimates could be inconsistent if, for example, equation (3.1) omitted a relevant variable. Slonimczyk (2008a) estimates equation (3.1) using a panel of matched CPS individuals. The fixed effects estimates, which control for any time-constant observed or unobserved characteristic of the individuals, yield results qualitatively similar to those obtained in cross-section studies. The estimates are attenuated, however, probably due to measurement error. But the hypothesis of equal

Table 3.3. Relative Factor Inequality Weights for Males: 1973–75, 1983–85, 1991–93, and 2000–02

	Linear Specification				Specification with Dummies				Specification with Dummies and Controls			
	1973–75	1983–85	1991–1993	2000–02	1973–75	1983–85	1991–1993	2000–02	1973–75	1983–85	1991–1993	2000–02
Mismatch Equation												
S_{age}	15.27	19.81	16.80	12.41	15.33	19.98	16.91	12.45	12.10	16.60	13.91	10.17
S_{Q^r}	12.00	15.02	20.38	22.62	11.93	15.09	19.74	21.71	10.95	13.75	17.95	19.97
S_{Q^d}	2.50	1.81	2.19	2.33	2.53	1.82	2.18	2.45	1.93	1.47	1.76	1.95
S_{Q^s}	0.31	0.49	1.17	1.93	0.65	0.69	1.58	2.40	0.68	0.69	1.54	2.34
$S_{controls}$									10.90	10.43	10.18	7.99
R^2	30.09	37.13	40.54	39.30	30.44	37.59	40.40	39.01	36.56	42.94	45.33	42.42
Standard Equation												
S^*_{age}	16.70	21.46	18.56	13.63	16.80	21.46	18.12	13.18	12.98	17.41	14.58	10.49
S^*_E	12.01	13.15	18.25	21.68	11.90	13.34	19.15	22.79	10.72	12.16	17.22	20.69
$S^*_{controls}$									11.79	11.54	11.43	9.21
R^2	28.71	34.61	36.81	35.30	28.71	34.80	37.28	35.97	35.49	41.11	43.23	40.39

Notes: Relative inequality weight for the constants are omitted. The factor inequality weight for age is the sum of the weights for 48 age dummies. In the second and third specifications, the inequality weights for required, surplus, and deficit qualifications and for years of education are the sum of the corresponding dummies. The extra controls are: non-white, married, public sector, part-time, industry (3 sectors) and 9 region dummies.

Table 3.4. Relative Factor Inequality Weights for Females: 1973–75, 1983–85, 1991–93, and 2000–02

	Linear Specification				Specification with Dummies				Specification with Dummies and Controls			
	1973–75	1983–85	1991–1993	2000–02	1973–75	1983–85	1991–1993	2000–02	1973–75	1983–85	1991–1993	2000–02
Mismatch Equation												
S_{age}	6.16	9.32	8.67	7.37	6.09	9.31	8.49	7.17	5.60	8.77	8.00	6.75
S_{Q^r}	19.25	21.59	25.66	25.32	17.60	20.17	23.57	24.08	16.20	19.08	21.91	23.39
S_{Q^d}	1.60	0.74	0.85	1.35	1.68	0.73	0.81	1.25	1.38	0.71	0.83	1.20
S_{Q^s}	0.60	1.17	1.70	3.23	0.75	1.33	1.96	3.70	0.57	1.16	1.73	3.49
$S_{controls}$									10.20	6.74	6.98	3.97
R^2	27.60	32.81	36.88	37.27	26.12	31.53	34.83	36.21	33.94	36.46	39.44	38.80
Standard Equation												
S^*_{age}	6.57	1.016	9.66	8.09	6.11	9.76	9.28	7.67	5.69	9.14	8.68	7.28
S^*_E	17.85	17.85	21.50	25.44	19.21	18.63	22.65	26.56	16.97	17.43	20.89	25.59
$S^*_{controls}$									9.19	6.75	7.27	4.16
R^2	24.42	28.00	31.16	33.53	25.32	28.39	31.93	34.23	31.86	33.32	36.83	37.03

Notes: Relative inequality weight for the constants are omitted. The factor inequality weight for age is the sum of the weights for 48 age dummies. In the second and third specifications, the inequality weights for required, surplus, and deficit qualifications and for years of education are the sum of the corresponding dummies. The extra controls are: non-white, married, public sector, part-time, industry (3 sectors) and 9 region dummies.

at the 1% level of significance for all periods both for males and females.[21] If the restrictions are not true *all* of the restricted estimates are biased, including those not apparently affected by the restrictions.[22] Note that the factor inequality weight of education is *not* the restricted factor inequality weight of required qualifications. By definition, the relative factor inequality weight of education (equation (3.3)) is:

$$\begin{aligned} S_E^* &= \frac{1}{\sigma^2(\ln W)} \cdot b_*^r \cdot \operatorname{cov}\big(Q^r + Q^s - Q^d, \ln W\big) \\ &= \frac{1}{\sigma^2(\ln W)} \cdot \operatorname{cov}\big(b_*^r Q^r + b_*^s Q^s + b_*^d Q^d, \ln W\big) \\ &= \frac{1}{\sigma^2(\ln W)} \sum_{i=r,s,d} \operatorname{cov}\big(b_*^i Q^i, \ln W\big) \\ &= \sum_{i=r,s,d} S_{Q^i}^* \end{aligned}$$

where again an asterisk denotes restricted estimates. Thus, while it is correct to compare b_*^E with b^r, one should compare S_E^* with $\sum S_{Q^i}$. Letting Ψ_E denote the true population level relative factor inequality weight of education, the bias in S_E^* when the restrictions are false is given by:

$$\begin{aligned} \mathbf{E}\big(S_E^*\big) - \Psi_E \;=\; & \frac{1}{\sigma^2\big(\ln W\big)} \cdot \Big\{ \Big[\mathbf{E}\big(b_*^r\big) - \beta^r\Big] \operatorname{cov}\big(Q^r, \ln W\big) + \\ & + \Big[\mathbf{E}\big(b_*^s\big) - \beta^s\Big] \operatorname{cov}\big(Q^s, \ln W\big) - \\ & \Big[\mathbf{E}\big(b_*^d\big) - \beta^d\Big] \operatorname{cov}\big(Q^d, \ln W\big) \Big\} \end{aligned} \tag{3.5}$$

---

required, surplus, and (minus) deficit qualifications coefficients could still be rejected at low levels of significance.

[21] The tests were conducted using the linear specification only. Testing each of the two restrictions separately gave the same result.

[22] The equation with the standard education variable does indeed yield restricted estimates for β^s and β^d. Of course, if b_* is the vector of restricted estimates for the qualifications variables: $b_*^E = b_*^s = -b_*^d$.

where $\mathbf{E}$, the expectations operator, is not to be confused with the education variable E. If in reality $\beta^r > \beta^s \approx -\beta^d$ the biases in b^r_* and b^d_* will be negative while the bias in b^s_* will be positive. Because in practice $\mathrm{cov}\big(Q^d, \ln W\big) < 0$, the overall bias in S^*_E will be negative.[23] Looking again at tables (3.3) and (3.4), in the period 2000–02 the differences between S_E and S^*_E in the linear specification were 5.2 and 4.5 percent for males and females respectively. In the other specifications the differences are quite lower. For males, the differences were 3.77 and 3.57 percent in the second and third specifications respectively. The same magnitudes for females were 2.47 and 2.49 percent. Of course, bias attenuation is to be expected in the more flexible specifications. Summing up, the available evidence suggests that the overall effect of education on earnings inequality is larger than what it would appear in the standard (restricted) approach. The magnitude of the additional explanatory power is significant if weighed against the restricted estimates and very large if compared to the explanatory power of the other factors included in the analysis (lumped together as "other controls").

An additional problem with S^*_E is that it does not permit analyzing the relative contributions of required, surplus, and deficit qualifications to explaining earnings inequality. If the restrictions were true and match premia did not exist, the relative contributions of these factors would closely follow their relative variabilities. The relative contribution of required qualifications would typically be around 73 and 79 percent of the total explanatory power of the qualifications factors for males and females respectively.[24] At the other extreme, if surplus and deficit years commanded zero returns all of the explanatory power of education would be due to the quali-

[23]The other two covariances are generally positive. Because the covariances of surplus and deficit qualifications with log earnings tend to be similar in absolute value, the inequality weight of required qualifications is always quite close to S^*_E.

[24]This relative contribution is simply $\frac{S^*_{Q^r}}{S^*_E} = \frac{\mathrm{cov}(Q^r, \ln W)}{\mathrm{cov}(E, \ln W)}$.

fications that are required on the job. Neither restriction appears to be true and reality seems to be quite close to a strict middle ground. The relative contribution of required qualifications averaged around 85 and 90 percent during 1973–2002 for males and females respectively.

3.3.3 Mismatch Premia

The factor inequality weights of surplus and deficit qualifications measure the proportion of the variation in earnings that is explained by skill mismatch. This point can be made clearer if the factor inequality weights are linked to the *mismatch premia*. Define the premia associated with having surplus or deficit qualifications as the average difference between the log wages mismatched workers actually earn and what they would earn if they only had the qualifications that are required on their jobs (which are assumed constant). Simply put:

$$\overrightarrow{\pi^V} = \frac{1}{K} \sum_{Q^s \neq 0} \Big[\big(\beta^r Q^r + \beta^s Q^s \big) - \beta^r Q^r \Big] = \beta^s \cdot \overrightarrow{Q^s} \tag{3.6}$$

$$\overleftarrow{\pi^U} = \frac{1}{H} \sum_{Q^d \neq 0} \Big[\big(\beta^r Q^r + \beta^d Q^d \big) - \beta^r Q^r \Big] = \beta^d \cdot \overleftarrow{Q^d} \tag{3.7}$$

where K and H are the total counts and $\overrightarrow{x}$ and $\overleftarrow{x}$ represent the average value of x for over- and under-educated workers respectively. Note that the mismatch premia depend on the average over- and under-education depth but *not* on over- and under-education rates. The other important component are the coefficients $\beta^{s,d}$. The higher the rewards to surplus qualifications and the penalties to deficit qualification, the higher the premia to mismatched workers.[25] A little more work will prove useful. Mean surplus and deficit qualifications are given by:

---

[25] We discuss only the empirically relevant case: $\beta^R > \beta^S > 0$, $\beta^R > -\beta^D > 0$, $\overrightarrow{Q^s} > 0$, and $\overleftarrow{Q^d} > 0$.

$$\overline{Q^s} = \frac{1}{N}\sum_{i=1}^{N} Q_i^s = \frac{1}{N}\sum_{Q^s>0} Q_i^s = \overrightarrow{Q^s}\frac{K}{N} \tag{3.8}$$

$$\overline{Q^d} = \frac{1}{N}\sum_{i=1}^{N} Q_i^d = \frac{1}{N}\sum_{Q^d>0} Q_i^d = \overleftarrow{Q^d}\frac{H}{N} \tag{3.9}$$

where N is the total number of individuals. Note that K/N and H/N are the over- and under-education rates, which from now on we denote as V and U. Equation (3.8) simply states that average surplus qualifications in the population is equal to a fraction V of mean surplus years among the over-qualified. A similar statement is true about under-qualified workers. The variances can be written as:

$$\sigma^2_{Q^s} = \frac{1}{N}\sum_{i=1}^{N}\left(Q_i^s - \overline{Q^s}\right)^2 = \frac{1}{N}\left[\sum_{Q^s>0}\left(Q_i^s - V\overrightarrow{Q^s}\right)^2 + \sum_{Q^s=0} V^2\left(\overrightarrow{Q^s}\right)^2\right]$$

$$\sigma^2_{Q^d} = \frac{1}{N}\sum_{i=1}^{N}\left(Q_i^d - \overline{Q^d}\right)^2 = \frac{1}{N}\left[\sum_{Q^d>0}\left(Q_i^d - U\overleftarrow{Q^d}\right)^2 + \sum_{Q^d=0} U^2\left(\overleftarrow{Q^d}\right)^2\right]$$

a little algebra yields:

$$\sigma^2_{Q^s} = \frac{1}{N}\sum_{Q^s>0}\left(Q_i^s\right)^2 - V^2\cdot\left(\overrightarrow{Q^s}\right)^2 = V\cdot\overrightarrow{\left(Q^s\right)^2} - V^2\cdot\left(\overrightarrow{Q^s}\right)^2 \tag{3.10}$$

$$\sigma^2_{Q^d} = \frac{1}{N}\sum_{Q^d>0}\left(Q_i^d\right)^2 - U^2\cdot\left(\overleftarrow{Q^d}\right)^2 = U\cdot\overleftarrow{\left(Q^d\right)^2} - U^2\cdot\left(\overleftarrow{Q^d}\right)^2 \tag{3.11}$$

Using again the formula for factor inequality weights in the orthogonal case (equation (3.4)) and substituting with equations (3.6–3.7) and (3.10–3.11), we get:

$$S_{Q^s} = \frac{1}{\sigma^2_{\ln W}}\left[V \cdot \overrightarrow{\left(\pi^V\right)^2} - V^2 \cdot \left(\overrightarrow{\pi^V}\right)^2\right] \tag{3.12}$$

$$S_{Q^d} = \frac{1}{\sigma^2_{\ln W}}\left[U \cdot \overleftarrow{\left(\pi^U\right)^2} - U^2 \cdot \left(\overleftarrow{\pi^U}\right)^2\right] \tag{3.13}$$

Thus, the factor inequality weights of surplus and deficit qualifications are direct simple functions of the mismatched individuals' premia and the mismatch rates. The higher the premia in absolute value, the more earnings inequality gets generated. In other words, equations (3.12–3.13) say that the contribution of mismatch to overall inequality is directly dependent on the extent to which workers' salaries differ from what they would earn were their qualifications adjusted to match requirements.[26]

---

[26]The factor inequality weights of surplus and deficit qualifications measure the contribution of each factor to explaining earnings inequality relative to an hypothetical situation in which, given the existing jobs and requirements, workers had qualifications that exactly matched those requirements. Because surplus and deficit qualifications are to some extent rewarded, those qualifications increase overall inequality. A different interpretation would be necessary if instead of equation (3.1) we had specified:

$$\ln W_{i,t} = X_{i,t} \cdot \gamma_t + \begin{bmatrix} E_{i,t} & Q^s_{i,t} & Q^d_{i,t} \end{bmatrix} \cdot \begin{bmatrix} \delta^E_t \\ \delta^s_t \\ \delta^d_t \end{bmatrix} + \epsilon_{i,t} \tag{3.14}$$

It is not difficult to show that the two specifications are equivalent, only that $\delta^E = \beta^r$, $\delta^s = \beta^s - \beta^r$, and $\delta^d = \beta^r + \beta^d$. The factor inequality weight of Q^s would in this case be negative! The reason is that the existence of over-education reduces inequality relative to a situation in which requirements are upgraded to meet the existing supply of skills, all other things equal. The corresponding *match* premia—the average difference between the counterfactual log wages that mismatched workers would receive if they became correctly matched due to changes in requirements and the log wages they actually get—are:

$$\overrightarrow{\pi'^V} = \frac{1}{K}\sum_{Q^s \neq 0}\left[\delta^E E - \left(\delta^E E + \delta^s Q^s\right)\right] = \left(\beta^r - \beta^s\right)\overrightarrow{Q^s} \tag{3.15}$$

$$\overleftarrow{\pi'^U} = \frac{1}{H}\sum_{Q^d \neq 0}\left[\delta^E E - \left(\delta^E E + \delta^d Q^d\right)\right] = -\left(\beta^r + \beta^d\right)\overleftarrow{Q^d} \tag{3.16}$$

The fact that these match premia and the mismatch premia defined above both have the same relationships with earnings inequality exemplifies well the general principle that when assessing a causal relationship it is fundamental to clearly state the counter-factual state. Equation (3.1) and equation (3.14) both lead to different but correct evaluations of the effect of mismatch on earnings inequality based on opposite counterfactuals. However, there is a problem with equation (3.14) because the resulting S'_E has no useful interpretation. Why would we want to know how much

Table 3.5. Mismatch Depth and Match Premia

	Males		Females	
	1973	2002	1973	2002
$\overrightarrow{Q^s}$	3.17	3.40	2.99	3.47
$\overrightarrow{\pi^V}$	9.1%	17.7%	12.7%	22.1%
$\overleftarrow{Q^d}$	4.09	4.18	3.59	3.33
$\overleftarrow{\pi^U}$	-19.0%	-21.7%	-16.2%	-18.1%

Coming back to empirics, tables (3.3–3.4) show that toward the end of the period surplus qualifications explained around 1.9 and 3.2 percent of earnings inequality for males and females respectively. The same figures for deficit qualifications were 2.3 and 1.4 percent. Part of the difference between the surplus and deficit qualifications weights come from the relative prevalence of the phenomena. We have discussed how by 2002 over-education rates were much higher than under-education rates. It is then surprising that, at least for males, the factor inequality weight of deficit qualifications is higher than the weight of surplus qualifications. The reason can be found in table (3.5). For males, the penalty associated with being under-qualified is on average substantially larger than the premium of being over-qualified. The situation is reversed for women, which explains why $S_{Q^s} > S_{Q^d}$ in this case.

The most interesting fact regarding the factor inequality weights of surplus and deficit qualifications is how the former has increased enormously while the latter decreased. We address this issue in the broader context of changes in the earnings distributions.

3.4 The Differences Question

In the previous section we estimated factor inequality weights to answer the question of which factors are most important in explaining earnings inequality at a point

earnings inequality would be explained by education if, given their current qualifications, all workers were correctly matched?

in time. Here we investigate whether these factors can explain changes in inequality over time.

The inequality weight of required qualifications increased consistently during the period, although growth was concentrated in the 80s. The increase was much more pronounced for males than for for females. Deficit qualifications explains relatively more of the inequality in earnings during the 70s than the 80s. This change is in line with what one would expect considering the declining trend of under-education over time. However, this factor's inequality weight rebounds (in the 80s for males and 90s for females) and ends at levels close to that of the beginning of the period. The explanatory power of surplus qualifications, on the contrary, increased monotonically throughout the period.

Two questions naturally arise from these findings. First, what accounts for these extraordinary changes in the factor inequality weights? For males, the explanatory power of required qualifications almost doubled in 30 years. Depending on the specification, their surplus qualification weight increased from a negligible 0.3 at the beginning of the period to the more substantial 1.9 percent in 2000–02. The increase in the weight for required qualifications was less dramatic for women probably in part because this factor was relatively more important at the beginning of period, while their surplus qualifications weight increased by 5 to 6 times. The second question is whether these factors' magnitudes are significant enough to explain the changes in earnings inequality reviewed above.

3.4.1 Differences in Factor Inequality Weights

It is possible to decompose the changes in factor inequality weights over time.[27] Logarithmically differentiating S_j (defined in equation (3.3)), we get:

[27] The decomposition is only exact for infinitesimal changes.

$$\hat{S}_j = \hat{a}_j + \sigma(\hat{z_j}) + \mathrm{cor}(z_j,\hat{\ln W}) - \sigma(\ln\hat{W}) \tag{3.17}$$

where $\hat{x}$ stands for the percentage rate of growth of x. As noted in Fields (2003), this decomposition has the problem that a_j and $\mathrm{cor}(z_j, \ln W)$ are not independent. One way around this problem is to look at the decomposition that would result if all the factors were orthogonal (see again equation (3.4)). We would then have the decomposition:

$$\hat{S}_j = 2 \cdot \hat{a}_j + 2 \cdot \sigma(\hat{z_j}) - 2 \cdot \sigma(\ln\hat{W}) \tag{3.18}$$

This decomposition will be inexact to the extent that factors are not orthogonal but would still provide a useful benchmark.

Tables (3.6) and (3.7) look at the rates of growth of factor inequality weights and their components for the period 1973–2002.[28] The required qualifications inequality weight grew faster for males than for females, so the gap in the explanatory power of this factor was practically closed by the end of the period despite the fact that initially women had a weight that almost doubled that of men. The growth rate of skill requirements is mostly accounted for by the growth in the degree of association with log earnings rather than by an increase in the standard deviation. Using the first decomposition (equation (3.17)) we see that the growth rates of the regression coefficient and the correlation between education and log earnings for males each account for more than 100 percent of the rate of growth in the inequality weight of skill requirements. The second decomposition (equation (3.18)) confirms this observation.[29]

[28]Here we focus on the specifications where education and the mismatch variables enter as linear terms. In the more flexible specifications some of the dummies have negative weights, which makes it impossible to calculate the rate of growth.

[29]Note that the second decomposition never adds up to 100% because factors are not really orthogonal.

The surplus qualifications weights increased at extraordinary rates of 182 and 173 percent for males and females respectively. In this case the decompositions are more balanced, with the growth in dispersion of surplus qualifications accounting for 20 and 26 percent of the growth in the inequality weights (decomposition 1) for males and females respectively. If we used decomposition 2, then the dispersion elements would appear to account for even a greater fraction of this growth. Finally, the deficit qualifications inequality weight fell during the period despite the fact that the association between deficit qualification and earnings increased. The decompositions show this point by assigning net negative contributions to the regression coefficient and the correlation between deficit qualifications and earnings.

While the decompositions given by equations (3.17) and (3.18) are useful, they beg the question when it comes to the surplus and deficit qualifications factors. What would be desirable is to link changes in S_{Q^s} and S_{Q^d} to changes in mismatch rates and depth. Differentiating equation (3.12) we get:

$$\frac{\partial S_{Q^s}}{\partial \overrightarrow{Q^s}} = \frac{2\left(\beta^s\right)^2 \cdot \overrightarrow{Q^s} \cdot V(1-V)}{\sigma^2_{\ln W}} \tag{3.19}$$

$$\frac{\partial S_{Q^s}}{\partial \beta^s} = \frac{2\left(\overrightarrow{Q^s}\right)^2 \cdot \beta^s \cdot V(1-V)}{\sigma^2_{\ln W}} \tag{3.20}$$

Thus, an increase in the over-education premium that comes either through an increase in surplus qualifications depth or through higher returns to surplus years of education would tend to increase overall inequality. In turn, the effect on earnings inequality of a change in the prevalence of over-education is given by:

$$\frac{\partial S_{Q^s}}{\partial V} = \frac{\overrightarrow{(\pi^V)^2} - 2V \cdot \left(\overrightarrow{\pi^V}\right)^2}{\sigma^2_{\ln W}} \tag{3.21}$$

which is always positive as long as mismatch rates are below 50%. The analysis for under-education is identical but note that because in practice $\beta^d < 0$ an increase

Table 3.6. Changes in Factor Inequality Weights and its Components: Males

	1973	2002	Growth Rate (%)	Decomp. 1 (as % of $\hat{S}$)	Decomp. 2 (as % of $\hat{S}$)
Required Qualif					
S_{Q^r}	11.5%	22.3%	66.1		
$\sigma(\ln W)$	0.49	0.57	15.4	-23.4	-46.7
b^r	0.07	0.12	51.9	78.5	156.9
$\sigma(Q^r)$	2.04	2.11	3.2	4.8	9.7
$\text{cor}(Q^r, \ln W)$	0.38	0.50	26.5	40.1	
Total				100.0	119.9
Surplus Qualif					
S_{Q^s}	0.3%	2.0%	182.1		
$\sigma(\ln W)$	0.49	0.57	15.4	-8.5	-16.9
b^s	0.03	0.05	59.5	32.6	65.3
$\sigma(Q^s)$	1.21	1.73	35.9	19.7	39.4
$\text{cor}(Q^s, \ln W)$	0.04	0.12	102.2	56.1	
Total				100.0	87.8
Deficit Qualif					
S_{Q^d}	2.5%	2.2%	-11.4		
$\sigma(\ln W)$	0.49	0.57	15.4	135.2	270.4
b^d	-0.046	-0.052	11.2	-98.4	-196.9
$\sigma(Q^d)$	1.912	1.341	-35.5	310.9	621.7
$\text{cor}(Q^d, \ln W)$	-0.140	-0.185	28.3	-247.7	
Total				99.9	695.3
Education					
S_E^*	11.6%	21.3%	60.6		
$\sigma(\ln W)$	0.49	0.57	15.4	-25.5	-50.9
b_*^r	0.05	0.08	42.2	69.7	139.3
$\sigma(E)$	3.14	3.02	-3.9	-6.4	-12.9
$\text{cor}(E, \ln W)$	0.34	0.50	37.7	62.2	
Total				100.0	75.5

Note: factor inequality weight are derived using a standard earnings equation and an equation with mismatch variables. The equations include a full set of age dummies but no other controls. The education and mismatch variables enter as linear terms. Growth rates are calculated as log differences between the end and starting periods.

in the returns to under-education leads to a fall in inequality. Summing up, both increases in the absolute value of mismatch premia and in mismatch rates should be expected to lead to increases in overall inequality. Columns 2 and 3 of tables (3.6–3.7) provide the regression coefficients for the qualifications variables in 1973 and 2002 for males and females respectively. Tables (3.1–3.2) contain the mismatch rates. Finally, table (3.5) presents my estimates for the depth of mismatch and the mismatch pre-

Table 3.7. Changes in Factor Inequality Weights and its Components: Females

	1973	2002	Growth Rate (%)	Decomp. 1 (as % of $\hat{S}$)	Decomp. 2 (as % of $\hat{S}$)
Required Qualif					
S_{Q^r}	19.0%	25.7%	30.5		
$\sigma(\ln W)$	0.44	0.53	18.7	-61.5	-123.0
b^r	0.10	0.13	29.0	95.1	190.3
$\sigma(Q^r)$	1.95	2.06	5.6	18.5	36.9
$\text{cor}(Q^r, \ln W)$	0.43	0.50	14.6	47.9	
Total				100.0	104.2
Surplus Qualif					
S_{Q^s}	0.5%	2.9%	173.4		
$\sigma(\ln W)$	0.44	0.53	18.7	-10.8	-21.6
b^s	0.04	0.06	40.9	23.6	47.1
$\sigma(Q^s)$	1.14	1.78	45.3	26.1	52.2
$\text{cor}(Q^s, \ln W)$	0.05	0.14	106.1	61.2	
Total				100.0	77.7
Deficit Qualif					
S_{Q^d}	1.9%	1.3%	-37.9		
$\sigma(\ln W)$	0.44	0.53	18.7	49.5	99.0
b^d	-0.05	-0.05	18.7	-49.3	-98.6
$\sigma(Q^d)$	1.39	1.00	-33.1	87.4	174.9
$\text{cor}(Q^d, \ln W)$	-0.14	-0.13	-4.7	12.4	
Total				100.0	175.2
Education					
S^*_E	18.2%	25.2%	32.3		
$\sigma(\ln W)$	0.44	0.53	18.7	-58.1	-116.1
b^r_*	0.07	0.09	28.8	89.4	178.7
$\sigma(E)$	2.65	2.73	3.2	9.8	19.6
$\text{cor}(E, \ln W)$	0.43	0.52	19.0	58.9	
Total				100.0	82.2

Note: factor inequality weight are derived using a standard earnings equation and an equation with mismatch variables. The equations include a full set of age dummies but no other controls. The education and mismatch variables enter as linear terms. Growth rates are calculated as log differences between the end and starting periods.

mia. The premia almost doubled for overeducated males and females. The growth was mostly due to increases in the returns to surplus education and not so much related to over-education depth. Because the prevalence of over-education also increased, the effect of rising premia on earnings inequality magnified. Under-education depth remained roughly constant for males and decreased for females. However, the significant increase in the penalty associated with deficit schooling led to an increase in the (negative) size of the under-education premia. The latter increase did not propagate, however, because under-education rates fell markedly during the period.

3.4.2 Differences in Overall Inequality Measures

The relative factor inequality weights estimated in the previous section have the good property of providing a unique decomposition of the level of earnings inequality up to a wide range of inequality measures. Unfortunately this property is lost once we move into the territory of changes in the earnings distribution. Different inequality indexes will lead to different answers regarding by how much and in what direction inequality moved. Thus, the decomposition must also be index-specific. The relative factor inequality weights are still useful in creating such decomposition (Fields, 2003). Given an inequality index I, the change in inequality can be written:

$$\boldsymbol{\Delta} I_{t_1,t_2} = \sum_{j=1}^{J+2} \left[S_{j,t_2} \cdot I_{t_2} - S_{j,t_1} \cdot I_{t_1} \right] \tag{3.22}$$

where $\boldsymbol{\Delta}$ is the difference operator. The contribution of factor j to the change in I is given by:

$$\begin{aligned} \Lambda^{I}_{j,t_1,t_2} &= \frac{S_{j,t_2} \cdot I_{t_2} - S_{j,t_1} \cdot I_{t_1}}{\boldsymbol{\Delta} I_{t_1,t_2}} = S_{j,t_2} + \frac{\boldsymbol{\Delta} S_{j,t_1,t_2}}{\boldsymbol{\Delta} I_{t_1,t_2} / I_{t_1}} \\ \sum_{j=1}^{J+2} \Lambda^{I}_{j,t_1,t_2} &= 100\% \end{aligned} \tag{3.23}$$

where the superscript denotes that Λ is specific to inequality index I. I refer to the Λ coefficients as *differential factor inequality weights*. The coefficients Λ depend positively on the magnitude of the change in the relative factor inequality weights and negatively on the rate of growth of the inequality measure. Note, however, that it is only the latter element that makes the differential factor inequality weights differ across inequality measures.

Table 3.8. Differential Factor Inequality Weights for the Gini Index and the Coefficient of Variation

Males: 1973–83, 1983–92, and 1992–2002

	Gini				CV			
	1973–83	1983–92	1992–2002	1973–2002	1973–83	1983–92	1992–2002	1973–2002
Total Variation								
	0.022	0.013	-0.006	0.030	0.034	0.023	-0.010	0.047
Mismatch Equation								
Λ_{age}	46.9%	-17.9%	149.6%	-2.2%	52.1%	-18.3%	160.4%	-4.1%
Λ_{Q^r}	30.4%	96.1%	-44.7%	74.4%	33.5%	96.9%	-50.0%	81.3%
Λ_{Q^d}	-3.5%	8.7%	1.8%	0.9%	-4.5%	8.8%	1.8%	0.8%
Λ_{Q^s}	0.9%	12.0%	-19.8%	9.9%	1.0%	12.1%	-21.5%	11.0%
Λ_{resid}	25.3%	1.1%	13.1%	16.9%	17.8%	0.5%	9.4%	11.1%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Standard Equation								
Λ^*_{age}	50.0%	-14.3%	166.4%	-1.6%	55.5%	-14.7%	178.4%	-3.6%
Λ^*_{E}	16.3%	90.4%	-78.2%	67.9%	17.0%	91.2%	-86.1%	74.1%
Λ^*_{resid}	33.7%	23.9%	11.8%	33.7%	27.4%	23.5%	7.6%	29.5%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Notes: Calculations based on linear specification for qualification variables. The differential factor inequality weights for age are the sum of the weights for the 48 age dummies.

Table 3.9. Differential Factor Inequality Weights for the Gini Index and the Coefficient of Variation

Females: 1973–83, 1983–92, and 1992–2002

	Gini				CV			
	1973–83	1983–92	1992–2002	1973–2002	1973–83	1983–92	1992–2002	1973–2002
Total Variation								
	0.012	0.010	-0.005	0.017	0.016	0.016	-0.009	0.024
Mismatch Equation								
Λ_{age}	52.7%	-0.2%	36.9%	26.4%	70.1%	-0.8%	40.5%	31.6%
Λ_{Q^r}	41.9%	114.3%	23.7%	90.9%	50.5%	120.4%	23.5%	108.8%
Λ_{Q^d}	-13.7%	1.6%	-14.0%	-4.5%	-19.5%	1.7%	-15.9%	-6.1%
Λ_{Q^s}	5.9%	16.8%	-35.3%	25.8%	7.9%	17.8%	-40.0%	32.1%
Λ_{resid}	13.3%	-32.5%	88.8%	-38.6%	-8.9%	-39.2%	92.0%	-66.4%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Standard Equation								
Λ^*_{age}	60.0%	3.6%	42.0%	32.3%	80.0%	3.2%	46.1%	38.9%
Λ^*_E	0.4%	96.8%	-105.2%	92.2%	-6.3%	102.0%	-121.3%	110.6%
Λ^*_{resid}	39.6%	-0.4%	163.2%	-24.5%	26.2%	-5.3%	175.1%	-49.5%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Notes: Calculations based on linear specification for qualification variables. The differential factor inequality weights for age are the sum of the weights for the 48 age dummies.

Tables (3.8–3.9) present, for males and females respectively, estimates for the differential factor inequality weights for the Gini coefficient and the coefficient of variation. Here I present results for the specification linear in the qualifications variables only.[30] Both measures of wage inequality behaved similarly during the period, as we already saw in figures (3.3–3.4). Overall wage inequality increased during the 70s and 80s and decreased slightly during the 90s. However, the rate of growth of the Gini coefficient was slightly higher than that of the coefficient of variation. The increase over the whole period was much stronger among males than females. What factors explain these changes?

The restricted differential weights of education are in the order of 70 percent for males and 100 percent for females.[31] Using what we learnt from the analysis of changes in factor inequality weights we can say that increases in the returns to education account for a major part of the increases in inequality. Again, these results are broadly in line with those found in Fields (2003). However, as discussed in the previous section the restricted factor inequality weights are probably biased.

According to the estimates that result from the mismatch equation, education accounts for substantially more growth in earnings inequality than what the restricted estimates suggest. The differential weights add up to 85 and 93 percent for males and 112 and 135 percent for females. For men, required qualifications accounts for only a small fraction of the difference between the unrestricted and the restricted estimates. For women, the required qualifications differential weight is actually smaller than the restricted education weight. Changes in deficit qualifications and under-education

[30] The results for the age and qualifications factors in the other two specifications are very similar. The other controls have a small and negative Λ for changes over the whole period. Detailed tables are available from the author upon request.

[31] For these measures of inequality, the explained fraction of the variability in log earnings (R^2) increased at a faster pace than inequality. As a consequence, the residual factors have negative differential weights and the differential weights of the non-residual factors add up to more than 100%.

penalties had very small impact on earnings inequality. The differences between the restricted and unrestricted estimates (and also the most remarkable point about tables (3.8–3.9)) are due to the very large differential weights of the surplus qualifications factor. As discussed above, S_{Q^s} grew at very high rates throughout the period, led both by the increases in over-education rates and premia. These changes accounted for 10–11 percent of the increases in the inequality measures for males and 26–32 percent of the increases in the measures for females. For males the growth in these inequality weights was almost completely concentrated in the 80s. The differential weights for women are larger in the 80s but also significant in the 70s. Because inequality according to these measures decreased during the 90s the differential weights for surplus qualifications are negative during this period.

In our investigation of the levels question we found that, although significant, the role of deficit and surplus qualifications in explaining earnings inequality was modest. Also, both factors carried around the same weight for male wage inequality. Surplus qualifications appeared as a relatively more important factor than deficit qualifications for females. The results in this section show that questions regarding changes in inequality lead to very different results. Both for males and for females the contribution of surplus qualifications toward explaining these changes far outweighs the contribution of deficit qualifications. The sheer sizes of the figures for Λ_{Q^s}, specially for women, suggest that the over-education phenomenon is very important in understanding the changes in the wage distribution in the last 3 decades.

3.4.3 Differences in Percentile Gaps

Tables (3.10–3.12) further investigate the effects of changes in factor inequality weights on the distribution of wages. Like the Gini and the CV, the 90–10 percentile gap is a measure of overall inequality. Thus, it is not surprising that the estimates in table (3.10) broadly confirm the findings obtained using the former measures.

However, all the differential factor inequality weights for females are lower for this measure than for the Gini or the CV. The exception is the differential residual factor, that now is positive and large. As reviewed above, the inequality measures for females expanded at lower rates than those of males. The exception is the 90–10 gap, whose rate of growth was much higher than that of the other measures for females. Moreover, the 90–10 gap among females is the only measure of overall inequality that increased during the 90s. While the factor inequality weights of age, and required and surplus qualifications grew throughout the period, the rate of growth of the 90–10 gap was more rapid. Therefore, the fraction of the change in inequality they can explain is lower in this case. Also, despite the fact that the residual factor weight decreased over the period, the rate of growth in the 90–10 gap for females was so high as to make the second term on the right of equation (3.23) very small in absolute value.

Tables (3.11–3.12) permit analyzing what happened to the right and left halves of the wage distributions separately. Both for males and for females the 90–50 percentile gap grew very rapidly, so the estimates for this measure of inequality somewhat resemble those obtained for the female 90–10 gap.[32] The unrestricted qualifications variables explain "only" 60 and 56 percent of the increases in inequality in the right-halves of the male and female wage distributions respectively.

The 50–10 percentile gaps grew slower than the 90–50 gap over the period. However, the slower growth rates are only consequential for males. In fact, for females the differential inequality weights for this measure do not differ much from those of the 90–50 percentile gap. For males, however, the increases in the weights of the qualifications variables over-explain the growth in this measure of inequality.

[32]Indeed the 90–50 gaps grew faster than the 90–10 gaps if rates of growth are calculated as in equation (3.23). These growth rates differ from the difference between the growth rates of the corresponding percentile wages, which are plotted in figures (3.3–3.4).

Table 3.10. Differential Factor Inequality Weights for 90–10 Percentile Gap

Males and Females: 1973–83, 1983–92, and 1992–2002

	Males				Females			
	1973–83	1983–92	1992–2002	1973–2002	1973–83	1983–92	1992–2002	1973–2002
Total Variation								
	0.137	0.122	-0.025	0.234	0.102	0.114	0.079	0.295
Mismatch Equation								
Λ_{age}	58.4%	-14.8%	281.8%	-3.8%	43.4%	3.5%	-6.4%	14.6%
Λ_{Q^r}	37.2%	89.2%	-109.2%	80.1%	37.4%	77.1%	26.7%	50.0%
Λ_{Q^d}	-5.6%	8.1%	1.4%	0.8%	-10.7%	1.3%	8.7%	-0.9%
Λ_{Q^s}	1.1%	11.0%	-40.8%	10.8%	4.8%	10.5%	21.3%	11.4%
Λ_{resid}	8.9%	6.4%	-33.1%	12.1%	25.0%	7.6%	49.8%	24.9%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Standard Equation								
Λ^*_{age}	62.1%	-11.4%	313.6%	-3.2%	49.4%	6.1%	-7.7%	17.3%
Λ^*_E	18.0%	83.9%	-174.0%	73.0%	3.9%	65.1%	87.9%	50.1%
Λ^*_{resid}	19.9%	27.5%	-39.5%	30.3%	46.8%	28.8%	19.8%	32.6%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Notes: Calculations based on linear specification for qualification variables. The differential factor inequality weights for age are the sum of the weights for the 48 age dummies.

Table 3.11. Differential Factor Inequality Weights for 90–50 and 50–10 Percentile Gaps

Males: 1973–83, 1983–92, and 1992–2002

	90–50				50–10			
	1973–83	1983–92	1992–2002	1973–2002	1973–83	1983–92	1992–2002	1973–2002
Total Variation								
	0.053	0.115	0.043	0.210	0.084	0.007	-0.068	0.024
Mismatch Equation								
Λ_{age}	66.4%	1.5%	-66.2%	4.0%	53.4%	-275.8%	62.7%	-73.3%
Λ_{Q^r}	42.0%	53.6%	60.7%	52.1%	34.2%	660.4%	-2.3%	329.6%
Λ_{Q^d}	-7.1%	5.1%	2.5%	1.5%	-4.7%	57.0%	2.1%	-5.5%
Λ_{Q^s}	1.3%	6.0%	14.4%	6.5%	1.0%	91.5%	-6.0%	48.9%
Λ_{resid}	-2.5%	33.8%	88.6%	35.8%	16.0%	-433.1%	43.5%	-199.8%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Standard Equation								
Λ^*_{age}	70.5%	4.1%	-73.9%	4.9%	56.9%	-259.1%	69.7%	-75.9%
Λ^*_{E}	19.1%	49.9%	78.2%	48.0%	17.2%	628.2%	-15.2%	296.3%
Λ^*_{resid}	10.3%	46.0%	95.7%	47.1%	25.9%	-269.2%	45.6%	-120.4%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Notes: Calculations based on linear specification for qualification variables. The differential factor inequality weights for age are the sum of the weights for the 48 age dummies.

Table 3.12. Differential Factor Inequality Weights for 90–50 and 50–10 Percentile Gaps

Females: 1973–83, 1983–92, and 1992–2002

	90–50				50–10			
	1973–83	1983–92	1992–2002	1973–2002	1973–83	1983–92	1992–2002	1973–2002
Total Variation								
	0.073	0.083	0.029	0.184	0.029	0.032	0.050	0.111
Mismatch Equation								
Λ_{age}	34.3%	4.7%	-13.8%	13.5%	66.2%	0.2%	-2.2%	16.4%
Λ_{Q^r}	33.0%	64.3%	27.2%	46.1%	48.5%	110.4%	26.4%	56.3%
Λ_{Q^d}	-7.6%	1.2%	12.5%	-0.5%	-18.2%	1.6%	6.4%	-1.4%
Λ_{Q^s}	3.8%	8.3%	30.8%	10.1%	7.4%	16.1%	15.7%	13.6%
Λ_{resid}	36.5%	21.4%	43.2%	30.8%	-3.9%	-28.4%	53.6%	15.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Standard Equation								
Λ^*_{age}	38.9%	6.9%	-16.1%	15.9%	75.5%	3.9%	-2.8%	19.7%
Λ^*_{E}	7.3%	54.3%	120.6%	46.1%	-4.8%	93.5%	68.9%	56.6%
Λ^*_{resid}	53.7%	38.8%	-4.5%	37.9%	29.3%	2.6%	33.9%	23.7%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Notes: Calculations based on linear specification for qualification variables. The differential factor inequality weights for age are the sum of the weights for the 48 age dummies.

In conclusion, the estimation of differential factor inequality weights to percentile gaps confirm that the qualifications variables and surplus qualifications in particular are important in explaining changes in the wage distribution. Nevertheless, some of these measures seem to grow at either faster or slower rates than the inequality weights. This discrepancies suggest that factors other than the ones included in the present analysis could be important in explaining the evolution of these measures.

3.5 Residual Inequality Analysis

The residual variance from the unrestricted equation (3.1) will necessarily be lower than that of the restricted version. In all studies of within-group inequality, however, residuals are obtained from an equation of the form:

$$\ln W_{i,t} = X_{i,t} \cdot \gamma_t + \delta_t^E E_{i,t} + \mu_{i,t} \tag{3.24}$$

where education is generally entered either as a linear term or as a more or less restricted set of dummies[33]. Comparing equations (3.24) and (3.14) and using what we know about the relationship between the δ and the β parameters, we get the following expression for the variance of the error term:

$$\sigma_\mu^2 = \left(\beta^s - \beta^r\right)^2 \sigma_{Q^s}^2 + \left(\beta^d + \beta^r\right)^2 \sigma_{Q^d}^2 + \sigma_\epsilon^2 \tag{3.25}$$

Looking at the expressions for $\sigma^2_{Q^s,d}$ in equations (3.10–3.11), it is clear that the difference between the residual variance that results from estimating the restricted equation (3.24) and the one that results from equation (3.1) will tend to be greater

[33]In the literature on residual inequality, equation (3.24) also incorporates a full set of interaction terms between education and age/experience. In this study I choose not to use interaction terms because the corresponding factor inequality weights are difficult to interpret. In preliminary explorations of the data I found that a full set of interaction terms would not change any of the main results and added very little to the explanatory power of the regression.

the higher the mismatch rates. The difference is also positively related to absolute value of the *match* premia defined in equations (3.15–3.16).

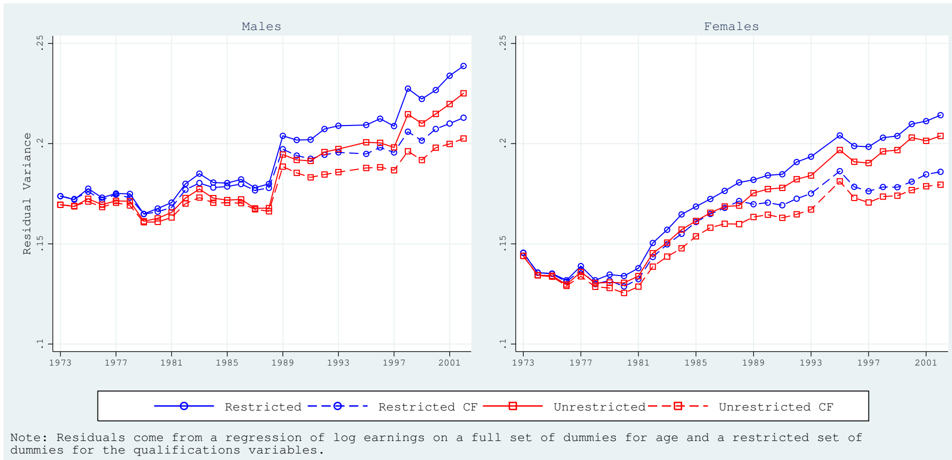


Figure 3.5. Residual Variance 1973–2002

Figure (3.5) plots the residual variance for both specifications (solid lines). The gap between the two series is seen to increase slightly in time. However, the effect is too mild to counteract the clear upward tendency of within-group inequality. We learnt from tables (3.3–3.4) that the residual factor—the fraction of overall earnings that cannot be accounted for by any of the factors—has actually *decreased* over time. This point does not come as much of a consolation since it is still true that most of the inequality in earnings has unknown sources. Decomposing education into required, surplus and deficit qualifications increases the R^2 of the regression but it does not significantly alter the known facts regarding residual inequality. Thus, we can say that the increase in residual inequality is not mainly due to the imposition of unjustified restrictions regarding the pricing of skills in the estimating equation.

3.5.1 Composition Effects

Why has residual inequality increased? Lemieux (2006a) offers the hypothesis that much of the increase in residual inequality is due to composition effects. The American labor force has experienced very significant changes in the course of the

Table 3.13. Residual Variance by Education Groups

Males	1973–75					2000–02				
	Residual Variance		Sample Prop.	V	U	Residual Variance		Sample Prop.	V	U
	Eq. (3.24)	Eq. (3.1)	(%)	(%)	(%)	Eq. (3.24)	Eq. (3.1)	(%)	(%)	(%)
HS Dropouts	0.169	0.161	31.8	1.6	54.4	0.146	0.148	12.6	1.3	42.4
High School	0.154	0.152	33.3	9.3	5.7	0.187	0.179	31.4	14.3	6.5
Some College	0.169	0.164	19.0	26.7	4.8	0.204	0.184	27.4	35.7	2.7
College Degree	0.195	0.188	10.6	46.5	2.8	0.253	0.235	18.8	64.3	1.2
Advanced Degree	0.299	0.287	05.4	57.7	0.0	0.283	0.268	9.8	66.5	0.0
Weighted Averages									Changes Eq. (3.24)	Changes Eq. (3.1)
Actual Prop.	0.173	0.168				0.208	0.196		0.035 [20.1%]	0.027 [16.3%]
Cf. 1973 Prop.	0.173	0.168				0.189	0.181		0.016 [9.2%]	0.013 [7.5%]
Females	1973–75					2000–02				
	Residual Variance		Sample Prop.	V	U	Residual Variance		Sample Prop.	V	U
	Eq. (3.24)	Eq. (3.1)	(%)	(%)	(%)	Eq. (3.24)	Eq. (3.1)	(%)	(%)	(%)
HS Dropouts	0.133	0.130	0.274	2.7	37.7	0.103	0.112	0.093	6.1	31.9
High School	0.134	0.129	0.420	13.8	6.3	0.158	0.151	0.307	14.4	10.9
Some college	0.132	0.124	0.174	23.4	3.8	0.186	0.171	0.315	31.4	3.1
College Degree	0.156	0.154	0.103	33.6	0.8	0.228	0.221	0.195	62.1	0.5
Advanced Degree	0.225	0.219	0.028	71.8	0.0	0.226	0.220	0.091	76.6	0.0
Weighted Averages									Changes Eq. (3.24)	Changes Eq. (3.1)
Actual Prop.	0.138	0.134				0.182	0.173		0.043 [31.3%]	0.039 [29.5%]
Cf. 1973 Prop.	0.138	0.134				0.157	0.153		0.019 [13.4%]	0.019 [14.2%]

Notes: Calculations based on regressions of log earnings on a full set of age dummies and a restricted set of dummies for the qualification variables. The numbers in square brackets represent the percentage change.

last 3 decades. The baby boom generation is coming near to retirement age so on average employed workers are more experienced today than they were in the past. The human capital model predicts that earnings profiles for workers with different levels of education will diverge as they get more experienced, so there are theoretical reasons to expect an older labor force to exhibit higher within-group inequality. The labor force is also more educated today, something apparent in figure (3.1) and tables (3.1–3.2). Because within-group inequality also increases with education, it is possible that the compositional changes may be the cause of the increase in residual variance. Lemieux finds that in 2000–02 a counterfactual residual distribution constructed using the labor force skill composition of 1973 would have around 11 and 20 percent lower variances for males and females respectively. Thus, most of the growth in within-group inequality during the period can be attributed to composition effects. Finally, Lemieux also reports that, both for men and women, the majority of the composition effect (around 75%) is due to the changes in education as opposed to changes in experience.

It is natural to wonder whether these results hold true for the residuals of the unrestricted equation. Looking at table (3.13), we see that with the exception of high school dropouts in 2000–02, the residual variances that result from the unrestricted regression are lower than those of the restricted regression for all groups. Based on equation (3.25), we expect groups with higher mismatch rates to experience more significant reductions. This is clearly the case for males. The simple correlation between the reduction in variance associated with the unrestricted specification and the total rate of mismatch are 0.94 and 0.72, for 1973–75 and 2000–02 respectively. For females the corresponding figures are 0.08 and 0.90. If mismatch rates grew uniformly with education attainment, composition effects would be attenuated for the unrestricted residuals since changing the participation of each group would have a milder effect on the weighted average. However, total mismatch rates do not uniformly grow with education. Under-education is more prevalent among high-school dropouts, while over-education tends to hit the opposite side of the skill distribution. It is the case, nevertheless, that under-education rates fell and over-education rates increased during the period. The resulting asymmetry between the two types of mismatch could result in something closer to a monotonous increase in overall mismatch with education. We could then still expect composition effects to be attenuated for the unrestricted equation.

Table (3.13) also contains the answer to this question.[34] Composition effects are still important when residuals proceed from equation (3.1). While the attenuation exists, it is very small in size relative to the changes in skill composition and the gaps in residual variances among groups. In the restricted specification, overall residual variance grew around 20 and 31 percent for males and females respectively. If the skill composition of the workforce had remained as in 1973–75, however, growth in within-group inequality would have been much lower: only around 9 and 13 percent again for males and females respectively[35]. Residual variances grew slightly less rapidly in the unrestricted specification but the size of the composition effects is practically the same.[36]

This conclusion is confirmed if the counterfactuals are estimated using many more "cells". The residual variance for equation (3.24) estimated on sample year t can be written:

$$\sigma^2_{\widehat{u}_t} = \sum_i \omega_{i,t} \cdot \widehat{u}^2_{i,t} \tag{3.26}$$

[34]Table (3.13) is similar to Lemieux's (2006a) tables 1A–B but I focus on education alone and do not disaggregate residual variances by experience groups.

[35]This statement assumes that had composition remained the same prices would have still changed in the way they did.

[36]This conclusion does not change if we used 2000-02 as the based period for counterfactual calculations.

where $\widehat{u}$ are the residuals and ω is the sample weight. The counterfactual variance that would result if characteristics were held constant at the levels of the base year (in our case 1973) can be obtained by a re-weighing procedure:

$$\overset{*}{\sigma}{}^{2}_{\widehat{u}_t} = \sum_i \overset{*}{\omega}_{i,t} \cdot \widehat{u}^2_{i,t} \tag{3.27}$$

$$\overset{*}{\omega}_{i,t} = \frac{1 - P_{i,t}}{P_{i,t}} \omega_{i,t} \tag{3.28}$$

where $\overset{*}{x}_t$ is the counterfactual value of x in year t when characteristics are held constant at their 1973 level. The counterfactual weight is simply the original weight multiplied by an adjustment factor. The adjustment is based on the estimated probability ($P_{i,t}$) that individual i is observed in year t and not in the base year. These probabilities of course depend on the individuals' characteristics. Here we are interested in the aforementioned changes in the composition of the labor force, so the relevant characteristics are education and age. To estimate the probabilities we use a logit model on sample containing only year t and the base year (1973). The outcome variable is a dummy signalling whether the individual is contained in the sample for year t. As in Lemieux (2006a), the right-hand-side contains a full set of age dummies and a restricted set of education dummies (same as above), as well as interaction terms between the education dummies and a quartic in age. The counterfactual variance re-weights the residuals so that the sample in year t represents the characteristic present in the base year. For example, because education attainment increased in time a highly educated individual in year 2002 will have $\overset{*}{\sigma}{}^{2}_{\widehat{u}_{2002}} < \sigma^2_{\widehat{u}_{2002}}$.

Figure (3.5) also shows the residual variances that result from applying the estimated counterfactual weights. It is clear that the reduction in residual variance associated with keeping characteristics at the 1973 level is much larger than the reduction that is obtained by using equation (3.1) instead of equation (3.24). The conclusion is that the importance of composition effects in explaining the growth of residual inequality is robust to the removal of the restrictions that do not allow the returns to required, surplus and deficit qualifications to differ.

3.6 Conclusions

The "common wisdom" about wage inequality in the U.S. is that it has grown in time led by increases in the relative demand for high skill workers, probably due to changes in technology that favor those workers vis-a-vis the less intensively trained. This paper questions some aspects of the standard story. First, the available evidence does not seem to support an overall increase in skill requirements. The DOT data presented here and in other studies suggests that requirements have grown very slowly during the period that elapsed between the last two editions (1977–1991). Most accounts of the skill-biased technical change hypothesis situate the beginning of the process in the mid 70s, so the DOT data on skill requirements seems to be in contradiction with this story. Second, while more educated workers do relatively better in the labor market, a substantial fraction of them end up in jobs whose requirements are below their acquired levels of skill. Over-qualification rates seem to have increased substantially while under-education seems to be less common. Changes in the depth of skill mismatch, while significant, have been less impressive.

Surplus qualifications are rewarded in the marketplace to some extent. Thus, over-educated workers would be worse off if placed on jobs whose requirements matched the skills they possess. However, they would be better off if this type of mismatch were eliminated through increases in the skill requirements of their jobs. The converse is true about under-educated workers. As a consequence, the contribution of the education factor toward explaining earnings inequality is more complex than what would appear at first glance.

This paper shows that skill mismatch is a relevant cause of inequality in real earnings in the U.S. and that a substantial fraction of the increase in overall and residual inequality during the period 1973–2002 was due to the increase in mismatch rates and mismatch premia. Surplus and deficit qualifications taken together account for 4.3 and 4.6 percent of the variance in earnings, around 15 percent of the total explained variance in 2002, for males and females respectively. While these figures might seem modest, the analysis of changes in the wage distribution shows that these factors are very important. Specifically, around 11 and 32 percent of the increase in the coefficients of variation of log earnings during the 30 years under analysis can be attributed to the growth in the explanatory power of surplus qualifications, again for males and females respectively.

CHAPTER 4

EMPLOYMENT AND DISTRIBUTION EFFECTS OF CHANGES IN THE MINIMUM WAGE

4.1 Introduction

This paper analyzes the effects of changes in the minimum wage on wage inequality, relative employment and the prevalence of mismatch (over-education) in the labor market.

Influential studies by DiNardo et al. (1995) and Lee (1999) suggest that changes in the minimum wage and other labor market institutions affect workers of different skill levels and that these changes may be more important for the observed increase in inequality than standard supply and demand arguments. This claim, however, faces important objections: a reduction in the minimum wage may increase wage inequality, but in a standard setting it should raise the demand for low-skill workers. Contrary to this prediction, low skill workers appear to have lost ground in terms of both wages and employment. The college premium has increased markedly since the early 1980s, but so has the relative employment of high skill-workers. Figures (4.1) and (4.2) show time series of the college premium and the relative supply of college workers, and the the federal minimum wage respectively.

The simultaneous increase in the relative wage and employment of high-skill workers has been interpreted as evidence of skill-biased technical change (Levy and Murnane, 1992; Acemoglu, 2002, e.g.). The presence of mismatch, however, implies that relative wages and employment can move in the same direction, even in the absence

[1]This chapter was co-authored with Peter Skott.

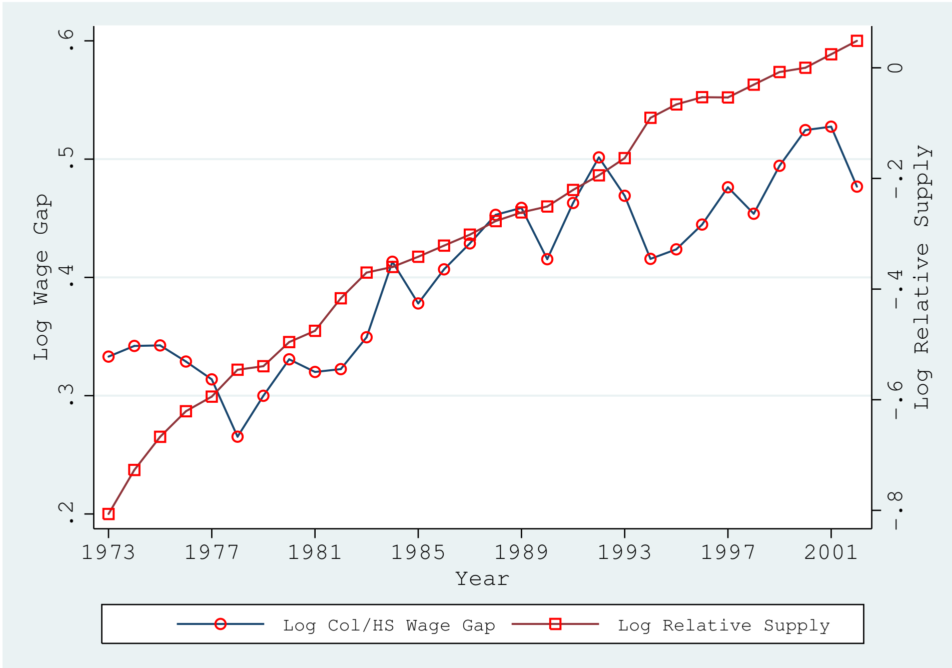


Figure 4.1. College-HS Wage Gap and Relative Supply

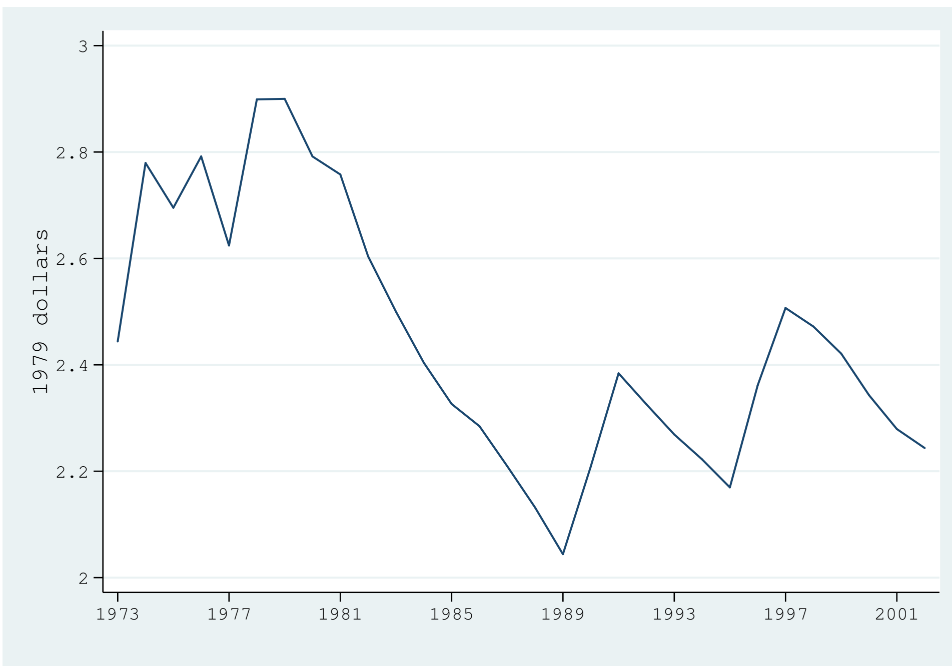


Figure 4.2. Log of Federal Minimum Wage (1979 dollars)

of any skill bias (Sattinger, 2006; Skott and Auerbach, 2004; Skott, 2005, 2006), and induced changes in the prevalence of mismatch may also contribute to an explanation of within-group or residual inequality, which has grown even more than inequality between groups (Katz and Autor, 1999). We use the theoretical framework in Skott (2006) and show that a fall in the minimum wage can generate a deterioration in the position of low-skill workers, both in terms of wages and employment.

The paper has links to another strand of literature. In a perfectly competitive labor market, a binding minimum wage increases both the average and the marginal cost of labor, forcing profit-maximizing firms to reduce employment. Contrary to this prediction, recent empirical studies point to instances where an increase in the minimum wage resulted in increased employment of low wage workers, a result that could be explained by monopsonistic effects (Card and Krueger, 1995; Dube et al., 2007). The monopsony model, literally interpreted to apply to single buyer markets, may have little relevance (for example see Stigler, 1946) but as argued by Manning (2003, 2004), labor markets can be monopsonistic, even if there is a multiplicity of buyers of labor. Indeed, the survey by Boal and Ransom (1997) describes several alternative multi-agent models that lead to many of the same conclusions as classic single-buyer monopsony. We contribute to this literature by showing that efficiency wages can generate economy-wide monopsony effects as well as skill mismatch: both the employment of low-skill workers and total employment may increase in response to a rise of the minimum wage.

To keep matters as simple as possible, we assume that high-skill workers can get two types of jobs (good and bad), whereas low-skill workers have only one type of employment opportunity (bad).[2] Monitoring of workers' effort is imperfect, contracts are incomplete, and workers cannot convincingly pre-commit to not shirking. One

[2] We will refer to good and bad jobs as high-tech and low-tech, respectively.

solution is for firms to use the threat of dismissal as a way to elicit effort (Shapiro and Stiglitz, 1984; Bowles, 1985). For this threat to work, both good and bad jobs must be rationed to ensure that employed workers receive a rent over and above their best alternative. Good jobs pay more than bad jobs, which in turn must pay more than unemployment. In equilibrium there will be both un- and under-employment (some high-skill workers have bad jobs that do not utilize their skills), and inequality between groups will depend not only on the wage gap between good and bad jobs, but also on the degree of mismatch.[3] As long as some matches of high-skill workers and bad jobs are sustained in equilibrium, changes in exogenous variables will affect not only wages and employment rates but also the degree of mismatch. These induced changes in the degree of underemployment of high-skill workers lie behind the monopsonistic effects. An increase in the minimum wage may reduce the employment of high-skill workers in low-tech jobs, and this deterioration of the employment conditions for high-skill workers relaxes the no-shirking condition in high-tech jobs and stimulates employment.

Monopsonistic effects have been introduced into efficiency wage models by Rebitzer and Taylor (1995) but our mechanism is very different. Rebitzer and Taylor assume that firms have fixed monitoring resources, so that the probability of detecting a shirking worker is decreasing in the total number of employees. Thus, firms are forced to increase wages, and with them the potential penalty of dismissal, *pari-passu* with employment. In other words, firms face an upward sloping labor (effort) supply curve, and a binding minimum wage may induce an increase in employment, just as in the classical monopsony case. Unlike Rebitzer and Taylor, we have two different types of workers, and this heterogeneity, in combination with the presence of mis-

[3]Mismatch may persist, also in the long run. If the wage gap between job types is wide enough, workers will invest in human capital even if there is a non-negligible probability of ending up under- or un-employed.

match, implies that monopsonistic features can arise even with exogenously given probabilities of detection.[4] Unemployment, mismatch and monopsonistic effects are generated by the same efficiency-wage mechanism.[5]

The significance of the theoretical analysis depends on the degree of mismatch. While measuring the latter variable has proven challenging, studies suggest that over-education is widespread in all OECD countries. Estimates range between 10 and 40%, and the evidence also shows large differences in the returns to education to different workers, depending on whether they are over- or under-qualified for their jobs (Sicherman, 1991; Groot and Maassen van den Brink, 2000).[6] Our own estimates in this paper produce over-education rates of about 15–25% in the US, and the degree of over-education changes substantially between 1973 and 2002 (the period for which we have data). Moreover, we find some support for monopsonistic effects of changes in the minimum wage: the minimum wage has a positive (but statistically insignificant) effect on the ratio of high- to low-tech jobs and a negative (but again statistically insignificant) effect on both unemployment and under-employment (the degree of mismatch).

Our analysis has implications for the estimation of the elasticity of substitution between different types of labor inputs. The presence of over-education and of substantial changes in the extent of mismatch implies that existing empirical studies are potentially misleading. This paper provides the first estimates of the elasticity of substitution between high- and low-tech *jobs*—as opposed to between high- and

---

[4]The model can be extended to include fixed monitoring resources, as in Rebitzer and Taylor. An appendix with this extension is available on request.

[5]This is unlike the analysis in Manning (2003, pp. 256–262), where efficiency wage elements and involuntary unemployment are added to models with monopsonistic features.

[6]Some studies have suggested that individual ability bias explains these results. Slonimczyk (2008b), however, shows that differences in the returns to surplus and required qualifications persist when fixed effects are introduced.

low-skill workers. Our estimates suggest that the degree of substitutability between inputs may be lower than indicated by Autor et al. (2008).

The paper is in five sections. Section 2 describes the basic efficiency wage model with endogenously generated mismatch. The effects of changes in a binding minimum wage are examined in Section 3. Section 4 presents the empirical evidence, and Section 5 concludes.

4.2 An efficiency wage model with endogenous mismatch

There are two types of job and two types of workers. Jobs are either high-tech or low-tech. Workers can be high-skill or low-skill, and the level of skill is the product of past decisions to invest in human capital, which are taken as given. Only high-skill workers can occupy high-tech positions, but both worker types compete for the low-tech positions.

Firms maximize profits subject to a production function that has only two inputs,

$$Y = F(N_H, N_L) \tag{4.1}$$

where N_H and N_L are the total number of high- and low-tech jobs that have been filled (with non-shirking workers). This specification assumes that high- and low-skill workers are perfect substitutes in low-tech jobs and, to avoid an extra parameter, that they are equally productive. There are constant returns to scale.

The first order conditions with respect to the employment levels yield:

$$\begin{aligned} w_H &= F_1(N_H, N_L) && (4.2) \\ w_L &= F_2(N_H, N_L) && (4.3) \end{aligned}$$

where it is important to note that the marginal products (F_i) correspond to *jobs*. If N_{ij} denotes the employment of worker type i in jobs of type j ($i = H, L$; $j = H, L$) then $N_H = N_{HH}$ and $N_L = N_{HL} + N_{LL}$).

Following Shapiro and Stiglitz (1984), an employed worker of type i in a job of type j gets a wage w_{ij} and instantaneous utility

$$u = \begin{cases} w_{ij} - e_{ij} & \text{if not shirking} \\ w_{ij} & \text{if shirking} \end{cases}$$

where e_{ij} is the worker's disutility associated with exerting effort. Workers are risk neutral and discount future outcomes at the rate ρ.

Firms set wages to ensure that workers' best response is to exert effort. Monitoring is costly and shirkers are detected (and fired) according to a positive but finite hazard rate (δ). The rate of job termination for non-shirking workers (p) is also positive and finite. Discount and termination rates are assumed constant across worker types.

These assumptions define three no-shirking conditions:

$$\begin{aligned} \rho V_{HH} &= w_{HH} - e_{HH} - p(V_{HH} - V_{HU}) \qquad (4.4) \\ &= w_{HH} - (p+\delta)(V_{HH} - V_{HU}) \\ \rho V_{HL} &= w_{HL} - e_{HL} - p(V_{HL} - V_{HU}) + q_{HLH}(V_{HH} - V_{HL}) \qquad (4.5) \\ &= w_{HL} - (p+\delta)(V_{HL} - V_{HU}) + q_{HLH}(V_{HH} - V_{HL}) \\ \rho V_{LL} &= w_{LL} - e_{LL} - p(V_{LL} - V_{LU}) \qquad (4.6) \\ &= w_{LL} - (p+\delta)(V_{LL} - V_{LU}) \end{aligned}$$

where the V_{ij} are the value functions associated with each of the three employment states and q_{ijk} are transition rates for workers of type i in jobs of type j, and transitioning into job type k. Equations (4.4) through (4.6) incorporate the assumptions that low-skill workers get only low-tech jobs and high-skill workers prefer high-tech jobs (the transition rates q_{HHL} and q_{LLH} are zero). If the no-shirking conditions are binding, equations (4.4)–(4.6) imply that

$$V_{HH} - V_{HU} = \frac{e_{HH}}{\delta} \tag{4.7}$$

$$V_{HL} - V_{HU} = \frac{e_{HL}}{\delta} \tag{4.8}$$

$$V_{LL} - V_{LU} = \frac{e_{LL}}{\delta} \tag{4.9}$$

There are no unemployment benefits or home production, and the flow of instantaneous utility is zero when unemployed. Thus, the value functions for unemployed workers are given by:

$$\rho V_{HU} = q_{HUH}(V_{HH} - V_{HU}) + q_{HUL}(V_{HL} - V_{HU}) \tag{4.10}$$

$$\rho V_{LU} = q_{LUL}(V_{LL} - V_{LU}) \tag{4.11}$$

Using equations (4.4)–(4.11) and assuming that the transition probabilities for a high-skill worker into high-tech jobs are the same independently of whether the worker is unemployed or under-employed ($q_{HUH} = q_{HLH} = q_{HH}$), we can solve for wages:

$$\begin{aligned} w_{HH} &= e_{HL}\frac{\delta + \rho + p + q_{HH} + q_{HUL}}{\delta} + \\ &\quad +(e_{HH} - e_{HL})\frac{\delta + \rho + p + q_{HH}}{\delta} \end{aligned} \tag{4.12}$$

$$w_{HL} = e_{HL}\frac{\delta + \rho + p + q_{HH} + q_{HUL}}{\delta} \tag{4.13}$$

$$w_{LL} = e_{LL}\frac{\delta + \rho + p + q_{LUL}}{\delta} \tag{4.14}$$

Given the termination rates for shirkers and non-shirkers and a constant supply of both types of workers (H, L), all transition probabilities (q) can be determined through steady state conditions that depend only on employment levels. In a steady state, the unemployment rates and the rate of mismatch are constant, and entries and exits from each of the employment states are balanced. Formally:

$$
\begin{aligned}
q_{HH}(H - N_H) &= pN_H && (4.15)\\
q_{HUL}(H - N_H - N_{HL}) &= pN_{HL} + q_{HH}N_{HL} && (4.16)\\
q_{LUL}(L - N_{LL}) &= pN_{LL} && (4.17)
\end{aligned}
$$

Using (4.15)–(4.17), the wage equations (the no-shirking conditions) can be written

$$
\begin{aligned}
w_{HH} &= e_{HL}\frac{\delta + \rho + p\frac{H}{H-N_H-N_{HL}}}{\delta} + (e_{HH} - e_{HL})\frac{\delta + \rho + p\frac{H}{H-N_H}}{\delta} && (4.18)\\
w_{HL} &= e_{HL}\frac{\delta + \rho + p\frac{H}{H-N_H-N_{HL}}}{\delta} && (4.19)\\
w_{LL} &= e_{LL}\frac{\delta + \rho + p\frac{L}{L-N_{LL}}}{\delta} && (4.20)
\end{aligned}
$$

The no-shirking conditions (4.18)–(4.20) define three distinct wage rates. However, at an interior solution with both high- and low-skill workers in low-tech jobs, we must have $w_{HL} = w_{LL} = w_L$ since otherwise profit maximizing firms would never hire both types of workers. Trivially, $w_H = w_{HH}$ since only high-skill workers have high-tech jobs.

Equations (4.18)–(4.20) can be combined with the first order conditions (4.2)–(4.3) to solve for equilibrium values of employment (N_H, N_{HL}, N_{LL}) and wages (w_H, w_L) in the absence of a binding minimum wage. Using (4.18)–(4.20) it is readily seen that the two groups of workers will have the same unemployment rates ($u_H = \frac{H-N_H-N_{HL}}{H} = \frac{L-N_{LL}}{L} = u_L$) if $e_{HL} = e_{LL}$. Empirically, unemployment rates for low-skill workers are higher than for high-skill workers, and we assume $e_{LL} > e_{HL}$. The same equations show that the two unemployment rates must move together. From the wage equations it follows, finally, that high-tech jobs pay a higher wage than low-tech jobs if $e_{HH} > e_{HL}$;[7] we assume this condition is met.

As shown by Skott (2006), this model can generate seemingly paradoxical effects. Neutral shifts in the production function may affect the relative wage and the relative employment rate of high-skill workers in the same direction and, moreover, since it hurts the employment prospects of low-skill workers, an increase in the supply of high-skill labor can lead to an increase in the skill premium.

4.3 Minimum wages

Now suppose that a minimum wage $\underline{w}$ is established and that this minimum wage is binding for low-tech but not for high-tech jobs. We are interested in the effects of an increase in $\underline{w}$ on employment and wages.

With constant returns to scale and perfect competition, an equilibrium must be characterized by zero profits. To satisfy this condition, an increase in one of the wage rates must be associated with a decline in the other wage.[8] By assumption the minimum wage is binding for low-tech jobs, and an increase in the minimum wage must therefore reduce the wage in high-tech jobs. Using the first-order conditions (4.2)–(4.3), the resulting decline in the wage ratio w_H/w_L generates an increase in the

[7] A similar result could be obtained with equal levels of effort disutility but different detection rates of shirkers ($\delta_{HL} > \delta_{HH}$).

[8] Assume that both wages at the new equilibrium were greater than or equal to wages at the original equilibrium (with at least one strict inequality). In this case firms would have been able to make positive profits at the original configuration of wage rates and the initial position could not have been an equilibrium.

employment ratio N_H/N_L. This general result is independent of the wage equations. Additional results, however, require assumptions about mismatch.

4.3.1 A standard model without mismatch

Without mismatch, the no-shirking condition for high-skill workers reads

$$w_H = e_{HH} \frac{\delta + \rho + p\frac{H}{H-N_H}}{\delta} \tag{4.21}$$

and the no-shirking condition for low-skill workers is replaced by the binding minimum wage

$$w_L = \underline{w} \tag{4.22}$$

Using (4.21), a decline in w_H implies a fall in N_H and since the employment ratio N_H/N_L rises, low-skill employment must also fall. These results do not depend on the efficiency-wage formulation. The same conclusions apply whenever the relevant "supply" curve for high-skill labor is upward sloping and independent of the minimum wage (a completely inelastic curve implies that high-skill employment is unaffected by an increase in the minimum wage while low-skill employment falls).

4.3.2 Minimum wages and induced mismatch

If the minimum wage is binding then, by definition, the no-shirking condition cannot be binding for both high- and low-skill workers in low-tech jobs. It may be binding for one or the other, but the minimum wage only has bite if the number of low-tech jobs could be increased without shirking, even with an unchanged wage. We consider two polar cases. In the first case, the no-shirking condition is always binding for low-skill workers; in the second case it is always binding for high-skill workers.

In his study of wage setting behavior, Bewley (1999) found that overqualified job applicants were common but that many employers were reluctant to hire them. Indeed, this "shunning of overqualified job applicants" is highlighted as one of two novel findings of the study (p.18). Attitudes to overqualified applicants differed somewhat between primary and secondary sector jobs, where secondary sector jobs are defined as short-term positions that are often part time. Both sectors received applications from overqualified workers, but for primary sector jobs 70 percent of firms expressed a "total unwillingness" to hire them, 10 percent were "partially unwilling" and only 19 percent were "ready to hire" overqualified applicants (pp. 282–83). Two main reasons account for for the negative attitude to overqualifications: a concern that applicants would quit again as soon as possible and a concern that applicants would be unhappy on the job. Secondary sector employers had fewer reservations, but only a minority (47 percent) "were ready to hire them" with 30 percent being "totally unwilling" and 23 percent "partially unwilling" (p. 324).

Bewley's findings support our first case: they suggest that firms may prefer low-skill workers in low-tech jobs if both high- and low-skill workers are available at the same wage cost. Büchel (2002), however, suggests that "over-educated workers are generally more productive than others" and that, because of this, "firms hire over-educated workers in large numbers." This claim would seem to support our second case.

4.3.2.1 Case 1: Mismatch with low-skill workers preferred in low-tech jobs

When firms prefer low-skill workers in low-tech jobs, high-skill workers will only be hired for low-tech jobs if the no-shirking condition is binding for low-skill workers. Thus, the no-shirking condition for low-skill workers is satisfied as an equality while the minimum wage exceeds the expression for w_{HL} in (**??**). Since the no-shirking condition for high-skill workers in low-tech jobs fails to be satisfied as an equality, equation (4.8) no longer holds. Instead—using (4.4), (4.5), (4.10) and $w_L = \underline{w}$—we have

$$V_{HL} - V_{HU} = \frac{\underline{w} - e_{HL}}{\rho + p + q_{HH} + q_{HUL}} = \frac{\underline{w} - e_{HL}}{\rho + p\frac{H}{H - N_H - N_{HL}}} \tag{4.23}$$

and the no-shirking conditions for high-skill workers in high-tech jobs and low-skill workers can be written,

$$\begin{aligned} w_H &= \frac{\delta(\underline{w} - e_{HL})}{\rho + p\frac{H}{H - N_H - N_{HL}}} \frac{\delta + \rho + p\frac{H}{H - N_H - N_{HL}}}{\delta} + \\ &\quad + (e_{HH} - \frac{\delta(\underline{w} - e_{HL})}{\rho + p\frac{H}{H - N_H - N_{HL}}}) \frac{\delta + \rho + p\frac{H}{H - N_H}}{\delta} &(4.24) \\ \underline{w} &= w_L = e_{LL} \frac{\delta + \rho + p\frac{L}{L - N_{LL}}}{\delta} &(4.25) \end{aligned}$$

Equation (4.25) implies an important result. It shows that N_{LL} will increase following a rise in the minimum wage, that is, low-skill workers will benefit both in terms of wages and employment.

The solution for N_H and N_{HL} is not quite as simple. The high-tech wage and the ratio of high-tech to low-tech jobs are determined, as before, by the first order conditions (4.2)–(4.3), and the values of N_H and N_{HL} can be derived using (4.24) and the definitional relation

$$N_H = \frac{N_H}{N_L}(N_{HL} + N_{LL}) \tag{4.26}$$

The effect of a rise in $\underline{w}$ on N_H is ambiguous. There may be a negative effect on the number of high-skill jobs, not surprisingly, but a positive effect on N_H can be obtained if N_{LL} is elastic and an increase in w_L generates a large decrease in N_{HL}. This possibility is illustrated numerically in Table 4.1.

Table 4.1. Employment and wage effects of changes in the minimum wage when firms prefer low-skill workers in low-tech jobs

$(L = H = 1, e_{LL} = 1.3, e_{HL} = 0.5, e_{HH} = 2, Y = 5N_H^{0.5}N_L^{0.5}, \rho = 0.1, \delta = 1, p = 0.2)$

$\underline{w}$	N_{LL}	N_{HL}	N_L	N_H	w_H	Ω	N	$\frac{w_{HA}}{\underline{w}}$	Θ
1.7	0.03	0.58	0.62	0.29	3.68	0.64	0.9	1.38	0.40
1.8	0.30	0.35	0.67	0.35	3.47	0.37	1.02	1.45	0.32
1.9	0.45	0.23	0.67	0.39	3.29	0.21	1.06	1.46	0.24
2.0	0.54	0.11	0.66	0.42	3.13	0.10	1.08	1.44	0.16
2.1	0.61	0.02	0.63	0.45	2.98	0.02	1.08	1.40	0.06

An increase in N_H is a necessary condition for other interesting effects. The employment ratio N_H/N_L must rise, but with an increase in N_H this condition can be satisfied, even with an increase in N_L. An increase in both N_L and N_H, moreover, implies that aggregate employment must also increase. These monopsonistic effects are made possible because a rise in minimum wages relaxes the no-shirking constraint for low-skill workers, and as the employment of high-skill workers in low-tech jobs decreases, there is a derived effect on the no-shirking condition for high-skill workers in high-tech jobs.

Table 4.1 also shows the effects on the degree of over-education (Ω), the average wage premium to high-skill workers ($\frac{w_{HA}}{w_L}$) and within group inequality (Θ).[9] The increase in $\underline{w}$ reduces over-education and within-group inequality. The average wage premium first increases but then falls again if the minimum wage is raised beyond a certain point.

[9] These variables are defined as follows:

$$\begin{aligned}
\Omega &= \frac{N_{HL}}{N_H + N_L} \\
\frac{w_{HA}}{w_L} &= \frac{\frac{N_{HL}}{N_H+N_{HL}} w_L + \frac{N_H}{N_H+N_{HL}} w_H}{w_L} \\
\Theta &= \sqrt{\frac{N_{HL}}{N_H + N_{HL}}\left(\frac{w_L - w_{HA}}{w_{HA}}\right)^2 + \frac{N_H}{N_H + N_{HL}}\left(\frac{w_H - w_{HA}}{w_{HA}}\right)^2}
\end{aligned}$$

4.3.2.2 Case 2: Mismatch when firms prefer high-skill workers in low-tech jobs

In this case firms will not hire low-skill workers unless the no-shirking condition is binding for high-skill workers in low-tech jobs. Empirically, some low-skill workers are employed. We therefore assume that the condition is binding and that wages must satisfy the following equations:

$$\begin{aligned} w_H &= e_{HL}\frac{\delta+\rho+p\frac{H}{H-N_H-N_{HL}}}{\delta}+(e_{HH}-e_{HL})\frac{\delta+\rho+p\frac{H}{H-N_H}}{\delta} && (4.27)\\ \underline{w} &= w_{HL}=e_{HL}\frac{\delta+\rho+p\frac{H}{H-N_H-N_{HL}}}{\delta} && (4.28)\end{aligned}$$

From profit maximization we know that an increase in $\underline{w}$ leads to a decline in w_H and an increase in N_H/N_L. Equations (4.27)–(4.28) now imply that N_H must fall (substitute (4.28) into (4.27) and use the fact that $w_H-\underline{w}$ decreases) and hence that N_L declines.

These implications are qualitatively the same as in the case without mismatch. The presence of mismatch, however, adds a few extra results. Using (4.28), it follows that a rise of $\underline{w}$ will increase aggregate employment of high-skill workers (N_H+N_{HL}). Hence, the decline in low-skill employment ($N_{LL}=N_L-N_{HL}$) is exacerbated, the proportion of mismatched high-skill workers ($N_{HL}/(N_H+N_{HL})$ and the degree of over-education (Ω) go up, and the wage premium, $w_{HA}/\underline{w}$ will fall. Total employment ($N=N_H+N_L$) must decrease since N_H/N_L increases and N_H falls.

According to this model, the fall in minimum wages since the 1970s should have led to increases in high-tech wages and the wage premium; the number of high-tech jobs should also have increased but over-education should have dropped, as should total

Table 4.2. Employment and wage effects of changes in the minimum wage when firms prefer high-skill workers in low-tech jobs

$(L = H = 1, e_{LL} = 0.2, e_{HL} = 0.5, e_{HH} = 2, Y = 5N_H^{0.5}N_L^{0.5}, \rho = 0.1, \delta = 1, p = 0.2)$

$\underline{w}$	N_{LL}	N_{HL}	N_L	N_H	w_H	Ω	N	$\frac{w_{HA}}{\underline{w}}$	Θ
1.61	0.86	0.39	1.25	0.52	3.88	0.22	1.77	1.81	0.39
1.64	0.50	0.48	0.99	0.42	3.81	0.34	1.41	1.62	0.41
1.67	0.03	0.62	0.65	0.29	3.74	0.66	0.94	1.40	0.41

employment of high-skill workers and within-group inequality; low skill workers should have seen an increase in employment. Numerical results are given in Table 4.2.[10]

4.4 Evidence

4.4.1 Measuring mismatch and match premia

The empirical relevance of the analysis in the previous section depends critically on the extent of mismatch in the labor market. There is agreement in the literature regarding the difficulty of measuring skill requirements. The best existing source for the U.S. is the Dictionary of Occupational Titles (DOT). The DOT reports expert assessment of more than 12,000 job titles. We take the General Education Development (GED) index as our measure of skill requirements. The GED ranks jobs in a scale of 1 to 6 (a GED of 4 roughly represents the skills acquired through high-school). Jobs with GED greater than 4 are considered high-tech. Unfortunately the very detailed job classification of the DOT is not available in any representative survey of earnings. We use the average GED over 3-digit occupations as a proxy measure. The analysis is thus restricted to the period 1973–2002, during which the 1970 and 1980 census

[10]With one exception, the benchmark parameters are the same as in Table 4.1. The exception is the cost of effort for low-skill workers which has been changed to $e_{LL} = 0.2$ (compared to $e_{LL} = 1.3$ in Table 4.1). The value of e_{LL} does not affect the solution for low-skill employment, but a lower value of e_{LL} is chosen to ensure that the no-shirking constraint is satisfied for low-skill workers at the implied levels of N_{LL} and $w_L = \underline{w}$.

occupational classifications were in use. During this period there were two data issues of the DOT: 1977 and 1991. Other years are obtained through linear extrapolation.

The skill requirements data were merged with the Current Population Survey (CPS) earnings files. We use the education item to identify low- (high school or less) and high-skill workers (at least some college). Figure (4.3) shows the distribution of the labor force across job and skill levels over the period. The graph confirms the well studied movement toward higher levels of education attainment. The share of employed workers with at least some college studies went from around 33% in 1973 to over 58% in 2002. Less well known is the steady increase in the share of high skill workers whose jobs have requirements below their skill level, at least according to the Bureau of Labor Statistics experts. At the beginning of the period only 14.7% of workers were in this category; toward the end of the period the percentage of over-educated workers had increased by 10 percentage points.

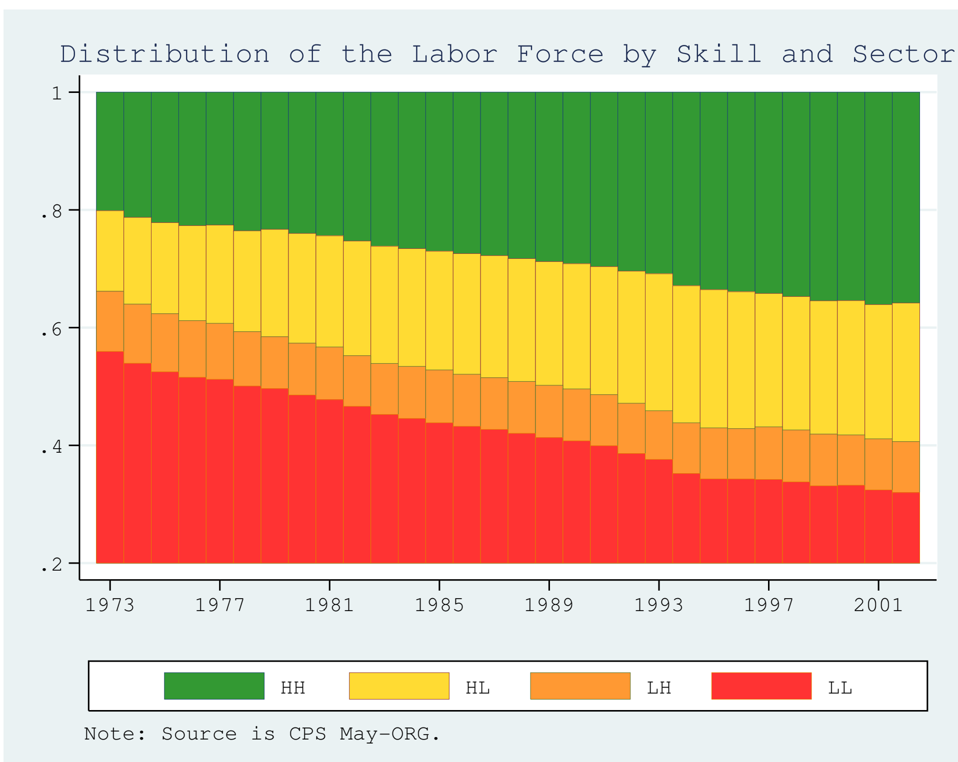


Figure 4.3. Employment Trends

Do job types matter for earnings conditional on education attainment? To answer this question we construct a wage sample from the CPS files. In 1973–78 earnings questions were asked to the whole CPS sample in May. Starting in 1979 earnings questions are asked every month to roughly a fourth of the sample (the outgoing rotation groups). Our earnings variable is real weekly earnings divided by usual weekly hours, unless a separate and higher hourly rate is also reported. Earnings are deflated using the CPI (1979 = 100). The wage sample contains all employed wage and salary workers between 18 and 65 years of age. We weight the CPS data by hours worked and the appropriate sampling weight. The CPS has undergone several changes that reduce its consistency over time; details on the necessary adjustments on earnings and other variables are provided in the appendix.

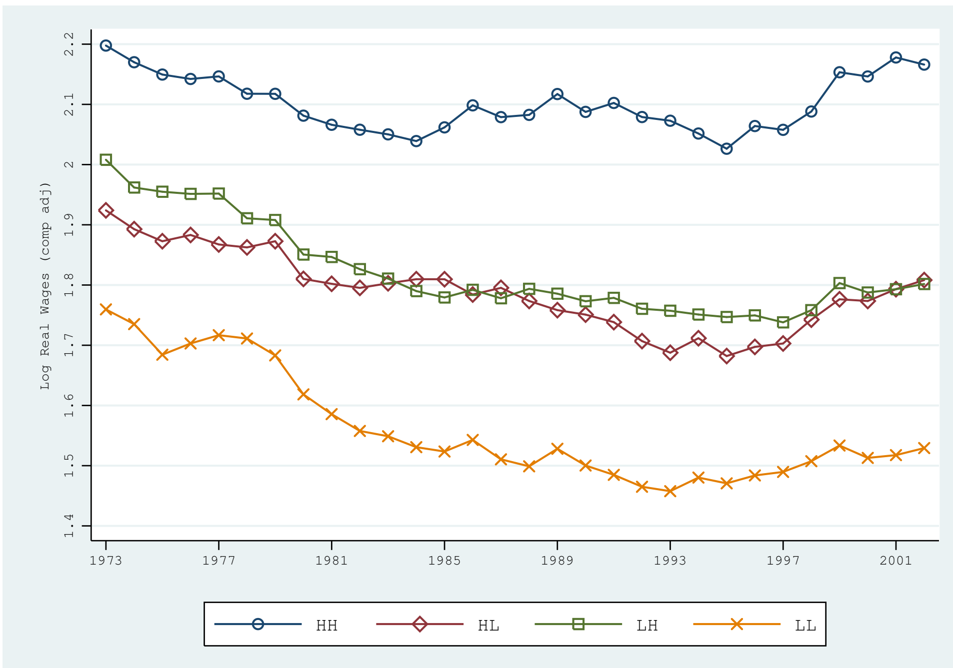


Figure 4.4. Earnings Series

Figure 4.4 shows average log wages for workers separated into the same four groups. The series have not been adjusted for compositional changes within each of the groups, so one should not rush into conclusions (but see below). Wages of high skill workers in high-tech jobs clearly stand out as higher than those of all other groups. The figure also shows low skill workers in high-tech jobs do better on average than over-educated workers.

4.4.2 Effects of the minimum wage on w_H/w_L and N_H/N_L

Wages and employment are determined by the interaction between wage setting and firms' labor demand. The latter is derived from the production function, and we assume the economy has a CES production function with two factors,

$$Y_t = [\alpha_t (a_t N_{H,t})^\rho + (1-\alpha_t)(b_t N_{L,t})^\rho]^{1/\rho}$$

where again $N_{H,t}$ and $N_{L,t}$ refer to *jobs* and not to worker types. The parameters a_t and b_t represent high-tech and low-tech labor augmenting technical change. The constant economy-wide elasticity of substitution is $\sigma = \frac{1}{1-\rho}$. Given the firm's FOC, it follows that

$$\frac{w_{H,t}}{w_{L,t}} = \frac{\alpha_t}{1-\alpha_t}\left(\frac{a_t}{b_t}\right)^\rho \left(\frac{N_{H,t}}{N_{L,t}}\right)^{\rho-1}$$

or

$$\log\frac{w_{H,t}}{w_{L,t}} = \log\frac{\alpha_t}{1-\alpha_t} + \rho\log\frac{a_t}{b_t} + (\rho-1)\log\frac{N_{H,t}}{N_{L,t}} \tag{4.29}$$

which can be rewritten as

$$\log\frac{w_{H,t}}{w_{L,t}} = \frac{1}{\sigma}\left[D_t - \log\frac{N_{H,t}}{N_{L,t}}\right] \tag{4.30}$$

where D_t measures technological shifts favoring high-tech jobs in log quantity units. Substituting a time trend for the unobserved variable D, equation (4.30) can be written

$$\log \frac{w_{H,t}}{w_{L,t}} = \frac{1}{\sigma} \left[a + b\ t - \log \frac{N_{H,t}}{N_{L,t}} \right] \tag{4.31}$$

This derivation of (4.31) is similar to that in Katz and Murphy (1992), except for the modifications arising from our distinction between job characteristics and worker types.[11]

Turning now to wage setting, a log linear version of the no-shirking conditions in the section 3 implies that

$$\begin{aligned} \log \frac{N_H}{N_L} &= \log \frac{N_H}{N_{LL} + N_{HL}} \\ &= \beta_0 + \beta_1 \log \frac{w_H}{w_L} + \beta_2 \log \underline{w} + \beta_3 \log \frac{H}{L} \end{aligned} \tag{4.32}$$

The form of this equation (but not the parameter values) is independent of whether firms prefer high- or low-skill workers in low-tech jobs.

From (4.31)–(4.32) it follows that

$$\log \frac{w_{H,t}}{w_{L,t}} = \frac{a - \beta_0}{\sigma + \beta_1} + \frac{b}{\sigma + \beta_1} t - \frac{\beta_2}{\sigma + \beta_1} \log \underline{w} - \frac{\beta_3}{\sigma + \beta_1} \log \frac{H}{L} \tag{4.33}$$

$$\log \frac{N_H}{N_L} = \frac{\sigma\beta_0 + \beta_1 a}{\sigma + \beta_1} + \frac{b\sigma\beta_1}{\sigma + \beta_1} t + \frac{\beta_2 \sigma}{\sigma + \beta_1} \log \underline{w} + \frac{\beta_3 \sigma}{\sigma + \beta_1} \log \frac{H}{L} \tag{4.34}$$

The adjustment speeds of both employment and wages in response to shocks may differ between high- and low-tech jobs. To correct for this, we include unemployment as a control for cyclical conditions (this again is similar to Autor et al. (2008)).

The results are in Tables 4.3–4.4. An increase in the minimum wage leads to a reduction in the wage premium and a rise in the ratio of high- to low-tech jobs.

[11] Also see Katz and Autor (1999); Autor et al. (2008).

The coefficient is highly significant at 1% in the wage equation but only statistically significant at a 10% level in the job equation. The effects of changes in relative supplies (the H/L ratio) are also as expected, while the positive trend in both regressions is consistent with skill-biased technical change (and/or power-biased technical change).[12]

Table 4.3. Reduced Form Regression for the Hi/Low-tech Log Wage Gap

COEFFICIENT	(1)	(2)
Time	0.011***	0.015***
	[0.002]	[0.002]
$\log \frac{H}{L}$	-0.267***	-0.400***
	[0.087]	[0.085]
$\log \underline{w}$	-0.401***	-0.377***
	[0.092]	[0.079]
u_L		0.525***
		[0.162]
Constant	0.222**	0.040
	[0.093]	[0.097]
Observations	30	30
R^2	0.914	0.939

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

4.4.3 Effects of minimum wages on unemployment and mismatch

The qualitative results in section 4.2 are consistent with all the specifications in section 3, independently of whether there is mismatch and of the precise mismatch assumptions in the case with a binding minimum wage. The specifications differ,

[12] The case for skill-biased technological change may be relatively weak (Howell, 1999; Card and DiNardo, 2002). Skott and Guy (2007) and Guy and Skott (2008) suggest that there is stronger evidence for "power-biased" technological change and that, like skill bias, a power bias can increase both wage and employment inequality.

Power-biased technical change produces shifts in the no-shirking conditions. Allowing for a time trend in these conditions, that is, letting

$$\beta_0 = b_0 + b_1 t$$

would not affect the reduced-form equations. The positive trend in the two equations, however, now reflect both skill-biased and power-biased technical change.

Table 4.4. Reduced Form Regression for the Log Job Composition Ratio

COEFFICIENT	(1)	(2)
Time	0.025***	0.019***
	[0.005]	[0.005]
$\log \frac{H}{L}$	-0.171	0.027
	[0.164]	[0.172]
u_L		-0.783**
		[0.328]
$\log \underline{w}$	0.246	0.211
	[0.173]	[0.160]
Constant	-0.806***	-0.534**
	[0.174]	[0.197]
Observations	30	30
R^2	0.974	0.978

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

however, in their predictions with respect to the effects of the minimum wage on unemployment and underemployment.

The no-shirking conditions and the definitional relation (26) yield reduced-form equations of the form[13]

$$\begin{aligned} u_L &= f(t, \underline{w}, w_H, \frac{N_H}{N_L}, \frac{H}{L}) \\ u_H &= g(t, \underline{w}, w_H, \frac{N_H}{N_L}, \frac{H}{L}) \\ \Omega &= h(t, \underline{w}, w_H, \frac{N_H}{N_L}, \frac{H}{L}) \end{aligned}$$

[13]The no-shirking conditions produce three independent equations. Hence, the three equations for u_L, u_H and Ω are not independent of the equation for N_H/N_L. We have

$$\log \frac{N_H}{N_L} = \log(1 - u_H - \Omega \frac{H+L}{H}) - \log((1-u_L)\frac{L}{H} + \Omega \frac{H+L}{H})$$

where the expression for u_L simplifies to $u_L = f(t, \underline{w})$ in case 1 (firms prefer low-skill workers in low-tech jobs) and the expression for u_H to $u_H = g(t, \underline{w})$ in case 2 (firms prefer high-skill workers in low-tech jobs).

Combining these equations with the first-order conditions (4.2)–(4.3), we get the following log-linearized reduced-form equations

$$\begin{aligned} u_L &= \gamma_0 + \gamma_1 t + \gamma_2 \log \underline{w} + \gamma_3 \log \frac{H}{L} \\ u_H &= \delta_0 + \delta_1 t + \delta_2 \log \underline{w} + \delta_3 \log \frac{H}{L} \\ \Omega &= \rho_0 + \rho_1 t + \rho_2 \log \underline{w} + \rho_3 \log \frac{H}{L} \end{aligned}$$

In case 1, we expect $\gamma_2 < 0$, $\gamma_3 = 0$, $\delta_2 \gtreqless 0$, $\delta_3 \gtreqless 0$, $\rho_2 < 0$, $\rho_3 > 0$;[14] in case 2, on the other hand, we would have $\gamma_2 > 0$, $\gamma_3 < 0$, $\delta_2 < 0$, $\delta_3 = 0$, $\rho_2 > 0$, $\rho_3 = 0$.[15]

Table 4.5 reports the estimates of these reduced form regressions. The regressions for unemployment include a lagged dependent variable while the regression for over-education uses unemployment as a control for cyclical fluctuations. All three equations show a negative effect of the minimum wage, as predicted by case 1. The effect is statistically insignificant, but certainly there is no evidence that the distributional costs of the decline in minimum wages (tables 3-4) have been compensated by increased employment. The evidence may be weak but it suggests the opposite: an increase in the minimum wage may raise employment and reduce inequality. The strongly significant effect of H/L on the degree of over-education also is consistent with case 1 (and not with case 2).

[14] The ambiguity of the sign of δ_2 in case 1 was discussed in section 3. The sign of δ_3 is ambiguous for related reasons. An increase in H/L reduces N_H/H but raises N_{HL}/H, and the unemployment rate can go either way. The analytics are messy, but simulations confirm the result.

[15] These parameter signs follow from equations (27)–(28).

Table 4.5. Reduced Form Regression for Unemployment/Underemployment

COEFFICIENT	(1) u_L	(2) u_H	(3) Ω
time	0.002	0.002	-0.002*
	[0.003]	[0.001]	[0.001]
lminwage	-0.016	-0.031	-0.010
	[0.084]	[0.042]	[0.028]
HL	-0.080	-0.082	0.180***
	[0.114]	[0.051]	[0.030]
L.ulowskill	0.708***		
	[0.178]		
L.uhighskill		0.580***	
		[0.175]	
ulowskill			0.214***
			[0.058]
Constant	-0.045	-0.046	0.333***
	[0.123]	[0.053]	[0.035]
Observations	29	29	30
R^2	0.502	0.512	0.978

Standard errors in brackets***. p<0.01, ** p<0.05, * p<0.1

4.4.4 The elasticity of substitution

Our analysis has implications for the elasticity of substitution. This elasticity has been estimated using only an equation derived from firms' first order conditions (that is, without any attention to wage setting) and without any attention to mismatch (for example Katz and Murphy (1992)). Using our notation, a single regression is run with $\log \frac{w_{HA}}{w_L}$ as the dependent variable and $\log \frac{N_H+N_{HL}}{N_{LL}}$ as the measure of relative employment.[16] We have replicated this procedure with our data set and time period. The results (which are available on request) are similar to those found in the literature. The only notable difference between our results and those in Autor et al. is that the effect of the minimum wage is strong and highly significant in our regression but weakly significant in theirs.

[16] More precisely, the dependent variable is the composition-adjusted log wage gap between college and high-school educated workers and the relative employment measure uses labor quantities in efficiency units. See appendix B for details.

From our perspective, there are two problems with these regressions. When there is mismatch, the theoretically correct specification regresses $\log \frac{w_{H,t}}{w_{L,t}}$ on $\log \frac{N_{H,t}}{N_{L,t}}$, rather than $\log \frac{w_{HA}}{w_L}$ on $\log \frac{N_H+N_{HL}}{N_{LL}}$. Secondly, by disregarding wage setting, the regressions implicitly assume that relative employment can be taken as exogenous. This exogeneity assumption may be reasonable if the labor market is competitive and the supplies of high- and low-skill labor are highly inelastic. It becomes highly questionable, however, if wage formation is governed by efficiency wages and the degree of mismatch is endogenously determined. Thus, the estimates of the elasticity of substitution in Autor et al. (2008) and other studies that follow the same approach may be biased.

Both of these problems can be addressed through the reduced form estimates in tables (4.3)–(4.4). The elasticity of substitution can be recovered from these reduced-form regressions: the implied value of σ can be found as the (negative of the) ratio of the coefficients on $\log \underline{w}$ (or on $\log H/L$). Both of these ratios are very low ($\frac{0.211}{0.377} = 0.56$ and $\frac{0.027}{0.40} = 0.068$, respectively). Note however that these magnitudes are calculated with substantial error. Using the delta method we can obtain 90% confidence intervals, which are $(-0.05, 1.17)$ and $(-0.57, 0.70)$ respectively. Trivially, neither estimate of the elasticity of substitution is statistically different from zero.[17] This low elasticity of substitution between labor inputs is consistent with the findings in (Card et al., 1999).

4.5 Conclusion

The theoretical model in this paper is highly stylized and clearly tells—at best—a small part of the story behind increasing inequality and the links between inequality

[17] We also conducted a Wald test of equality of both estimates of the elasticity of substitution. The null of equality cannot be rejected at conventional levels of significance (the test statistic is 0.87 and is asymptotically distributed as a Chi-squared with one degree of freedom).

and the minimum wage. Several results, however, stand out and may play a role in a more elaborate account of the observed changes.

We have shown that if firms prefer to fill low-tech jobs with low-skill workers rather than with over-educated high-skill workers then

- "aggregate monopsonistic elements" arise naturally in a model with mismatch
- these monopsonistic elements imply that a fall in the minimum wage can have adverse effects on aggregate employment as well as on the degree of mismatch and thus the degree of *under*employment of high-skill workers.
- both within and between group inequality may rise when the minimum wage falls, and
- low-skill workers suffer a double blow of falling employment as well as falling wages.

The evidence reported in section 4 suggest that these theoretical results may be empirically relevant. There is strong evidence of mismatch in the labor market, and the degree of mismatch has been increasing, especially in the 1970s and 1980s. Moreover, the monopsonistic implications of the theoretical model are supported by US data for 1973–2002. Our regressions suggest that the fall in the minimum wage led to a deterioration of the relative wage and employment of low-skill workers and an increase in the underemployment of high-skill workers.

APPENDIX A

CPS DATA

A.1 Overall description of the CPS and the MORG files

The Current Population Survey (CPS) is the major U.S. government household survey of employment and labor force participation. The CPS is the source of numerous high-profile economic statistics including the unemployment rate. It is conducted monthly by the U.S. Bureau of the Census for the Bureau of Labor Statistics (BLS).

The CPS sample is a probability sample selected to be representative of the civilian, non-institutional population of the United States 16 years of age and older. Because of its very large size—currently about $60,000$ households are interviewed each month—the CPS allows for fairly fine-grained analysis of labor market trends. An adult (the reference person) at each household is asked to report on the activities of all other persons in the household. Thus, there is a record in the file for each adult person.

Each household entering the CPS is administered 4 monthly interviews, then ignored for 8 months, then interviewed again for 4 more months before leaving the sample permanently. In other words, in any given month every interviewed household has a "month in sample" ranging from 1 to 8 and an "interview month" that jumps discontinuously after the fourth interview. The survey design is such that each month an equal number of households belong to each of the eight groups, as defined by the month in sample variable. This rotation structure assures that 75% of the sample remains the same from one month to the next and 50% from one year to the same month the following year[1], which permits controlling for individual fixed effects if necessary.

Between 1973 and 1978, earnings questions were asked to the whole sample in May. Starting in 1979, questions regarding usual weekly earnings and usual weekly hours of work were asked every month but only to households in rotations 4 and 8. These households constitute the *outgoing rotation groups*. Each year the BLS gathers all these interviews into a single Merged Outgoing Rotation Groups (MORG) file. A consequence of this construction is that an individual appears only once in any file year, but may reappear (if found for interview) in the following year. To make the use of the CPS data easier, the NBER has compiled annual extracts of the files starting from 1979, making variable names and other aspects of the code uniform across all files (see Feenberg and Roth, 2007).

A.2 Weights

The CPS has a very complex sample design, whose main purpose is to attain national and state representativeness and make sure that employment statistics are highly accurate. The sample is drawn once a decade to meet the reliability criterion that the coefficient of variation on the national monthly unemployment rate is 1.9% assuming a 6% unemployment rate[2]. At the state level (including the District of Columbia) the coefficient of variation should be at most 8%. To meet these criteria the CPS sample selection follows a state-based and multi-stage process. Sample weights are constructed for each individual in order to:

---

[1]This is a rough approximation since technically what stays in the sample is the address and not the people living in it. Thus, the correspondence falls below 50% due to people moving out to a new location and due to non-response in the second year (See U.S. Census Bureau, 2002, 2006).

[2]The actual requirement is that a difference of 0.2% in the unemployment rate for two consecutive months be significant at the 90 percent confidence level (for details see U.S. Census Bureau, 2006; Polivka, 2000, p. 12).

1. Reflect the sample design.
2. Account for non-participation by some selected households.
3. Chance differences between the sample and known population parameters.

The weighting procedure involves four components. The first component is the base weight, which is equal to the inverse of the probability of selection. The second component is an adjustment to the base weight to account for intentional over- and under-sampling of certain sections of the population, as well as unintentional sampling rate differences that arise because size measures used in selecting Primary Sample Units (PSUs) and blocks within PSUs are not perfectly accurate. At this point the weights essentially reflect the probability of selection of the household. The third weight component is a non-interview adjustment, which inflates the weights to account for the fact that not all households selected agree to participate in the survey. Since the rate of survey non-response differs by demographic and economic subgroups, the non-interview adjustment is calculated separately for similar groups of households from the same sample areas. Note that non-response in this instance is *complete* non-response. In contrast, *item* non-response occurs when a household is interviewed but respondents only provide partial information. The difference is important because, whereas complete non-response leads to exclusion of the household from the sample, item non-response is generally corrected through the allocation of an imputed value (see the section on earnings imputation). The fourth and final weight component is designed both to reflect the fact that survey estimates of the distribution of demographic characteristics can differ by chance from known population distributions and to account for the fact that the size of the population changes over the course of a decade. Thus, this last adjustment entails assuring that the survey estimates of various demographic sub-populations agree with independently obtained adjusted census counts. This is accomplished by means of ratio adjustments[3]. The overall result of this process is the *final weight* variable used for most calculations using CPS data. Since 1979, most CPS files have included separate weights for the outgoing rotations. These weights were generally referred to as "earning weights" on files through 1993, and generally called "outgoing rotation weights" on files for 1994 and subsequent years (U.S. Census Bureau, 2006, 10-13). In addition to the already mentioned ratio adjustments, these weights also reflect additional constraints that force them to sum to the composite estimates of employment, unemployment, and not-in-labor-force each month. An individual's outgoing rotation weight will be approximately four times his or her final weight. This new weight is the one used in most studies of the U.S. earnings distribution.

A.3 Variables used

The CPS interview is divided into three basic parts:

1. Household and demographic information.

2. Labor force information.

3. Supplemental information in months that include supplements.

The questions in the first section lead to the construction of a household roster, as well as the gathering of basic information regarding the relatedness among people living in the household. In addition, the questionnaire asks respondents for other demographic data for each household member including: birth date, marital status, level of education, race, ethnicity, nativity, etc.

---

[3]There are five ratio adjustments in the CPS estimation process: the first-stage ratio adjustment, the national coverage adjustment, the state coverage adjustment, the second-stage ratio adjustment, and the composite ratio adjustment leading to the composite estimator.

A.3.1 Education Attainment

Educational attainment for each person in the household age 15 or older is obtained. Prior to 1992, the questions on educational attainment asked first about the highest grade or year of school each person had attended and, then, if they had completed that grade or year of school. Specifically, the item before 1992 had two parts. The first part asked, "What is the highest grade or year of regular school…has ever attended?". This was followed with the question, "Did…complete the grade?" The first question was topcoded at 6 years of college.

Starting in January 1992 the BLS switched from the years of schooling measure to a credential oriented measure[4]. The new (single) question is: "What is the highest level of school…has completed or the highest degree…has received?" In the new item, response categories for lower levels of schooling were collapsed into several summary categories. A new category ("12th grade, No Diploma") was added. But the major change in the item occurred in the categories for high school completion and beyond. Beginning with the response, "High School Graduate–high school diploma or the equivalent (for example, GED)," the categories identify specific degree completion levels, rather than years of schooling. Five different levels of degree attainment are identified: Associate (academic and vocational), Bachelors, Masters, Professional, and Doctoral degrees. A residual category of "some college but no degree" also is included.

Research based on the continuous variable "highest grade completed" must be adapted to the new way to represent educational attainment. There are two basic ways to "bridge" the old and new questions. First, it is possible to linearize responses to the new (categorical) question to provide a measure that is comparable to the

[4]Kominski and Siegel (1993) describe the history of the education question since the 1940 census. The "years of schooling completed" measure served for 50 years before the change in the 1990s.

"highest grade completed." Second, a categorical recoding scheme can be applied to both questions.

The resulting imputed highest grade completed for the linearization method can be seen in Table A.1. The method is based on matched data for March 1991 and 1992, where the same individual has answered both the old and the new educational attainment questions. The imputed value for highest grade completed is the median (and modal) highest grade completed of those individual answering that category in 1992. A slight adjustment is made for the lower categories (the method is explained in detail in Jaeger, 1997a,b).

Table A.1: Imputations of Highest Grade Completed for New Educational Attainment Question

CPS Code 92-97	CPS Code 1998-	Description	Imputed Grade Comp. 92-97	Imputed Grade Comp. 1998-
31	31	Less than 1^{st} grade	0	0
32	32	1^{st}, 2^{nd}, 3^{rd}, or 4^{th} grade	2.5	2.5
33	33	5^{th} or 6^{th} grade	5.5	5.5
34	34	7^{th} or 8^{th} grade	7.5	7.5
35	35	9^{th} grade	9	9
36	36	10^{th} grade	10	10
37	37	11^{th} grade	11	11
38	38	12^{th} grade, no diploma	12	12
39		High School Graduate–high school diploma, or the equivalent (e.g., GED)	12	
	39.1	High School Diploma		12

Table A.1 continued from previous page

CPS Code 92-97	CPS Code 1998-	Description		Imputed Grade Comp. 92-97	Imputed Grade Comp. 1998-
	39.2.1	GED and	Less than 1^{st} grade		0
	39.2.2	GED and	1^{st}, 2^{nd}, 3^{rd}, or 4^{th} grade		2.5
	39.2.3	GED and	5^{th} or 6^{th} grade		5.5
	39.2.4	GED and	7^{th} or 8^{th} grade		7.5
	39.2.5	GED and	9^{th} grade		9
	39.2.6	GED and	10^{th} grade		10
	39.2.7	GED and	11^{th} grade		11
	39.2.8	GED and	12^{th} grade, no diploma		12
40		Some college but no degree		13	
	40.1	Some college:	Less than 1 year		12
	40.2	Some college:	Freshman year completed		13
	40.3	Some college:	Sophomore year completed		14
	40.4	Some college:	Junior year completed		15
	40.5	Some college:	Four+ years but no diploma		16
41		Associate's degree in college–occupational/vocational school program		14	
42		Associate's degree in college–academic program		14	

Table A.1 continued from previous page

CPS Code 92-97	CPS Code 1998-	Description		Imputed Grade Comp. 92-97	Imputed Grade Comp. 1998-
	41/42.1	Associate degree +	Less than 1 year of college		12
	41/42.2	Associate degree +	1^{st} year of college completed		13
	41/42.3	Associate degree +	2^{nd} year of college completed		14
	41/42.4	Associate degree +	3^{rd} of college completed		15
	41/42.5	Associate degree +	Four+ years but no diploma		16
43	43.2	Bachelor's degree (e.g, B.A., A.B., B.S.)		16	16
	43.1.1	Bachelor's degree +	Graduate or Professional courses (0-5)		17
	43.1.2	Bachelor's degree +	Graduate or Professional courses (6-)		18
44		Master's degree (e.g., M.A., M.S., M.Eng., M.Ed., M.S.W., M.B.A.)		18	
	44.1	Master's degree:	1 year program		17
	44.2	Master's degree:	2 years program		18
	44.3	Master's degree:	3 years program		18
45	45	Professional school degree (e.g., M.D., D.D.S., D.V.M., L.L.B., J.D.)		18	18

Table A.1 continued from previous page

CPS Code 92-97	CPS Code 1998-	Description	Imputed Grade Comp. 92-97	Imputed Grade Comp. 1998-
46	46	Doctoral degree (e.g., Ph.D., Ed.D.)	18	18

Source: Jaeger (1997a) The post-1998 CPS codes are shown in summary form (actual codes are split into several variables).

In addition to using a measure of the highest grade completed, it is sometimes necessary to group individuals by more aggregated educational attainment categories. The four categories most often used are high school dropouts, high school graduates, individuals with some college, and college graduates. However, because the old question provides no information about whether 12^{th} graders graduated and obtained a diploma, the recommended category to use is "12^{th} grade" rather than "high school graduate"(see Jaeger, 1997b, p. 37). Table A.2 presents the categorization that provides the highest degree of matching between the recoded variables.

In 1998 the BLS began collecting data from an expanded set of educational attainment questions. Contingent on the answer to the base question ("highest grade received") eight additional questions were added. Individuals who respond that they have completed a high school diploma or equivalent are now also asked whether that diploma is a traditional high school degree or GED. If the latter is the case, an additional question asks for the highest grade that was completed prior to earning the GED. Similarly, individuals who answer they attended college but did not obtain a degree or that they completed an Associate Degree are now asked how many years of college they completed. Respondent who completed a Bachelor's degree are now asked if they took any graduate or professional school courses after graduating. If

Table A.2. Categorical Recoding Scheme for Old and New Educational Attainment Questions

Recoded Category	Old Question codes: highest grade attended		New Question Codes
	Not Completed	Completed	
High school dropout	0-12	1-11	31-37
Twelfth grade	...	12	38-39
Some College	13-16	13-15	40-42
College Graduate	17-18	16-18	43-46

Source: Jaeger (1997a)

so, they are asked whether they took 6 or more of those courses. Finally, individuals with a master's degree are now asked how many years their master's program takes.

The recommended imputation method (see Jaeger, 2003) can be found in the last column of Table A.1. For individuals who took the GED, the imputed highest grade completed is the grade that they finished before receiving their GED, while traditional high school degree recipients are assumed to have completed 12^{th} grade. Imputations for individuals who took some college course for credit or either type of an Associate's degree go from 12^{th} to 16^{th} grade completion. The highest grade completed is still topcoded at 18 to keep the series consistent. The NBER MORG files contain a variable that follows the imputation strategy described.

A.3.2 Employment Status and Type of Worker

Labor force information is obtained after the household and demographic information has been collected. One of the primary purposes of the labor force information is to classify individuals as employed, unemployed, or not in the labor force. Other information collected includes hours worked, occupation, and industry and related aspects of the working population. Those in the labor force or who have been in the labor force within the last 5 years (1989-1993 only, in the labor force or worked within the last year for 1994 onwards) are classified according to the type of employer.

The class of worker variable divides the eligible population among private employees (profit and non-profit), state employees (federal, state, and local), the self-employed (incorporated and not incorporated), and those that work without pay. Beginning with the major CPS redesign in January 1994, both actual and usual hours of work have been collected. Published data on hours of work relate to the actual number of hours spent "at work" during the reference week[5]. For example, persons who normally work 40 hours a week but were off on the Memorial Day holiday would be reported as working 32 hours, even though they were paid for the holiday. For persons with more than one job, the published figures correspond to the total number of hours worked at all jobs during the week. From 1994 on, the redesigned CPS allowed respondents to indicate their "hours vary". Typically 6-7% of workers respond their hours vary without being more specific. However, the vast majority of them indicate whether they are usually employed part-time or full-time. The NBER excludes these workers from the MORG extracts, which might lead to systematic bias in estimates if the distribution of hourly earnings for workers with varying hours differs from that of workers who report exact hours. A feasible alternative to exclusion involves merging the NBER files with "raw" CPS files that contain data on hours vary workers and then imputing hours to those workers using a regression approach. This approach does not seem to affect the distribution of hours much (see Schmitt, 2003, for details).

A.3.3 Industry and Occupation

For the employed, the industrial and occupational (I&O) information corresponds to the job held in the reference week. A person with two or more jobs is classified according to the job at which he or she worked the greatest number of hours. The unemployed are classified according to their last jobs. The universe for I&O is all

[5]The *reference week* is conventionally defined as the 7-day period, Sunday through Saturday, that includes the 12th of the month (U.S. Census Bureau, 2002, 2006).

private workers for pay, as defined by the class of worker variable. The I&O classification of CPS data has changed dramatically over the decades. Since 2000 the survey utilizes the 2000 North American Industrial Classification System (NAICS) industry codes and the census 2000 Standard Occupational Classification (SOC). Prior to that, the CPS had used the 3-digit Industry Classification Code and Occupational Classification from the 1980 and 1990 census (1983-2002, though there were some changes in 1992). The earliest classifications used in MORG files are those of the 1970 census (1979-1982, also 3 digits). The main changes in these different classifications are reviewed in a subsection below.

A.3.4 Earnings

As already mentioned, information on what people earn at their main jobs is collected for those who are receiving their fourth and eighth monthly CPS interviews only[6]. The BLS processing of the CPS treats earnings of "hourly" and "non-hourly" workers differently. Until 1994, hourly workers are those paid by the hour. After the 1994 redesign, the MORG files report hourly earnings for any worker "hourly paid" or otherwise, for whom it was easiest to report earnings by the hour. One difference between hourly and non-hourly workers involves the top-coding (discussed below). In addition, a second important inconsistency is that hourly workers' earnings are reported as "straight time" pay per hour, which excludes overtime, tips, and commissions (OTC). For all workers, the BLS also reports *weekly* earnings including OTC. Thus, there are two earning variables: an hourly earnings variable that excludes OTC (non-hourly workers have missing values in this variable) and a weekly earnings variable that includes OTC (with non-missing data for the vast majority of employed workers). Compounding this problem, the 1994 redesign changed the way the earn-

[6] Yearly income and other questions are asked to the whole sample as part of the March demographic supplement. A comparison of the different income data sources can be found in Katz and Autor (1999).

ings questions are administered. Before 1994, all respondents were asked to report their usual earnings before taxes and other deductions and to include any overtime pay, commissions, or tips usually received [7]. After the 1994 redesign, respondents may report earnings in the time period they prefer (hourly, weekly, biweekly, monthly, or annually). Based on additional information collected during the interview, earnings reported on a basis other than weekly are converted to a weekly amount in later processing. Data are collected for wage and salaried workers (excluding the self-employed who respond that their businesses were incorporated). Individuals also are asked a specific question to determine if they receive overtime pay, tips or commissions. If individuals indicate that they do receive OTC earnings, a lead-in is included in the earnings amount question reminding respondents to include them (Polivka, 1996). The most straightforward strategy to correct for this inconsistency is to add OTC earnings to the straight hourly pay for hourly workers. A feasible way to do this involves creating an hourly wage estimate using the weekly earnings variable, divided by "usual weekly hours worked". These estimates would thus include OTC. In practice, however, a large share of the resulting estimated hourly wages are lower than the reported straight time hourly wages. In those cases, the latter data are preferred. A second strategy, available only after the 1994 redesign, is to use the new questions that specifically ask about earnings from OTC and to add those amounts to the straight time hourly wage (Schmitt, 2003). The procedure involves comparing the hourly wage that results from the straightforward strategy to the one that results from using the post-1994 data and picking the larger amount. In order to control for outliers in the form of impossibly high OTC amounts, the straightforward strategy

[7] The term "usual" was as perceived by the respondent. If the respondent asked for a definition of usual, however, interviewers were instructed to define the term as more than half the weeks worked during the previous 4 or 5 months.

estimate is kept in cases when the post-1994 OTC-inclusive estimate exceeds some upper-bound (normally 4 times the straightforward strategy wage rate).

A.3.4.1 Topcoding

As already explained, earnings in the MORG files are reported at two intervals: on an hourly basis for "hourly workers" and on a weekly basis for all other workers. A problem creating a consistent hourly earnings series is that while hourly workers' wages are generally topcoded at $99.99 per hour (a threshold rarely crossed), weekly earnings are topcoded at much lower thresholds. As a result, an important share of workers' earnings data is censored and replaced with the value of the topcode. Even more problematically, the topcode has been changed at discrete intervals (see table A.3). During periods when the topcode remains constant (1979-88, 89-97, and 98 onwards), the share of topcoded earnings increases monotonically. In order to provide a method to address distortions in mean earnings caused by topcoding, it has been usual practice to fit a Pareto distribution to the upper tail of the earnings data[8]. The method involves estimating the α parameter of the distribution using the available earnings data and then obtaining and estimate for mean earnings above the topcode (for details see: Polivka, 2000; West, 1986). This is the preferred approach [add estimated mean to table].

A.3.4.2 Non-response, proxy-response, edits and imputations

In cases when earnings data is not provided by a respondent[9] the Census Bureau allocates a value using a "cell hot deck" imputation method. The census creates cells based on the following seven categories: gender (2 cells), age group (6), race (2),

[8]A simpler approach involves multiplying the topcode by a constant (ranging from 1.3 to 1.5) and replacing topcoded data with this "estimated" mean. Schmitt (2003) recommends fitting a log-normal to the whole distribution of earnings, rather than a Pareto to the tale.

[9]This case is referred to as item non-response, as opposed to full non-response. Item non-response generally leads to imputation. Full non-response leads to deletion from the sample.

Table A.3. CPS Topcode

(Weekly Earnings in nominal dollars)

Year	Topcode	Share Topcoded (%)
1979	999	1.3
1980	999	1.6
1981	999	2.3
1982	999	3.2
1983	999	4.2
1984	999	5.2
1985	999	6.1
1986	999	7.4
1987	999	8.7
1988	999	10.3
1989	1,923	1.1
1990	1,923	1.3
1991	1,923	1.6
1992	1,923	1.8
1993	1,923	2.0
1994	1,923	2.8
1995	1,923	3.1
1996	1,923	3.3
1997	1,923	3.9
1998	2,884	1.5
1999	2,884	1.7
2000	2,884	2.0
2001	2,884	2.2
2002	2,884	2.5

Source: Schmitt (2003, p.33)

education group (3), occupation (13), hours worked (8), and receipt of tips, commissions, or overtime (2), a matrix of 14,976 possible combinations[10]. All cells are kept "stocked" with a donor, insuring that an exact match is always found. The donor in each cell is the most recent person surveyed by the census with reported earnings and all the characteristics. When a new person with those characteristics is surveyed and reports earnings, the census replaces the previous occupant of the cell. The search for a donor reaches as far back as necessary within a given survey month and then to previous months and years. When surveyed individuals do not report earnings, their earnings are imputed the value of (nominal) earnings reported by the current donor occupying the cell with exact match characteristics. The "edited" earnings variable that results contains both respondents and allocated earnings. Allocation flags designate which individuals have reported earnings and which imputed earnings. In the period 1979-1988 the proportion of allocated earnings was in the range of 10-17%, with no clear tendency upward or downward. Beginning in January 1989, earnings allocation flags included in the MORG files are unreliable (only around 4% of workers are designated as having imputed earnings). Because the files contain an "unedited" weekly earnings variable as well, it is possible to device an alternative imputation flag (designating those with missing unedited earnings and valid edited earnings[11]). Based on this method, about 15% of workers had earnings imputed during 1989-92 and almost 17% in 1993. After the 1994 redesign, two new cells ("hours full-time" and "hours part-time") were included in the cell hot-deck. As a consequence, there are no usable earnings allocation flags for January 1994 through August 1995 (the

[10]The selection categories have not been identical over time.

[11]An additional reason to ignore the original flag is that some workers designated as allocated have non-missing unedited and edited weekly earnings whose values are equivalent. See Hirsch and Schumacher (2004, p.703), for details.

unedited weekly earnings variable is not provided[12]). For the period September 1995 through 1998, 22%-24% of individuals had imputed earnings. The series of earnings questions became more complex following the redesign, increasing the rate of non-response (Polivka, 1996). The rate of non-response kept growing as is today above 30%. The Census and the BLS include earnings of both respondents and non-respondents in published tabulations of earnings and other outcomes. Researchers typically do the same in the belief that biases due to imputation are low and the efficiency gains of a larger sample are high. However, the hot-deck procedure might introduce different types of biases. For example, in standard earnings equations or estimated wage differentials there will be attenuation or "match" bias toward zero for characteristics that are not imputation match criteria (e.g., union status). The attenuation is a first order problem, roughly equal to the sample proportion with imputed earnings, and independent of possible response bias (the earnings of donors being systematically different from those of recipients within a cell[13]). Bollinger and Hirsch (2006) analyze the effect of match bias and, most important, suggest possible solutions. The recommended approach involves excluding the allocated earnings from the sample when studying any characteristic that is not an explicit match criteria or that is only imperfectly matched[14].

---

[12]According to Schmitt (2003, p.17), the reason flags are not provided is that no earnings were imputed during the period (the hot-deck was being filled with donors).

[13]Bollinger and Hirsch (2007) study the issues of response bias and proxy respondents.

[14]Education, for example, enters the matching criteria at a level more aggregated (educational groups) than that available.

A.3.5 Union Membership

Beginning in January 1983, the CPS asked questions regarding union membership and coverage to the outgoing rotation groups[15]. The universe for whom union membership figures can be compiled is all employed civilian wage and salary workers, ages 16 and over. There are two union status questions. The first asks for *membership* to a union or an employee association similar to a union. Respondents who answer negatively to the first question are asked whether they are *covered* by a union employee association contract. Normally union coverage is considered a wider category encompassing both union members and non-members whose working conditions are regulated by a collective contract negotiated by a union.

A.4 Changes in the I&O Classification Codes

From 1940 to 1990 the basic structure of the industry classification system used in the censuses of population was generally the same. The census system in each of these years was based on the structure of the Standard Industrial Classification (SIC) updated during each decade and used throughout the Federal Government during that time period. The occupational classification had a similar structure from 1940 to 1960. For 1970 the occupation classification was enlarged by almost 50 percent from 297 categories in 1960 to 441 categories in 1970 because of requests from data users for more detail. In the 1960 system eight large "not elsewhere classified" (n.e.c.) categories contained one third of the labor force. The task in preparing the 1970 classification was to search these large "n.e.c" categories for occupational groups that could be identified separately. The revision for 1980 added another 74 new categories but deleted 12 allocation (semi-imputed) categories for a net increase of 62. The 1980

[15] Prior to that, for the period 1973-81, the May supplement of the CPS asked the union question to the whole sample. There is no union data in the CPS during 1982. For details see Hirsch and Macpherson (2003).

occupation classification also was a major departure from earlier censuses because of the adoptions of the Standard Occupational Classification (SOC) by federal agencies, which became the model for the 1980 census classification. The 1990 system had 501 categories, also based on the 1980 SOC. There was not much change, therefore, between the 1980 and 1990 census occupational classifications (Scopp, 1989, 2003; Levine et al., 1999).

After 1990, however, the crosswalk tables converting the industry and occupation data from one of these past censuses to the classification systems of the previous or subsequent census became increasingly more necessary because for the first time both the standard industry and occupation classifications underwent major revisions in the same decade. The 1987 SIC was replaced in 1997 by the North American Industrial Classification System (NAICS), and the 1980 SOC was replaced in 1998 by a completely revamped SOC. The latter then evolved into a slightly modified update in 2000. The 1997 NAICS and 2000 SOC, respectively, provided the structure for the Census 2000 I&O classifications. Here we focus on the changes to the occupational classification.

In early classification systems too much emphasis was placed on the industry in which one worked. While it is true that the work setting can influence the job, it is the hallmark of more recent classification systems that characteristics of the work performed comes first. The 2000 SOC replaced the 1980 SOC to reflect the dramatic changes in the US labor force over the previous two decades. Like the 1980 SOC, the new classification covers all occupations in which work is performed for pay or profit, including work performed in family-operated enterprizes. Occupations are classified based on work performed and on required skills, education, training, and credentials. The committee in charge of the new SOC decided to completely rearrange the structure of the classification rather than to start with the old SOC and simply try to make improvements. The world of work was arranged into "job families," in which people who work together are classified together, regardless of their skill level. For example, physicians, registered nurses, and medical laboratory technicians are all in the same SOC and census major group in 2000. On the other hand, while first-line supervisors are found in the same major groups as the workers they supervise, higher management levels are not. Managers are in their own major group. The following list shows the 23 major occupational groups of the revised SOC (the 1980 SOC used 22):

1. Management occupations
2. Business and financial operations occupations
3. Computer and mathematical occupations
4. Architecture and engineering occupations
5. Life, physical, and social science occupations
6. Community and social services occupations
7. Legal occupations
8. Education, training, and library occupations
9. Arts, design, entertainment, sports, and media occupations
10. Healthcare practitioners and technical occupations
11. Healthcare support occupations
12. Protective service occupations
13. Food preparation and serving related occupations
14. Building and grounds cleaning and maintenance occupations
15. Personal care and service occupations
16. Sales and related occupations
17. Office and administrative support occupations
18. Farming, fishing, and forestry occupations
19. Construction and extraction occupations
20. Installation, maintenance, and repair occupations
21. Production occupations
22. Transportation and material moving occupations
23. Military specific occupations

These major groups include 98 minor groups, 452 broad occupations, and 822 detailed occupations. Occupations with similar skills or work activities are grouped at each of the four levels of hierarchy to facilitate comparisons. For example, the major group, life, physical, and social science occupations, is divided into four minor groups-life scientists, physical scientists, social scientists and related workers, and life, physical, and social science technicians. "Life scientists" contains broad occupations, such as agriculture and food scientists, as well as biological scientists. The broad occupation, biological scientists, includes detailed occupations such as biochemists and biophysicists as well as microbiologists. The following example shows the hierarchical structure of the 1998 SOC:

19-0000 Life, physical, and social science occupations (major group)

19-1000 Life scientists (minor group)

19-1020 Biological scientists (broad occupation)

19-1021 Biochemists and biophysicists (detailed occupation)

19-1022 Microbiologists (detailed occupation)

19-1023 Zoologists and wildlife biologists (detailed occupation)

Broad occupations often include several detailed occupations that are difficult to distinguish without further information.

APPENDIX B

DATA PROCESSING FOR CHAPTER 4

B.1 Basic Processing of May/ORG CPS and DOT Data

Data on skill requirements comes from the Dictionary of Occupational Titles 4th Edition (1977) and revised 4th Edition (1991). We use the dataset compiled by Levy and Murnane (1992) that contains weighted averages of three GED scores (language, reasoning, and math) by occupation and sex using both the 1970 and 1980 3-digit occupational classifications. Only the highest GED is binding so we drop the other two. Scores for years other than 1977/91 are linearly extrapolated. The 1970 and 1980 Census occupational classifications are available in the CPS only during the period 1973–2002. Thus, we use the May CPS for 1973–78 and the merged outgoing rotation groups for 1979–2002. The general inclusion criteria are: age in the range 18–65, to have worked in the past, and potential experience between 1 and 40 years (this inclusion criteria will be referred to as counts sample). Calculations that involve earnings are done using the standard earnings weight multiplied by usual weekly hours.

Our wage variable is the log of real hourly earning in 1979 dollars (deflated using the CPI-U-RS). Hourly earnings are weekly earnings divided by usual weekly hours with the exception of cases in which a separate higher hourly wage is reported. After 1994 individuals are allowed to answer that their hours vary. We use a simple regression imputation approach to assign hours to those individuals. No allocated earnings are utilized, however. During the period 1989–93 the allocation flags fail to identify most imputed earnings. Following Lemieux (2006a), we use the unedited earnings variable to identify and drop unflagged allocated earnings. Topcoded earnings are winsorized using a 1.4 factor.

B.2 Construction of Relative Wage Series

We use the same method as in Autor et al. (2008) to calculate composition-adjusted relative wage series. Only full-time employed wage and salary workers are considered. For each year and sex, we regress log hourly wages on schooling dummies (DO, HS, SC, CO, AD), a quartic in potential experience, a minority (non-white) indicator, 9 regions, and interaction terms of the experience quartic with the high-tech dummy and four education dummies (HS, SC, CO+). We divide the sample into 40 cells: 2 sexes, 5 education levels, 4 experience ranges (1–10, 11–20, 21–30, 31-40). For each cell and year we compute the predicted log wage for whites in the most frequent region and at the mid-point of the experience range. The composition adjusted annual series of college and high-school wages results from a fixed-weight average of these predicted log wages. The fixed weights are equal to the proportion of hours supplied by the cell over the 30 year period, calculated using a count sample of all employed for pay workers (inclusive of self-employed). This adjustment should take care of distortions to measured relative wages that result from differential changes in the composition of workers at different educational levels.

The procedure to calculate the composition-adjusted hi/low-tech log wage gap is analogous. We include a hi-tech ($GED \geq 4$) indicator among the regressors. We also add high/low-tech to the criteria for cell formation (80 cells rather than 40).

B.3 Construction of Relative Employment and Relative Supply Measures

In order to estimate relative employment of College/HS equivalents, we divide the count sample into 400 cells (same as before but with 40 single-year experience categories) and compute total hours supplied per cell per year for all employed workers in the sample. Broader employment aggregates can be obtained in efficiency units. Hours supplied by high-school males with 10 years of experience are taken as the unit of account. The mean relative wage of other cells with respect to high-school males with 10 years of experience is used as a conversion factor for their hours. College equivalent hours are hours in efficiency units supplied by college graduates, advanced degree workers, and half the hours supplied by workers with some college. All other hours supplied are considered high-school equivalent. The relative employment of high/low-tech workers is obtained following an analogous procedure. The sample is divided into 800 cells (2 high/low-tech, 2 sexes, 5 education groups, 40 experience groups). Efficiency units are computed in terms of low-tech high-school males with 10 years of experience.

The overall supply ratio (H/L) is calculated as the ratio of the simple count of college to high-school equivalents.

BIBLIOGRAPHY

Daron Acemoglu. Technical change, inequality, and the labor market. *Journal of Economic Literature*, 40(1):7–72, 2002.

James Albrecht and Susan Vroman. A Matching Model with Endogenous Skill Requirements. *International Economic Review*, 43(1):283–305, Feb 2002.

David Autor, Lawrence Katz, and Melissa Kearney. Trends in U.S. Wage Inequality: Revising the Revisionists. *The Review of Economics and Statistics*, 90(2):300–323, May 2008.

David H. Autor, Frank Levy, and Richard J. Murnane. The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics*, 118:1279–334, 2003.

Thomas K. Bauer. Educational mismatch and wages: A panel analysis. *Economics of Education Review*, 21(3):221–229, Jun 2002.

Stephen Bazen, Claudio Lucifora, and Wiemer Salverda, editors. *Job Quality and Employer Behavior*. Palgrave Macmillan, 2005.

Gary Becker. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education.* The University of Chicago Press, New York, 1964.

Ivan Berg. *Education and Jobs: The Great Training Robbery.* Beacon, Boston, 1971.

Truman F. Bewley. *Why Wages Don't Fall During a Recession*. Harvard University Press, 1999.

William M. Boal and Michael R. Ransom. Monopsony in the Labor Market. 35(1): 86–112, Mar 1997.

Christopher R. Bollinger and Barry T. Hirsch. Match bias from earnings imputation in the current population survey: The case of imperfect matching. *Journal of Labour Economics*, 24(3):483–519, 2006.

Christopher R. Bollinger and Barry T. Hirsch. How well are earnings measured in the current population survey? bias from nonresponse and proxy respondents. *North American Summer Meetings of the Econometric Society*, Jun 2007. Available at `https://zeus.econ.umd.edu/cgi-bin/conference/download.cgi?db_name=NASM2007&paper_id=733`.

John Bound and George Johnson. Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations. *The American Economic Review*, 82 (3):371–392, 1992.

Samuel Bowles. The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian Models. *The American Economic Review*, 75(1):16–36, 1985.

Samuel Bowles and Herbert Gintis. The problem with human capital theory–a marxian critique. *American Economic Review*, 65(2):74–82, May 1975.

Samuel Bowles and Herbert Gintis. *Schooling in Capitalist America*. Basic Books, New York, 1976.

Samuel Bowles and Herbert Gintis. Schooling in Capitalist America Revisited. *Sociology of Education*, 75(1):1–18, Jan 2002.

H. Braverman. *Labor and Monopoly Capital*. Monthly Review, New York, 1974.

Felix Büchel. The effects of overeducation on productivity in germany the firms' viewpoint. *Economics of Education Review*, 21(3):263–275, Jun 2002.

Peter Cappelli. Are skill requirements rising? evidence from production and clerical jobs. *Industrial and Labor Relations Review*, 46(3):515–530, Apr 1993.

David Card and JohnE. DiNardo. Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of Labor Economics*, 20(4):733–783, 2002.

David Card and Alan B. Krueger. *Myth and Measurement. The New Economics of the Minimum Wage*. Princeton University Press, Priceton, New Jersey, 1995.

David Card, Francis Kramarz, and Thomas Lemieux. Changes in the relative structure of wages and employment: A comparison of the united states, canada, and france. *The Canadian Journal of Economics*, 32(4):843–877, 1999.

Arnaud Chevalier. Measuring over-education. *Economica*, 70:509–531, 2003.

Clifford C. Clogg and James W. Shockey. Mismatch Between Occupation and Schooling: A Prevalence Measure, Recent Trends and Demographic Analysis. *Demography*, 21(2):235–257, May 1984.

Elchanan Cohn. The Impact of Surplus Schooling on Earnings: Comment. *The Journal of Human Resources*, 27(4):679–682, 1992.

Elchanan Cohn and Shahina P. Khan. The wage effects of overschooling revisited. *Labour Economics*, 2(1):67–76, Mar 1995.

John DiNardo, Nicole Fortin, and Thomas Lemieux. Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach. *Econometrica*, 64(5):1001–44, 1995.

Peter Dolton and Anna Vignoles. The incidence and effects of overeducation in the u.k. graduate labour market. *Economics of Education Review*, 19(2):179–198, Apr 2000.

Arindrajit Dube, T. William Lester, and Michael Reich. Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties. *SSRN eLibrary*, 2007.

Greg J. Duncan and Saul D. Hoffman. The incidence and wage effects of overeducation. *Economics of Education Review*, 1(1):75–86, 0 1981.

Daniel Feenberg and Jean Roth. CPS labor extracts 1979–2006. *NBER*, Jan 2007. Available at `http://www.nber.org/morg/docs/cpsx.pdf`.

Gary Fields. *Accounting For Income Inequality and its Change: A New Method, with Application to the Distribution of Earnings in the United States*, volume 22, pages 1–38. Research in Labor Economics, 2003.

Richard B. Freeman. *The Over-educated American.* Academic Press, New York, 1976.

Claudia Goldin and Lawrence F. Katz. The origins of technology-skill complementarity. *The Quarterly Journal of Economics*, 113(3):693–732, Aug 1998.

Francis Green, Steven McIntosh, and Anna Vignoles. 'Overeducation' and Skills – Clarifying the Concepts. *CEP Discussion Paper*, 435, Sep 1999.

Wim Groot. Overeducation and the returns to enterprise-related schooling. *Economics of Education Review*, 12(4):299–309, Dec 1993.

Wim Groot and Henriëtte Maassen van den Brink. Overeducation in the labor market: A meta-analysis. *Economics of Education Review*, 19(2):149–158, Apr 2000.

Frederick Guy and Peter Skott. Information and communications technologies, coordination and control, and the distrbution of income. *Journal of Income Distribution*, 17(3–4):71–92, 2008.

Michael J. Handel. Skills Mismatch In The Labor Market. *Annual Review of Sociology*, 29(1):135–165, 2003.

Joop Hartog. Over-education and earnings: Where are we, where should we go? *Economics of Education Review*, 19(2):131–147, Apr 2000.

Daniel E. Hecker. Reconciling conflicting data on jobs for college graduates. *Monthly Labor Review*, pages 3–12, Jul 1992.

Barry T. Hirsch and David A. Macpherson. Union membership and coverage database from the current population survey: Note. *Industrial and Labor Relations Review*, 56(2):349–354, Feb 2003.

Barry T. Hirsch and Edward J. Schumacher. Match bias in wage gap estimates due to earnings imputation. *Journal of Labour Economics*, 22(3):689–722, 2004.

David Howell. Theory-driven facts and the growth in earnings inequality. *Review of Radical Political Economics*, 31(54), 1999.

David R. Howell and Edward N. Wolff. Trends in the growth and distribution of skills in the u.s. workplace, 1960-1985. *Industrial and Labor Relations Review*, 44 (3):486–502, 1991.

David A. Jaeger. Estimating the returns to education using the newest current population survey education questions. *Economic Letters*, 78:385–394, 2003.

David A. Jaeger. Reconciling the old and new census bureau education questions: Recommendations for researchers. *Journal of Business & Economic Statistics*, 15 (4):300–9, Jul 1997a.

David A. Jaeger. Reconciling educational attainment questions in the cps and the census. *Monthly Labor Review*, pages 36–40, Aug 1997b. Technical Note.

Jack Johnston and John DiNardo. *Econometric Methods*. McGraw-Hill, New York, fourth edition, 1997.

Lawrence Katz and David Autor. Changes in the wage structure and earnings inequality. In Orley Ashenfelter and David Card, editors, *Handbook of Labor Economics*, volume 3A, pages 1463–1555. Elsevier Science B.V., Amsterdam, 1999.

Lawrence Katz and Kevin Murphy. Changes in Relative Wages, 1963–87: Supply and Demand Factors. *The Quarterly Journal of Economics*, (107):35–78, Feb 1992.

Robert Kominski and Paul M. Siegel. Measuring education in the current population survey. *Monthly Labor Review*, 116(9):34–8, Sep 1993. Research Summaries.

David Lee. Wage inequality in the us during the 1980s: Rising dispersion or falling minimum wage? *Quarterly Journal of Economics*, (114):941–1024, 1999.

Thomas Lemieux. Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill? *The American Economic Review*, 96(3):461–498, Jun 2006a.

Thomas Lemieux. The Mincer Equation Thirty Years after Schooling, Experience, and Earnings. chapter 11. Springer Verlag, 2006b.

Ch. Levine, L. Salmon, and D. H. Weinberg. Revising the standard occupational classification system. *Monthly Labor Review*, pages 36–45, May 1999.

Frank Levy and Richard Murnane. U.S. earning levels and earnings inequality: A review of recent trends and proposed explanations. *Journal of Economic Literature*, (30):1333–1381, 1992.

Brigitte C. Madrian and Lars J. Lefgren. A Note on Longitudinally Matching Current Population Survey (CPS) Respondents. *NBER Technical Working Paper*, (247), 1999.

Alan Manning. *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press, 2003.

Alan Manning. Monopsony and the Efficiency of Labour Market Interventions. *Labour Economics*, 11:145–163, 2004.

Stephen A. Marglin. What do bosses do?: The origins and functions of hierarchy in capitalist production. *Review of Radical Political Economics*, 6:60–112, Jul 1974.

Séamus McGuinness. Overeducation in the Labour Market. *Journal of Economic Surveys*, 20(3):387–418, Jul 2006.

Jakob Mincer. *Schooling, Experience, and Earnings*. NBER, New York, 1974.

Paul Osterman. Skill, training, and work organization in american establishments. *Industrial Relations*, 34(2):125–146, Apr 1995.

Anne E. Polivka. Using earnings data from the monthly current population survey. *Bureau of Labor Statistics*, Oct 2000. Available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=261190`.

Anne E. Polivka. Data watch: The redesigned current population survey. *Journal of Economic Perspectives*, 10(3):169–180, Summer 1996.

G Psacharopoulos. Returns to Education: an updated international comparison. *Comparative Education*, 17(3):321–341, 1981.

James B. Rebitzer and Lowell J. Taylor. The consequences of minimum wage laws. Some new theoretical ideas. *Journal of Public Economics*, 56(2):245–255, Feb 1995.

S. Rubb. Overeducation in the labor market: A comment and re-analysis of a meta-analysis. *Economics of Education Review*, 22(6):621–629, Dec 2003.

Michael Sattinger. Overlapping labour markets. *Labour Economics*, 13:237–257, 2006.

Michael Sattinger. Assignment models of the distribution of earnings. *Journal of Economic Literature*, 31(2):831–880, 1993.

John Schmitt. Creating a consistent hourly wage series from the current population survey's outgoing rotation group, 1979-2002. *CEPR Working Paper*, Aug 2003. Available at `http://www.ceprdata.org/cps/CEPR_ORG_Wages.pdf`.

Theodor W. Schultz. *Investment in Human Capital*. The Free Press, New York, 1971.

T.S. Scopp. The relationship between the 1990 census and census 2000 industry and occupation classification systems. *U.S. Census Bureau Technical Paper*, (65), 2003.

T.S. Scopp. The relationship between the 1970 and 1980 industry and occupation classification systems. *U.S. Census Bureau Technical Paper*, (59), 1989.

C. Shapiro and J. Stiglitz. Equilibrium Unemployment as a Worker Discipline Device. *The American Economic Review*, 74(3):433–444, 1984.

Kristina J. Shelly. The future of jobs for college graduates. *Monthly Labor Review*, pages 13–21, Jul 1992.

A. F. Shorrocks. Inequality Decomposition by Factor Components. *Econometrica*, 50 (1):193–211, Jan 1982.

Nachum Sicherman. "Overeducation" in the Labor Market. *Journal of Labor Economics*, 9(2):101–122, Apr 1991.

Frank Siebern-Thomas. Job Quality in European Labour Markets. In Bazen et al. (2005), chapter 2, pages 31–66.

Peter Skott. Fairness as a source of hysteresis in employment and relative wages. *Journal of Economic Behavior and Organization*, 57(3):305–331, Jul 2005.

Peter Skott. Wage Inequality and Overeducation in a Model with Efficiency Wages. *Canadian Journal of Economics*, 39(1):94–123, Feb 2006.

Peter Skott and Paul Auerbach. *Wage Inequality and Skill Assymmetries*, chapter 2, pages 27–54. Interactions in Analytical Political Economy: Theory, Policy and Applications. M.E. Sharpe, 2004.

Peter Skott and Frederick Guy. A Model of Power-Biased Technological Change. *Economics Letters*, 95(1):124–131, Apr 2007.

Peter J. Sloane. Much ado about nothing? What does the Overeducation Literature Really Tell us? chapter 2, pages 11–48. Edwad Elgar, Northampton, MA, 2003.

Fabian Slonimczyk. Earnings Inequality and Skill Mismatch in the U.S: 1973–2002. *Working Paper*, 2008a.

Fabian Slonimczyk. Skill Mismatch and Earnings: A Panel analysis of the U.S. Labor Market, 1983–2002. *Working Paper*, 2008b.

Michael Spence. Job Market Signaling. *The Quarterly Journal of Economics*, 87(3): 355–374, Aug 1973.

Kenneth I. Spenner. Deciphering Prometheus: Temporal Change in the Skill Level of Work. *American Sociological Review*, 48(6):824–837, 1983.

Kenneth I. Spenner. The Upgrading and Downgrading of Occupations: Issues, Evidence, and Implications for Education. *Review of Educational Research*, 55(2): 125–154, 1985.

G. Stigler. The Economics of Minimum Wage Legislation. *The American Economic Review*, 36:358–365, 1946.

Lester Thurow. *Generating Inequality*. Basic Books, New York, 1975.

Yuping Tsai. Returns to Overeducation: A Longitudinal Analysis in the U.S. Labor Market. *SSRN eLibrary*, 2007. Available at `http://ssrn.com/paper=1087427`.

U.S. Census Bureau. Current population survey: Design and methodology. *Technical Paper*, 63V, 2002. Available at `http://www.bls.census.gov/cps/tp/tp63.htm`.

U.S. Census Bureau. Current population survey design and methodology. *Technical Paper*, 66, 2006. Available at `http://www.census.gov/prod/2006pubs/tp-66.pdf`.

U.S. Department of Labor. *Dictionary of Occupational Titles*. Government Printing Office, 4th edition, 1977.

U.S. Department of Labor. *Dictionary of Occupational Titles*. Government Printing Office, revised 4th edition, 1991.

Stephen Vaisey. Education and its discontents: Overqualification in america, 1972-2002. *Social Forces*, 85(2):835–864, Dec 2006.

Richard R. Verdugo and Naomi Turner Verdugo. The impact of surplus schooling on earnings: Some additional findings. *The Journal of Human Resources*, 24(4): 629–643, 1989.

Sandra A. West. Estimation of the mean from censored income data. *Bureau of Labor Statistics Report*, 1986. Available at `http://www.amstat.org/sections/SRMS/Proceedings/papers/1986_126.pdf`.

Edward Wolff. Technology and demand for skills. chapter 2, pages 27–56. Edward Elgar, Northampton, MA, 2000.

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, MA, 2002.