The Wage Effects of Job Polarization: Evidence from the Allocation of Talents

Michael J. Boehm *

October 2013

Abstract Over the last decades, the United States and other developed countries have experienced profound job polarization whereby employment in high-skill and low-skill occupations increased at the expense of employment in middle-skill occupations. This paper examines the wage effects of job polarization: first, have the relative wages of workers in middle-skill occupations declined as the prevalent demand-side explanation for job polarization predicts? Second, has the relative price paid per unit of effective labor in the middle-skill occupations dropped with polarization? Third, can job polarization explain the changes in the overall wage distribution over that time period? I answer these questions by comparing over time two representative cohorts of young workers in the United States for whom I can consistently measure relative skills in occupations. My results show that the answer to the first two questions is yes while the answer to the third question is partly yes and partly inconclusive.

Keywords: Job Polarization; Wage Inequality; Talent Allocation; Roy Model 

JEL Classification Numbers: J21, J23, J24, J31

* University of Bonn and Centre for Economic Performance (LSE), mboehm1@uni-bonn.de. This is a fundamental revision of my job market paper entitled “Has Job Polarization Squeezed the Middle Class? Evidence from the Allocation of Talents”. I am extremely grateful to my supervisor Steve Pischke. I very much thank Yona Rubinstein, Alan Manning, Luis Garicano, David Autor, David Dorn, Pedro Carneiro, John Van Reenen, Guy Michaels, Esteban Aucejo, Barbara Petrongolo, Claudia Steinwender, Johannes Boehm, Georg Graetz, Martin Watzinger, and Pinar Hosafci as well as participants at various seminars and conferences for their comments. I also thank Joseph Altonji, Prashant Bharadwaj, and Fabian Lange for sharing their data and code, and Steve McClaskie for support with the NLSY data.
1 Introduction

Over the last two decades, the labor market in the United States has experienced a profound polarization of employment. In particular, the aggregate share of jobs in middle-skill production and clerical occupations has declined by almost seven percentage points since the end of the 1980s while the share of jobs in high-skill professional and managerial occupations as well as low-skill services occupations has surged. This has coincided with a u-curved change in wage inequality, whereby earnings in the middle of the wage distribution have stagnated or even fallen while earnings at the top and at the bottom have increased substantially (compare Acemoglu and Autor 2011 and see figures 1 and 2 for the data used in this paper).

Pioneering research by Autor, Levy, and Murnane (2003) and other\footnote{Goos and Manning (2007), Autor, Katz, and Kearney (2006), Michaels, Natraj, and Van Reenen (2013), Acemoglu and Autor (2011), Autor and Dorn (2013).} has shown that the main driver of job polarization was a rapid improvement of computer technology which could replace the “routine” work that is intensively carried out in middle-skill occupations\footnote{Offshoring of production and international trade, especially with China, seem to be an additional and related driver of job polarization (e.g. Acemoglu and Autor 2011, Autor, Dorn, and Hanson 2012).}. It thus seems natural to hypothesize that the u-curved change in the wage distribution was shaped by the same demand-side forces (e.g. Autor, Katz, and Kearney 2006, Acemoglu and Autor 2011, Autor and Dorn 2013). However, to date there exist hardly any empirical studies that have established a close link between job polarization on the one hand and workers’ wages and the wage distribution on the other.

The purpose of this paper is therefore to analyze the wage effects of job polarization. In particular, I ask the following questions: first, have the relative wages of workers in the middle-skill occupations declined as the demand-side explanation for job polarization predicts? Second, how have the prices paid for a constant unit of effective labor in the high-, middle-, and low-skill occupations changed with polarization? Third, what was the effect of job polarization on the overall wage distribution and, in particular, could it have generated the u-curved change of inequality that we observe in the data?

In order to answer these questions, I construct two representative cross-sections
of 27 years old male workers at the end of the 1980s and at the end of the 2000s from the cohorts of the National Longitudinal Survey of Youth (NLSY79 and NLSY97). The advantage of the NLSY over more commonly used data sources is that it provides consistent, early-determined, and multidimensional measures of worker talents—such as mathematical, verbal, and mechanical test scores and risky behaviors—which predict occupational choices and wages. This allows me to study the wages of the kind of individuals who are more and less likely to work in the middle-skill compared to the high- and the low-skill occupations before and after polarization took place.

My results indicate a strong effect of job polarization on wages. First, the wages of middle-skill occupation workers compared to high- as well as low-skill occupation workers declined firmly between the two cohorts of the NLSY. Workers who are very likely to enter the middle-skill occupation in the end of the 1980s suffered even an absolute decline in their average wages over this twenty year period. Second, the wage rate that is paid for a unit of effective labor in the middle-skill occupation decreased substantially compared to the wage rates in both of the other occupations. Third, job polarization seems to have driven the change at the top of the wage distribution while its effect at the bottom of the distribution turns out somewhat inconclusive.

The paper builds its analysis step-by-step, making additional assumptions and using increasingly specialized techniques as they become necessary. I start by analyzing the sorting of workers into occupations in the end of the 1980s using multinomial choice regressions in the NLSY79. I find that, conditional on the other talents, workers with high math talent are likely to choose the high-skill occupation, workers with high mechanical talent are likely to choose the middle-skill occupation, and workers with high verbal talent are likely to choose the high- or the low-skill occupation but not the middle. This indicates a systematic sorting of workers into occupations according to their relative talent endowments and it enables me to construct predicted probabilities for every worker in both datasets to enter the high-, middle-, and low-skill occupation in the end of the 1980s.

I use these predicted probabilities to study how the wages of workers who were likely to choose to work in the high-, middle-, or low-skill occupations in the end of the 1980s have changed over time. Effectively, this amounts to a wage regression on the occupational propensities interacted with a dummy for the period at the end of
the 2000s. Under the assumption that the distribution of fundamental worker skills conditional on the detailed talents has been constant before and after polarization took place, these regressions identify the changing returns over time to working in the high- and low-skill occupation versus the middle-skill occupation in the end of the 1980s. Moreover, the empirical strategy deals with the fact that the composition of skills in each occupation changes under job polarization. The latter renders the naive approach of a raw comparison of occupational wages meaningless, because it confounds changing wage rates with a changing selection bias in occupations over time.

The results from the regressions on occupational propensities show that the wages of middle-skill occupation workers declined substantially over the polarization period. I find that the positive wage effect associated with a one percentage point higher propensity to work in the high- compared to the middle-skill occupation almost doubled from .31 to .60 percent. The negative wage effect associated with a one percentage point higher propensity to work in the low-skill occupation attenuated from -1.65 to -.95 percent. Moreover, workers with a high propensity to enter the middle-skill occupations in the 1980s suffered even an absolute decline in their expected real wages. These findings are robust to controlling for absolute skill measures, such as educational attainment, which supports the idea that it is relative occupational skills rather than absolute skills whose returns have changed over time.

Next, I set up a Roy (1951) model of occupational choice in order to interpret the results so far within a more explicit economic framework. In the model every worker possesses a vector of talents which combine into skills in each occupation and which are partly observed in the data. The wage offered to a worker in a given occupation is then the product of this skill and an occupation-specific skill price (or wage rate). Given that workers choose their jobs according to where they earn the highest wage, the occupational choice regressions on observable talents yield predictors about workers’ relative skills in the different occupations. Moreover, job polarization can

3 I provide supporting evidence for the identification assumption by showing that the level and cross-correlation of talents have not changed between the NLSY79 and the NLSY97, and that the distribution of the predicted propensities, which approximate relative occupational skills, has been constant.

4 I use the regression equation formulation of workers’ skills and wages from Heckman and Seldacek (1985) as a specific example throughout the analysis.
be introduced into this model of labor supply as a decline in the relative prices that are paid per unit of skill in the middle- compared to the high- and the low- skill occupations. In this case, employment in the middle will fall and the relative wages of workers who are likely to enter the middle-skill occupations will decrease. This is the above empirical finding and the answer to the first question of the paper.

More importantly than interpreting the reduced-form regression results, I use the model to estimate the underlying changes in occupation-specific skill prices. I do this by deriving the following result: for every talent, the change over time in its return depends solely on the changing relative skill prices, how it is associated with occupational choice, and how the association changes over time. Because I know the changing returns to talents and the occupational choices they predict, the result yields empirical moment conditions which I can use to estimate the changing skill prices. Moreover, since the NLSY provides more talents—three test scores plus the risky behaviors and other demographics—than the two unknown relative prices, I also directly obtain an over-identifying restrictions test of the model itself.

Implementing this procedure, I find that the over-identifying restrictions test does not reject the polarization model in the data. Moreover, the relative wage rate in the high- compared to the middle-skill occupation increases by 20.1 percent since the end of the 1980s and the relative wage rate in the low-skill occupation increases by 31.4 percent. The former rate is estimated relatively precisely while the latter is imprecise and not statistically significant. Nonetheless, overall these results paint a picture of substantially declining relative wage rates for a constant quality of work in the middle-skill occupation over time. This answers the second question of the paper.

Next, I generate a counterfactual wage distribution that is due to polarization’s effect on the occupation-specific skill prices. To do this, I assign the estimated relative wage rate changes to each worker in the NLSY79 according to his occupation. It turns out that the counterfactual distribution more or less matches the increase of wages at the top of the actual distribution compared to the middle. In contrast, despite a slight increase, it does not reproduce much of the surge in wages at the bottom of the actual distribution compared to the middle. Moreover, the increase of average wages

---

5 I need to make another linear approximation assumption. This is explained in detail in section along with the estimation procedure.
in the low-skill occupation is substantially higher in the counterfactual than in the actual wage distribution.

However, this does not yet imply that job polarization in general cannot generate the u-curved change in the actual wage distribution. The reason is that the counterfactual so far leaves out the potential effect on wages and the wage distribution of workers reallocating out of the declining middle-skill occupation. Without additional assumptions, there are no clear predictions from the Roy model about this effect. As a last exercise, I therefore calibrate occupational switches and wage effects in response to polarization in order to see whether they can make the counterfactual wage distribution match the actual distribution in the bottom as well. It turns out that under relatively strong assumptions, this can in principle be achieved.

Overall, therefore the answer to the third question of the paper remains partly inconclusive. On the one hand, there is a strong impact of polarization on inequality at the top of the wage distribution which matches its actual change quite well. On the other hand, the increase in wages in the bottom of the actual wage distribution can only be reproduced under assumptions which are relatively strong and not directly testable in the data. In fact, given these results, it seems possible that part of the surge at the bottom of the actual distribution is driven by other factors than job polarization—for example policy-related and institutional factors such as the minimum wage or de-unionization (e.g. Machin and Van Reenen 2008, Autor, Manning, and Smith 2010, Firpo, Fortin, and Lemieux 2011).\footnote{This notion is supported by evidence that job-polarization does not necessarily go together with a u-curved change of the wage distribution in other countries (e.g. Machin and Van Reenen 2008, Dustmann, Ludsteck, and Schönberg 2009).}

As far as I am aware, there exist only a couple of recent papers that empirically analyze the wage effects of job polarization—asking complementary questions and employing complementary methodology to those of this paper. First, Firpo, Fortin, and Lemieux (2011) use a decomposition method to assess the contribution of different factors to the change in the wage distribution over the last three decades. Their results indicate an important role for technology and de-unionization in the 1980s and the 1990s, and for offshoring from the 1990s onward. Second, Cortes (2012) uses panel data to study the wages over time of workers who stay in the middle-skill occupation compared to workers who stay in the high- and the low-skill occupation and compared
to workers who transition into these occupations. He finds that the formers’ relative wages declined compared to the latter two groups. Third, Autor and Dorn (2013) use local labor market data in the U.S. to, among other questions, study the changes in the wage distribution of differentially routine-specialized local labor markets over time. They find that more routine-specialized labor markets experienced a more u-curved change in wage inequality.

In addition to these studies, my paper is closely linked to the larger literature on job polarization and in particular the studies that make explicit predictions about polarization’s effect on wages. For example, a recent survey article by Acemoglu and Autor (2011) provides a Ricardian assignment model of skills to tasks where polarization amounts to the substitution of machines for certain tasks previously performed by labor. The model predicts that those workers who have a comparative advantage in tasks for which the relative market price decreases will be displaced and find their relative wages decline. Moreover, the replacement of tasks in the middle of the skill distribution at the same time generates job polarization and a u-curved change of inequality within their framework. These are exactly the predictions that I examine in my paper. Other models that generate similar predictions include Autor, Levy, and Murnane (2003), Autor, Katz, and Kearney (2006), Cortes (2012), and Autor and Dorn (2013). Feng and Graetz (2013) generate job and wage polarization in a model where firms’ automation choices are endogenous. They argue that these predictions result from automation in general and are thus not a unique consequence of the ICT revolution.

However, in one aspect the Roy framework in my paper is more general than the existent models in the literature on polarization. These models either feature a one-dimensional distribution of worker skills with comparative advantage changing along this dimension (Acemoglu and Autor 2011, Cortes 2013) or restrict the population distribution of skill in the low-skill occupation to be homogenous (Autor, Katz, and Kearney 2006, Autor and Dorn 2013). In contrast, my model allows an unrestricted

---

7In addition, there are a couple of studies that use individual-level panel data to examine the wage impact of the exposure to international trade (e.g. Ebenstein, Harrison, McMillan, and Phillips 2009, Autor, Dorn, Hanson, and Song 2013).

8The exception is Autor, Levy, and Murnane (2003) who have two unrestricted dimensions of skills, but they assume them right away and thus have no notion how talents map into them. Firpo, Fortin, and Lemieux (2011) also have a Roy model as a conceptual framework, but they do not use
mapping from workers' fundamental talent endowments into different skills for each occupation. This is conceptually appealing because it lets multidimensional talents be priced differently across occupations but it is also empirically relevant as workers sort into occupations according to their relative talent endowments in the data. Moreover, there is substantial overlap between the wage distributions across occupations, which smoothes the impact of changing occupational skill prices on the overall wage distribution and which the restricted models do not allow for. Finally, changing occupational skill prices lead to individuals moving up or down in the wage distribution, which affects the impact of job polarization on different quantiles. Again, the restricted models do not allow for this.

The paper continues as follows. Section 2 demonstrates that job polarization and wage inequality in the NLSY are similar to what is found in the commonly used Current Population Survey (CPS). Section 3 shows that workers sort themselves systematically into occupations according to the talent measures available in the NLSY. Section 4 presents the empirical results on the returns to occupation-specific skills. These results are interpreted within the Roy model in section 5. Section 6 estimates the change in occupation-specific skill prices in the model and tests its over-identifying restrictions. Using the price estimates, section 7 assesses the impact of job polarization on the overall wage distribution. The last section concludes.

2 Data and Empirical Facts

This section establishes the stylized facts of job polarization and the u-curved change in wage inequality in my data. Median real wages for 27 year old males rise only marginally, which is also in line with evidence on prime age workers presented in other papers.

I use data from the National Longitudinal Survey of Youth (NLSY) cohorts of 1979

---

9My paper is also related to, and inspired by, studies that use the Roy model in relation to an earlier literature on skill-biased technological change (e.g. Gould 2002, Mulligan and Rubinstein 2008, Yamaguchi 2012).

10Moreover, after Altonji, Bharadwaj, and Lange (2012), this is the first study to analyze labor market outcomes across the two cohorts of the NLSY. While Altonji, Bharadwaj, and Lange examine the effect of changes in overall skill supply on wage levels and inequality in the economy, my paper analyzes the effect of shifts in skill demand across occupations.
and 1997, which contain detailed information on individuals’ fundamental talents that is not available in other datasets. Moreover, the two cohorts are specifically designed to be comparable to one another. When possible, I compare my results to the more standard Current Population Survey Merged Outgoing Rotation Groups (CPS) over the same period.

The individuals in the NLSY surveys are born between 1956 and 1964 and between 1980 and 1984, respectively. I restrict my attention to 27 year olds, which is the oldest age that I have enough data in the NLSY97 for to analyze, and to males. The sample selection and attrition weighting is done closely in line with a recent paper using both of the NLSY cohorts by Altonji, Bharadwaj, and Lange (2012). Labor supply by hours worked and real hourly wages are defined as in Lemieux (2006). The details of the sample construction can be found in section A of the appendix intended for the web (henceforth web appendix A). Table 1 accounts for how I end up with a sample of 3,054 and 1,207 individuals in the NLSY79 and the NLSY97, respectively.

For the overall (male) labor force, the wage distribution change from the end of the 1980s to the end of the 2000s is characterized by a u-curve, i.e. wages increase substantially at the top of the distribution and somewhat less at the bottom but hardly at all in the middle. Moreover, there is job polarization in the sense that employment in the middle-skill occupations decreases and employment in the high-skill and low-skill occupations increases. For the details of these facts, see the survey paper by Acemoglu and Autor (2011).

I start with the stylized fact about the wage distribution in my data. Figure 1 graphs the change in log real wages by distribution quantile in the NLSY79 and the NLSY97 and for the comparable years and age group in the CPS. We see that the changes in the NLSY and the CPS align well for both cohorts. This establishes the well-known U-shape in the wage distribution for the NLSY.

\[11\] At the time of the analysis, NLSY97 data was available up to 2009. The periods that I compare are thus 1983-1991 and 2007-2009. One may be concerned that the latter is a crisis period and therefore peculiar. However, Jaimovich and Siu (2012) show that job polarization is closely linked to the business cycle in general. Moreover, the 2007 data for the NLSY and the CPS were collected well before the crisis affected the labor market while the facts for that year alone are qualitatively similar to the whole of 2007-2009.

\[12\] This is sometimes referred to as the squeeze of the middle class.

\[13\] The increase at the top for 27 year olds is not as pronounced as previous papers have found for prime age males (e.g. Acemoglu and Autor 2011). This is not surprising, since the wage trajectory for high-skilled workers is steep around the age of 27 and thus the differences, and their changes,
The second important fact is job polarization. Early papers on this topic by Goos and Manning (2007) and Autor, Katz, and Kearney (2006) combined occupations into three broad skill groups according to their initial wages or education levels and found that relative employment in the middle-skill group dropped substantially during the last decades while employment in the high- and the low-skill group increased. Recent studies on job polarization, however, preferred delineating occupations more explicitly along the lines of their routine and non-routine task content (e.g. Acemoglu and Autor 2011, Cortes 2012, Jaimovich and Siu 2012). Specifically, in these papers,

- managerial, professional, and technical occupations are grouped as high-skill (or non-routine cognitive);
- sales, office and administrative, production, and operator and laborer occupations as middle-skill (or routine);
- and protective, food, cleaning and personal service occupations as low-skill (or non-routine manual).

This delineation has the advantage of being conceptually appealing in terms of Autor, Levy, and Murnane (2003)’s well-known routinization model. At the same time it preserves the ranking into high-, middle-, and low-skill occupations in terms of average wages and education levels, and it generates the same polarization facts as the early papers. I follow this emerging standard in the literature, while my results below are qualitatively similar if I use either of the alternative groupings according to occupational wages or education levels.

Figure 2 graphs the percentage point change of employment in the high-, middle-, and low-skill occupations for the NLSY79 and NLSY97 and compared to the CPS. We can see that the employment share of middle-skill occupations is declining substantially while the employment share of the high- and the low-skill occupations is rising. These facts establish job polarization in the NLSY, which is close to what can be found for 27 year olds in the CPS.

Before moving on, figure 3 plots the change in average real wages in the high-, middle-, and low-skill occupations in the NLSY and, for comparison again, the are likely to be larger at older ages.
CPS. While wages in high-skill occupations have increased robustly in levels and compared to the other two occupations, wages in low-skill occupations have lost somewhat further ground against wages in middle-skill occupations in the NLSY and also slightly in the CPS. One might find this surprising under the demand side explanation for job polarization, which should decrease employment and wages in the middle at the same time. Yet, just as the size of occupations, the composition of skills in occupations does not stay constant when relative demands change. Appropriately accounting for this effect is the main contribution of my paper. In order to do that, I study which kind of workers sort into the high-, middle-, and low-skill occupations in the next section.

3 Talent Sorting into Occupations

Workers do not choose their occupations at random. This section uses choice regressions to establish and quantify systematic occupational sorting depending on workers’ talent endowments.

3.1 Measures of Talent

The NLSY data provides a long array of characteristics of its respondents. Out of these, I focus on variables that are early determined, that are relevant for occupational choices and wages, that should approximate different dimensions of skill, and that can be compared over the two cohorts.

Table 2 reports labor force averages of NLSY variables that fulfill the four criteria.

\footnote{Note that the small differences between wages, occupational employment, and occupational wages in the NLSY and the CPS are unlikely to stem from systematic sample attrition or non-test-taking in the NLSY. This is because sample attrition or non-test-taking are much lower in the NLSY97 than the NLSY97, while the differences between CPS and NLSY are equally large for the two cohorts. Further, note again that the scope of the NLSY and the CPS are different. The CPS is supposed to be representative of the resident population in the survey year while the NLSY is supposed to be representative of those individuals in the survey year who were between 14 and 21 years old in 1979 and between 12 and 16 in 1997, respectively.}

\footnote{Also other studies find a further decrease in low-skill compared to middle-skill wages (Goos and Manning 2007). Autor and Dorn (2013) find that relative wages in clerical occupations rise while quantities fall.}

\footnote{Thus, the popular non-cognitive skill measures of locus of control and self-esteem have to be left out of the analysis because they are not available in the NLSY97.}
(“early skill determinants”) and some demographic variables and contemporary skill determinants that are available in more standard datasets. In terms of the early skill determinants, I construct intuitive composite measures of mathematical, verbal, and mechanical talent by combining test scores on mathematics knowledge, paragraph comprehension and word knowledge, and mechanical comprehension and auto-and shop information, respectively. In addition, I report the AFQT score, which is commonly taken as a measure of general intelligence.\footnote{All these measures are taken from the Armed Services Vocational Aptitude Battery of tests (ASVAB) which consists of ten components: arithmetic reasoning, word knowledge, paragraph comprehension, mathematics knowledge, general science, numerical operations, coding speed, auto and shop information, mechanical comprehension, and electronics information. The breakup into mathematical, verbal, and mechanical talent is very similar to what a factor analysis of test scores suggests. AFQT is essentially the average of arithmetic reasoning, word knowledge, paragraph comprehension, and mathematics knowledge.}

The advantages of the early skill determinants—and in particular the composite measures of mathematical, verbal, and mechanical talent—compared to the contemporary skill determinants—and in particular measures of education—for my study are threefold: First, the early skill determinants are largely exogenous to an individual’s actual occupational choice as they are hardly malleable and determined before entry into the labor market. This is in particular the case for the NLSY97 cohort who are tested at ages 12 to 16 and who are choosing different jobs due to polarization. Second, the test scores are finer measures of individual differences in skill than education, which has a lot of bunching at points like high school graduate (12 years of education) or college graduate (16 years of education). This is a sizeable advantage when I want to use test scores to compare similarly skilled individuals over the two cohorts. And finally, the test scores provide proxies for multiple dimensions of individuals’ skills. Thus, they can be used to determine comparative advantage as I show in the next subsection.

Before moving on, we see from table\textsuperscript{2} that the level of AFQT, which is a measure of IQ, does not change in the male labor force over the two cohorts. In addition, table\textsuperscript{3} reports that the cross-correlation of the composite test scores and AFQT remained virtually the same. This supports my identification assumption in the following that the tests measure similar dimensions of talent over the two cohorts and that “within test score groups” individuals can be considered on average the same across cohorts.
Hispanics in the population and the labor force increased by about 8 percentage points while the share of Whites decreased by a similar extent and the share of Blacks stayed roughly constant. Therefore, I control for race in all my analyses in the following.

3.2 Sorting into Occupations

Figure 4 depicts average mathematical, verbal, and mechanical talent in the three occupation groups in both cohorts. We see that the levels of all three talents are much higher in the high-skill occupation than in the middle-skill occupation which, in turn, is higher than the low-skill occupation. Thus, there is a clear ordering of absolute advantage in occupations independent of the talent considered. This underlines the appropriates of the classification of high-, middle-, and low-skill occupations.

Yet, in the absence of restrictions to enter occupations, workers’ choice should not be governed by their absolute but by their comparative advantage and thus depend on their relative skills (for details, compare Sattinger 1993). We see in figure 4 that average mathematical talent in the high-skill occupation is higher than average verbal or mechanical talent, while average mechanical talent is considerably higher in the middle-skill occupation than mathematical or verbal talent. Verbal talent is higher than mathematical and mechanical talent in the low-skill occupation.

This strongly suggests sorting according to comparative advantage and multidimensional skills as in the well-known Roy model—with workers who have high math talent choosing the high-skill occupation, workers who have relatively high mechanical talent choosing the middle-skill occupation, and workers who have relatively high verbal talent choosing the low-skill occupation—and it is opposed to the one-dimensional skill notion that is often adopted in this literature. It is also intuitive, since high analytical skills are required to pursue a career in managerial, professional, or technical jobs while individuals who have relatively strong mechanical skills or a practical inclination may prefer to work in production or clerical jobs. Verbal skills may be relatively helpful to communicate in personal and protective service occupations. In this case, the uniform absolute ranking of occupations in the three talents should stem from the high cross-correlations between them as seen in table 3.

To test the idea of sorting according to comparative advantage I run multinomial
choice regressions. Let \( \{K_i\} \) be a set of indicator variables that take the value of 1 when individual \( i \) works in occupation \( K \in \{L, M, H\} \) and zero otherwise. The timing is such that \( t = 0 \) when the members of the NLSY79 are 27 years old and \( t = 1 \) when the members of the NLSY97 are 27 years old. I model the conditional choice probabilities as multinomial logit (MNL):

\[
p_K(x_{it}, t) = \frac{\exp(b_{K0t} + b_{K1tx_{1it}} + \ldots + b_{KJtx_{Jit}})}{\sum_{G=H,M,L} \exp(b_{G0t} + b_{G1tx_{1it}} + \ldots + b_{GJtx_{Jit}})},
\]

(1)

where \( p_K(x_{it}, t) \) denotes the probability in point in time \( t \) to enter occupation \( K \) for an individual of talent vector \( x_{it} \), and \( x_{jit} \) represents an element of that talent vector.

Maximum likelihood estimation of equation (1) yields the coefficients of this model and it provides conditional probabilities (“propensities”) to enter each occupation based on the observable talents. As I explain in more detail in section 5, these propensities can be interpreted as individuals’ predicted relative skills in an occupation as opposed to the other two occupations. However, note that the descriptive choice regressions do not in general identify any parameters of the economic model that I introduce then.

Table 5 reports the results from the multinomial choice regressions. These extract the marginal effect of an additional unit of each talent on occupational choice when the respective other talents are held constant. For ease of discussion, focus on the first column which gives the sorting into high- and low-skill occupations relative to the omitted middle-skill occupation in the NLSY79. Conditional on the other talents, a one unit higher math score is associated with an about 4.7 percent higher probability to enter the high-skill versus the middle- or the low-skill occupation. A one unit higher mechanical score is associated with a 1.4 and 2.3 percent lower probability to enter the high- and the low-skill occupation as opposed to the middle-skill occupation, respectively. In contrast, a one unit higher verbal score decreases the probability to enter the middle- as opposed to the high- or the low-skill occupation by about two percent. Thus, the idea of sorting according to comparative advantage is strongly supported by these regressions—with workers who have (conditionally) high math skills moving into the high-skill occupation, workers with conditionally high mechanical and low verbal skills moving into the middle, and those workers
with low math and mechanical skills moving into the low-skill occupation. Also, the results underscore the importance of measuring multiple dimensions of skill for linking occupational demand to workers’ comparative advantage in my data. They are the same when looking at the NLSY97 in figure 4 and in column three of table 5.

Finally, the regressions in columns two and four of table 5 are run for creating the propensities to enter occupations based on observables that are used in the following. The test scores are split into terciles in order to also allow for a u-curve change in demand for skill levels as suggested by the one-dimensional skill models. Moreover, normalized measures of illicit activities and engagement in precocious sex are added. The regressions omit parental education because it is not available for about a third of respondents. However, the results below are qualitatively robust to adding parental education, omitting the risky behavior measures, or using the regressions in columns one and three for creating propensities.

Table 4 reports the predicted values from the NLSY79 sorting regression in column two of table 5. We see that the distribution of occupational propensities is remarkably stable over the two cohorts. Hence, the joint distribution of observable talents which are relevant for occupational choice is virtually unchanged. This is in line with the constant correlations across test scores in table 3. It also makes it more plausible that the (conditional) distribution of the relevant unobservables has been stable and thus further supports the identification assumption in the next section.

4 Polarization’s Relative Wage Effect

The prevalent view is that job polarization is a demand-side phenomenon. Thus, the wages of middle-skill workers should fall over time compared to the wages of high- and low-skill workers. I exploit the sorting results of the last section to study the wages of workers who have different probabilities to enter the high-, middle-, and low-skill occupations.
4.1 Empirical Analysis

In order to do this I estimate ordinary least squares (OLS) regressions for pooled data of the form

\[ w_{it} = \alpha_0 + \alpha_1 p_H(x_{it}, 0) + \alpha_2 p_L(x_{it}, 0) + \alpha_3 \times NLSY97 + \alpha_4 p_H(x_{it}, 0) \times NLSY97 + \alpha_5 p_L(x_{it}, 0) \times NLSY97 + \varepsilon_{it}, \]

where NLSY97 is a dummy for whether a particular observation is from the NLSY97 (in fact that \( t = 1 \)), and \( p_H(x_{it}, 0) \) and \( p_L(x_{it}, 0) \) are the probabilities to choose the high- and the low-skill occupation in the NLSY79, i.e. before polarization took place. Hence, the approach is to study the change in average wages for types of workers that have different propensities to work in high-, middle-, and low-skill occupations. The parameters of interest are the changing relative returns to a higher probability in the NLSY79 of working in the high- and the low-skill occupation compared to the middle skill occupation \( \alpha_4 \) and \( \alpha_5 \).

Of course, the occupational choice probabilities are not directly available in the data and they have to be estimated in a preceding step in the NLSY79 along the lines of the previous section. The parameter estimates are then used to predict \( p_H(x_{it}, 0) \) and \( p_L(x_{it}, 0) \) for each individual in the NLSY79 and the NLSY97. This makes the estimation of (2) a two-step procedure. In fact, I am using two-step estimation procedures throughout this paper since my empirical strategy exploits measuring relative skills in occupations with respect to observable talents and then relating these relative skills to changes in the returns to talents.

In terms of the two-step procedure used here, two clarifications are in order. First, different functional form assumptions can be used to specify \( p_K(x_{it}, 0) \). A linear probability model, i.e. OLS regression, provides the best linear estimator for the probabilities but some predicted values from it will be above one or below zero, that is, they are not probabilities themselves. Therefore, many researchers would prefer a multinomial logit or probit model. I report the results from the multinomial logit that I ran in table 5 but my results do not change if I use the other options to specify \( p_K(x_{it}, 0) \).

Second, the standard errors in the second stage regression (2) have to reflect the
fact that $p_H(x_i, 0)$ and $p_L(x_i, 0)$ are estimates and thus possess sampling variation. Among others, Murphy and Topel (1985) provide a procedure to do this, which is however somewhat tedious.\[18\] Therefore I compute bootstrapped standard errors instead, which are also asymptotically consistent.

Table 6 reports the results from wage regression (2). Unsurprisingly, in column one we see that a higher propensity to enter the high-skill occupation compared to the omitted middle-skill occupation is associated with a significantly higher wage. The reverse is true for the propensity to enter the low-skill occupation.

Polarization should however change the returns to propensities over time, which are indicated in the table by “x NLSY97”. We see that the coefficients change strongly and significantly in the expected direction. For the propensity to enter the high-skill occupation, the coefficient almost doubles (from .31 to .60) while the coefficient for entering the low-skill occupation rises by almost a third (from -.65 to -.95). The level of the change in the low-skill coefficient is twice that of the high-skill coefficient, which may come as a surprise. However, note that it is also much less precisely estimated. Moreover, when scaling the size of the estimates by the respective standard deviations of the propensities reported in table 4 the change in the effect of the propensity to enter the high-skill occupation is larger: a one standard deviation increase in the high- and low-skill propensities, respectively, is associated with a 11.3 percent higher and 5.2 percent lower wage in the NLSY97 compared to a 5.9 percent higher and 8.4 percent lower wage in the NLSY79.\[19\]

For illustration of the effect of different propensities to enter the three occupations, figure 5 plots the predictions from linear wage regressions on each propensity at a time together with their probability densities.\[20\] In the top left sub-figure we see the positive effect of having a higher propensity to enter the high-skill occupation in the NLSY79 indicated by the upward-sloping line. This effect increases further in the

\[18\] Two stage least squares or joint estimation (in ML or GMM) of step one and two in a standard statistical package would be a convenient option to get the correct standard errors automatically. However, this is not feasible here as for the individuals in the NLSY97 the regressors are estimated in a different dataset.

\[19\] For the NLSY79 multiply the coefficients on the propensities to enter the high- and low-skill occupations of 0.31% and -1.65% by the standard deviations of these propensities of 19.1 and 5.1. For the NLSY97 multiply the coefficients on the propensities to enter the high- and low-skill occupations of 0.60% and -0.95% by the standard deviations of these propensities of 18.8 and 5.5.

\[20\] The coefficients and standard errors from these wage regressions on each propensity separately are not reported in a table for saving space.
NLSY97 as the dashed line is even steeper. In the top right sub-figure, we see that there is a strong negative effect of the propensity to enter the low-skill occupation, which is however less severe in the NLSY97. Moreover, we see again that the range of propensities to enter the low-skill occupation is limited in the data. Finally, for the propensity to enter the middle-skill occupation there is already a negative effect in the NLSY79 but this becomes substantially more negative in the NLSY97. For individuals with a very high propensity to enter the middle, which is quite frequent in the data, expected real wages even decline during the two decades between the NLSY79 and the NLSY97. This is indicated by the crossing of the two lines.

The identification of changes in returns to propensities in regression (3) is based on the assumption that for a given vector of observed talents \( x \) workers are in expectation the same in terms of their unobserved occupational productivities over the two cohorts. Tables 2 and 3 provided support for this assumption as they showed that the level and cross-correlation of observable early skill determinants is very similar in the NLSY79 and NLSY97. In addition, table 4 showed that the distribution of predicted propensities is very similar in the NLSY79 and NLSY97, i.e. that the distribution of relative occupational skills according to the observable measures has not changed over the two cohorts. Combined, these pieces of evidence lend substantial support to my identification assumption. Given this identification assumption, the changes of the propensity coefficients provide the increase in average wages that is associated with relative advantage in the high- or the low-skill occupation compared to the middle.

The result in column one of table 6 does not exclude the possible influence of other factors than polarization on wages of workers with relative advantage in the high- or the low-skill occupation. In particular, skill-biased technological change (SBTC) that is independent of occupational demand constitutes an alternative hypothesis to polarization and may thus have an important effect on talent returns. According to this view, relative advantage in high-, middle-, and low-skill occupations is not important because returns to skills rise across the board. When allowing for SBTC in regression (2) with all the talents included on top of polarization, the identification

\[ \text{21}\text{The racial distribution does however change over the cohorts. I control for race in all my analyses.} \]
will have to rely on the functional form of $p_H(x_{it},0)$ and $p_L(x_{it},0)$, because the same variables that are used for estimating the propensities are directly entered into the wage regression. This may potentially lead to near multicollinearity of the explanatory variables in the regression and imprecise estimates. In additional regressions, I thus use education indicators as absolute skill measures.

The remaining columns of table assess the potential importance of the SBTC hypothesis versus polarization. Column two adds to the regression a dummy of whether the individual completed a four-year college or more. On the one hand, we see that the level of the coefficient on the propensity to enter the high-skill occupation drops all the way to zero but that the changes in both coefficients are remarkably stable. On the other hand, the level of return to college is large and highly significant while its change does not significantly increase once I control for the propensities. The result is similar if I control for four different degree dummies (high school dropout and graduate, some college, and at least four year college) in column three. This indicates that Mincerian returns to education are important to explain wages in the cross-section, but that they have much less power than relative skills in occupations to explain the change in wages that took place over the twenty years from the NLSY79 to the NLSY97.

Finally, the regression reported in column four of the table adds the same specification of talents that I use to estimate the occupational propensities in the first place. The parameter estimates on the propensities remain in the right direction and become even stronger but they also become imprecise and insignificant, which is due to the high degree of multicollinearity between the regressors in this specification. Therefore, the regression is not as informative as the preceding ones.

4.2 Evaluation of the Results

The above results show that the relative returns to propensities of working in high-as well as low-skill occupations have risen strongly between the two NLSY cohorts. Individuals with a very high probability to work in the middle-skill occupations in the 1980s have even seen a an absolute decline in their average real wages. Moreover,

\footnote{The coefficient estimates on the degree dummies and the talents included in column three and four of the table do not provide additional insight and are not reported in order to save space.}
these returns changes seem to be driven by relative occupation-specific skills rather than an across-the-board increase in the returns to skill as suggested by an alternative skill-biased technological change hypothesis. Taken together, this is a statistically and economically strong result on the long-term wage effect of job polarization which is fully in line with the prevalent demand-side explanation of polarization and it answers the first question of the paper.

At this point, the natural next question to ask is: how much of the change in the overall wage distribution can the changing returns to propensities explain? Figure 6 provides the answer. It plots the actual and the predicted change in the wage distribution when the changing coefficient values from the regressions reported in columns one and four of table 6 are assigned to workers’ wages in the NLSY79. As we can see, the propensities to enter occupations with their functional form restriction do not do a worse job in matching the wage distribution than a very flexible specification of the same talents that are included in estimating the propensities. However, both options do not explain much of the change in the wage distribution.

The reason is the following: while there is considerable and detailed variation in the predicted probabilities of entering the three occupations exploited in the wage regressions (e.g. compare table 4 or figure 5), these propensities based on observable talents cannot explain a lot of the variation in overall occupational sorting (e.g. compare the pseudo R-squared in table 5). The observables also only explain a small share of the variation in wages in the cross-section. Therefore, it is unsurprising that the changing returns to observables, without considering unobservables, cannot explain much of the change in the overall wage distribution.

Hence, in order to assess the effect of job polarization on the overall wage distribution, I need to get hold of the changing returns to unobservables. This is possible only in a more explicit modeling framework which allows me to use the evidence about the observables in order to estimate quantities that are informative about the unobservables. Such a framework also has the advantage of providing a more explicit economic interpretation of the results so far. Moreover, the estimation of the model yields a direct statistical test of the question whether the changing returns to observables are actually driven by polarization as opposed to other factors.
5 Theoretical Framework

This section develops a Roy (1951) model of occupational choice in order to interpret
the empirical results so far within an explicit economic framework. Moreover, esti-
mation of key parameters of the model in the next section will allow me to assess the
impact of job polarization on the overall wage distribution.

5.1 General Setup

Let each worker \( i \) choose the occupation that offers him the highest wage:

\[
W_{it} = \max\{W_{Hit}, W_{Mit}, W_{Lit}\},
\]

where \( \{H, M, L\} \) again index the high-, middle-, and low-skill occupation, respec-
tively. These wages are composed of the product of \( i \)’s skill to carry out work in
occupation \( K \in \{H, M, L\} \) \( (S_{Kit}) \) and the market price that prevails for that work in
point in time \( t \) \( (\Pi_{Kt}) \).

As we have seen in section 3, workers choose systematically different occupations
according to their talents. This suggests that the \( S_{Kit} \)s depend on workers’ talents
in different ways. Thus, the same mixture of talents yields different levels of skill in
different occupations. In addition, two workers who have the same level of skill in one
occupation will not generally have the same level of skill in the other two occupations.
This is different from a one-dimensional model of skill.

An illustrative way to formalize these ideas is Heckman and Sedlacek (1985)’s
linear factor formulation of log wages:

\[
w_{Kit} = \pi_{Kit} + s_{Kit} = \pi_{Kit} + \beta_{K0} + \beta_{K1}x_{1it} + \ldots + \beta_{KJ}x_{Jit} + u_{Kit},
\]

where the small \( s_{Kit} \) and \( \pi_{Kit} \) denote the log of occupation \( K \) specific skill and price,

---

23 The model can be set up more generally with a decision rule according to utilities instead
of wages. Moreover, a richer model could also feature dynamic occupational choice according to
(expected) life-time utility, (occupation-specific) skill acquisition through experience, and costs of
occupational mobility. The predictions about polarization’s effect on occupational choices and wages
of this richer setup would be qualitatively the same as in the static and purely pecuniary model,
while the estimation in my data would only be possible under restrictive assumptions that make it
no more realistic than the static and pecuniary model.
$x_{it} = [x_{1it}, \ldots, x_{jit}, \ldots, x_{Jit}]'$ are the observed talents, the $\beta_{Kj}$s are the corresponding linear projection coefficients, and $u_{Kit}$ is an orthogonal regression error which represents the unobserved component of skill in occupation $K$. Note that this specification for $s_{Kit}$ is just an intuitive example and that all the results in the following hold for a general dependency of occupation-specific skills on talents.

One can now interpret the sorting results of section 3 within this framework. For the sake of brevity, I only give a crude intuition: Suppose that the productivity of the math talent in the high-skill occupation is high in relative and in absolute terms (i.e. the $\beta_{Hj}$ corresponding to math is a large number), the productivity of mechanical (verbal) talent is relatively high (low) in the middle-skill occupation, and that the talents are not particularly productive in the low-skill occupation. Moreover, suppose that the intercept $\beta_{K0}$ in the low-skill occupation is high while it is lower in the middle-skill occupation and lowest in the high-skill occupation. Thus, many workers can do the low-skill occupation decently, while a subset of individuals with relatively high mechanical talent can do the middle-skill occupation well, and only few individuals with high relative and absolute math talent can do the high-skill job well.

Suppose, in addition, that the talents are substantially positively correlated in the population as shown in table 3. Then, we will find the evidence about talent sorting reported in figure 4 and table 5: workers in the high-, middle-, and the low-skill occupations have relatively high math, mechanical, and verbal talents, respectively. In addition, average talents and wages are highest in the high-skill occupation and they are lowest in the low-skill occupation but still there exists substantial dispersion of skills within occupations such that middle-skill (low-skill) occupation workers obtaining higher wages than high-skill (and middle-skill) occupation workers are not uncommon. The latter empirical fact cannot be generated in a one-dimensional skill setup or in a setup with homogenous skills in the low-skill occupation while it is easily explained in the Roy model. Moreover, the Roy model allows for workers moving up or down in the wage distribution if the relative compensation in their occupation increases or decreases, respectively. This is also not allowed for in the other setups and it will turn out to be an important factor in the discussion about the change in the overall wage distribution in section 7.
But how should we exactly think about job polarization within the model? The previous literature as well as my empirical findings strongly suggest that polarization is a demand-side phenomenon. The natural way to model a demand shift for work in occupations, given a more or less constant supply, is that it changes the relative market equilibrium prices for occupation-specific skills.

\[
\Delta(\pi_H - \pi_M) > 0 \text{ and } \Delta(\pi_L - \pi_M) > 0.
\]

(5)

If (the relative) \( \pi_{Mt} \) falls, the wage that every worker could earn in the middle-skill occupation \( w_{Mit} \) will fall. Hence, some of the individuals who previously preferred working in the middle-skill occupation will now switch to either the high- or the low-skill occupation. This immediately generates job polarization as seen in figure

The above are all economically intuitive interpretations on the facts about workers’ occupational choices within and across NLSY cohorts. However, the focus of this paper is on the wage effects of polarization about which this framework can provide important additional empirical predictions. Since the argument in the following is rather involved and the general case requires complex notation, I use a maximally simplified version of the model from now on. The results can be extended to the general case for the empirical analysis as shown in web appendix B.

5.2 A Simplified Model to Study the Wage Effects of Polarization

In order to strip the model of equations (3)-(5) to its essence, assume there are only two occupations, middle \( M \) and nonmiddle \( N \), with \( \Delta(\pi_N - \pi_M) > 0 \) under polarization. Moreover, there is only one observable talent \( x_i \) with mean zero (\( E(x_i) = 0 \)) and variance one (\( Var(x_i) = 1 \)), and \( \beta_{K0} \) is zero. I indicate the difference between \( N \) and \( M \) occupation variables by a tilde, i.e. \( \tilde{\pi}_t \equiv \pi_{Nt} - \pi_{Mt}, \tilde{\beta} \equiv \beta_N - \beta_M, \) and \( \tilde{u}_i \equiv u_{Ni} - u_{Mi}. \) I suppress the index \( t \) for \( x_i \) and \( u_{Ki} \) because the only variables that change in the model are the prices \( \pi_{Ni} \) and \( \pi_{Mi} \) and their functions. Wages in

\[24\] This assumption (or result) is similar in the other models on job polarization (see Autor, Levy, and Murnane 2003, Autor, Katz, and Kearney 2006, Cortes 2012, Acemoglu and Autor 2011, Autor and Dorn 2013).
occupations $K \in \{N, M\}$ become:

\[ w_{Kit} = \pi_{Kt} + s_{Ki} = \pi_{Kt} + \beta_K x_i + u_{Ki}. \] (6)

How do the wages of workers who have a comparative advantage in the middle occupation change over time? Since I do not observe the same individual workers in both points in time (the counter-factual), the prediction from the Roy model will have to be in terms of conditional moments with respect to observable talents. Let $K_{it}$ be an indicator variable that takes the value of 1 when individual $i$ works in occupation $K$ and zero otherwise and consider his expected wage conditional on his observable $x_i$:

\[ E(w_{it}|x_i) = E(w_{Mit}|x_i, N_{it} = 1) + p_N(x_i, \tilde{\pi}_t) \left[ E(w_{Nit}|x_i, N_{it} = 1) - E(w_{Mit}|x_i, N_{it} = 0) \right], \]

where the notation

\[ p_N(x_i, \tilde{\pi}_t) = Pr(\tilde{u}_i > -(\tilde{\pi}_t + \tilde{\beta} x_i)) \]

now emphasizes the fact that the probability to enter occupation $N$ is a function of the differences in price per unit of skill between the two occupations. All of the economics of the Roy model can be found in this equation because the probability $p_N(x_i, \tilde{\pi}_t)$ and the conditional wages $E(w_{Kit}|x_i, K_{it})$ are determined by the worker’s optimal choice given his skills and the prices that he faces. Note that $\tilde{\beta} x_i$ is the expected relative skill given $x_i$ and, for a given $\tilde{\pi}_t$, $p_N(x_i, \tilde{\pi}_t)$ is a monotone function of it. The propensity to enter occupation $N$ for worker $i$ estimated from the data can thus be interpreted as a predictor of his relative skill in occupation $N$.

Under the price change of polarization $\Delta(\pi_N - \pi_M) > 0$, the change in the conditional expected wage from $t = 0$ to $t = 1$ can be approximated as a sum of three components:

\[ \Delta E(w_{it}|x_i) = \Delta \pi_M + p_N(x_i, \tilde{\pi}_0) \Delta(\pi_N - \pi_M) + \]

\[ + \Delta p_N(x_i, \tilde{\pi}_t) \left[ E(w_{Nit}|x_i, N_{it} = 1) - E(w_{Mit}|x_i, N_{it} = 0) \right] + \]

\[ + p_N(x_i, \tilde{\pi}_0) \Delta E(s_{Ni}|x_i, N_{it} = 1) + p_M(x_i, \tilde{\pi}_0) \Delta E(s_{Mi}|x_i, N_{it} = 0) \]

23
The first component is the direct price effect, the second the effect of moving out of occupation $M$ (since workers react optimally to the relative price shifts $\Delta p_N(x_i, \tilde{\pi}_t) \geq 0$), and the third a composition effect of skills within occupations. I call the first component the price or wage rate effect and subsume the second and third components under the name reallocation effect. However, without additional assumptions on the distribution of the unobserved skill vector $u_i$, one cannot make a prediction on the relative size of these two effects for workers with different observable talents $x_i$. For example, we do not know how the selection of skills into middle and nonmiddle occupations, $E(s_{Ki}|x_i, K_{it} = 1)$ with $K_{it} = M_{it}$ and $K_{it} = N_{it}$, respectively, changes under polarization. Thus, as mentioned above, it is not clear in which direction relative occupational wages will move.$^{25}$

For this reason, I take a different approach in my paper by starting out from a clear prediction on relative wages for marginal shifts in the $\pi_{Ki}s$ and then applying it beyond the margin. Consider the change in worker $i$'s wages for a marginal shift in prices:

\[
\begin{cases}
  d\pi_N & \text{if } N_{it} = 1 \\
  d\pi_M & \text{if } N_{it} = 0,
\end{cases}
\]

where $d$ denotes a marginal change. Thus, due to the optimality of workers’ occupational choice and the envelope theorem,$^{26}$ the effect on wages of a marginal change in $\pi_{Ki}s$ is only the direct price effect

\[
dE(w_{it}|x_i) = d\pi_M + p_N(x_i, \tilde{\pi}_t)d(\pi_N - \pi_M). \tag{8}
\]

According to prediction (8), under the polarization hypothesis, workers who are ceteris paribus more likely to enter the nonmiddle occupation are expected to see their relative wages increase. This prediction also holds qualitatively beyond the margin.

$^{25}$Even with a distributional assumption, say normality, $E(w_{Ki}|x_i, K_{it} = 1)$ and its change remain hard to interpret economically as there is no simple expression for the expectation of the maximum of correlated normal random variables. Results on the truncated normal provided for example in Heckman and Sedlacek (1985) apply only to the bivariate case; so for my more general three-occupation case things get very complicated. Hsieh, Hurst, Jones, and Klenow (2013) use an extreme value distribution to solve the problem, but this comes at the cost of the strong assumption that individuals’ skills are uncorrelated across occupations.

$^{26}$A version of the envelope theorem also holds for optimization problems where agents’ choices are discreet (e.g. see Milgrom and Segal 2002).
That is, the expected overall wage gain from polarization rises with the initial probability to work in the nonmiddle occupation. Note that the change in worker $i$’s expected wage is the sum over his marginal expected wage changes along the adjustment path from $\pi_0$ to $\pi_1$. Hence, we can integrate prediction (8) from $t = 0$ to $t = 1$ to obtain:

$$E(w_{i1}|x_i) - E(w_{i0}|x_i) = \Delta \pi_M + \int_{\tilde{\pi}_0}^{\tilde{\pi}_1} p_N(x_i, \tilde{\pi}_t) d\tilde{\pi}_t, \quad (9)$$

where the structure of $p_N(x_i, \tilde{\pi}_t) = Pr(\tilde{u}_i > -(\tilde{\pi}_t + \tilde{\beta} x_i))$ illustrates that on the adjustment path of prices, the ranking of $p_N(x_i, \tilde{\pi}_t)$ with respect to $x_i$ remains unchanged.

Web appendix B shows that the result (9) carries over to the three occupation case analyzed in the empirics. This is the same result as in Acemoglu and Autor (2011) and other papers about the wages of workers who have a comparative advantage in tasks for which the relative market price decreases—obtained from a general model of labor supply with multidimensional skills. Hence, the empirical findings of section 4 on the changing returns to working in the high-, middle-, and low-skill occupations are fully in line with the predictions of the Roy model: the model predicts that individuals who have a higher propensity to work in the nonmiddle occupations experience a higher increase in average wages over time. Moreover, it illustrates that the changing returns to occupational propensities include the direct price effect as well as the reallocation effect of workers moving into the nonmiddle occupations.

Another way of deriving equation (9) is illustrative: Concentrate on a specific worker $i$ first and note again that $\tilde{\pi}_t \equiv \pi_{Nt} - \pi_{Mt}$, $\Delta \tilde{\pi}_t > 0$, and $N_{it}$ is an indicator for working in occupation $N$ such that $w_{it} = w_{Mit} + N_{it}(w_{Nit} - w_{Mit})$. Defining the relative price that makes $i$ indifferent as $\tilde{s}_i = -(s_{Nt} - s_{Mt})$, we get:

$$w_{i1} - w_{i0} = \Delta \pi_M + N_{i1}(w_{Nit} - w_{Mit}) - N_{i0}(w_{Nit} - w_{Mit})$$

$$= \Delta \pi_M + \begin{cases} 
\Delta \pi_N - \Delta \pi_M = \tilde{\pi}_1 - \tilde{\pi}_0 & \text{if } N_{i0} = 1, N_{i1} = 1 \\
\tilde{\pi}_1 + \tilde{s}_i = \tilde{\pi}_1 - \tilde{s}_i & \text{if } N_{i0} = 0, N_{i1} = 1 \\
0 & \text{if } N_{i0} = 0, N_{i1} = 0
\end{cases}$$

$$= \Delta \pi_M + \int_{\tilde{\pi}_0}^{\tilde{\pi}_1} N_{it} d\tilde{\pi}_t.$$
However, the model allows me to go beyond the interpretation of the results so far. Intuitively, in section 4, I have only estimated the changing returns to observable characteristics $E(w_{i1}|x_i) - E(w_{i0}|x_i)$ due to $\Delta \tilde{\pi}_t$. One way to also get at the changing returns to unobservables is to estimate the changing prices directly, since these apply to both observables and unobservables as seen in equations (4) and (6). I show in the next section that result (9) provides a robust and straightforward way to do so.

Before moving on, it is appropriate to discuss in more detail the crucial assumption that polarization solely amounts to changes in occupation-specific skill prices $\pi_{Kt}$ in the model. This has two components. First, the conditional distribution of talents does not change over the two cohorts, i.e. the distribution of $u_{Kit}$ conditional on the vector $x_{it}$ does not depend on $t$. This was already assumed and defended in the previous sections.

Second, only the demand levels—and thus market prices—for occupation-specific work but not the production of work are changing. Hence, while the $\pi_{Kt}$s are affected by polarization, the $\beta_{Kj}$s are not. This assumption is probably not entirely true. For example, Autor, Levy, and Murnane (2003) and Spitz-Oener (2006) show that the task content of occupations has been changing over time.

So far, for the purpose of interpreting the reduced-form estimates, the assumption was still largely innocuous. Most other papers in the literature have made the same or a similar assumption in their stylized models. However, when estimating, this is an important restriction imposed on the data.\footnote{Cortes (2012) applies the same restriction for estimation.}

Therefore, I concede that reducing job polarization to changing demand for work in three broad occupation groups is a potentially strong assumption. Nonetheless, it should capture the main thrust of this phenomenon and it allows me to estimate the core parameters of the model in a straightforward and robust way in the next section.
6 Estimating the Change in Occupation-Specific Prices

In this section I estimate the change in occupation-specific skill prices using a minimum distance estimation technique on the moment conditions implied by the model of section 5. The intuition for this is that the return to a talent should change depending on which occupational choice it predicts and how that changes. The approach also directly provides a test of the model restrictions. I explain the method with the help of my simplified setup. Web appendix C details how this can be extended to the general three-occupation model for the estimation.

6.1 Methodology

The overall change in worker $i$’s expected relative wage is the integral over his marginal expected wage changes along the adjustment path from $\pi_0$ to $\pi_1$ as shown in equation (9), which is reproduced here for convenience:

$$E(w_{i1}|x_i) - E(w_{i0}|x_i) = \Delta \pi_M + \int_{\tilde{\pi}_0}^{\tilde{\pi}_1} p_N(x_i, \tilde{\pi}_t)d\tilde{\pi}_t. \tag{9}$$

In this equation, I want to estimate the distance between $\tilde{\pi}_1$ and $\tilde{\pi}_0$ (i.e. $\Delta \tilde{\pi}$) and possibly $\Delta \pi_M$. I know $E(w_{it}|x_i)$ and $p_N(x_i, \tilde{\pi}_t)$ in points in time $t = 0$ and $t = 1$ in the sense that I can consistently estimate them from my primary data. I do not know, however, $p_N(x_i, \tilde{\pi}_t)$ within the interval $t\epsilon(0,1)$ and I will need to make an assumption on it.

The estimation problem can be illustrated in a graph. In figure 7 I want to back out the distance on the x-axis between $\tilde{\pi}_1$ and $\tilde{\pi}_0$ while I know the starting and the end point (the thick dots $A_1$ and $A_2$) of the function (the arch) over which I need to integrate and the value of the integral (the shaded area). I thus need to make an assumption about the shape of the curve connecting $A_1$ and $A_2$. This curve has to be weakly monotonically increasing, as with higher $\tilde{\pi}_t$ the number of workers in occupation $N$ will increase, but it can be concave as in the picture or convex or both.

The first assumption that comes to mind is to simply impose that it is a horizontal line through the point $A_1$, which implies no reallocation of workers due to the price
change and thus to plug $p_N(x_i, \tilde{\pi}_t) = p_N(x_i, \tilde{\pi}_0)$ into (9). In the figure, the difference between $E(w_{i1}|x_i)$ and $E(w_{i0}|x_i)$ is then assumed to be only the rectangle $a$. This results in the marginal prediction (8) holding exactly for the discrete price change as well and the regressions in section 4 on the occupational propensities identifying it directly. Of course, this is not a good assumption.

A seemingly attractive alternative to this would be to assume that $\tilde{u}_i$ is normally distributed (for simplicity assume $\tilde{\sigma} = 1$), which modifies (9) to

$$E(w_{i1}|x_i) - E(w_{i0}|x_i) = \Delta \pi_M + \int_{\tilde{\pi}_0}^{\tilde{\pi}_1} \Phi(\tilde{\pi}_t + \beta x_i) d\tilde{\pi}_t,$$

where $\Phi(.)$ denotes the distribution function of the standard normal. For this to be helpful, I need to know the structural parameter $\tilde{\beta}$ from the model. I could in principle estimate it from a probit model or a Heckman two stage regression. But then I am estimating the price change by relying on a distributional assumption in (9) and, in order to implement it, estimating the necessary parameter $\tilde{\beta}$ relying on the distributional assumption in the first stage. This appears to be no improvement to outright structurally estimating the Roy model with a normality assumption in both cross-sections and comparing the estimated $\tilde{\pi}_0$ and $\tilde{\pi}_1$.

I therefore instead decide for an approach which makes full use of the empirical evidence in $t = 0$ and $t = 1$. I linearly approximate

$$p_N(x_i, \tilde{\pi}_t) \approx p_N(x_i, \tilde{\pi}_0) + \frac{p_N(x_i, \tilde{\pi}_1) - p_N(x_i, \tilde{\pi}_0)}{\tilde{\pi}_1 - \tilde{\pi}_0}(\tilde{\pi}_t - \tilde{\pi}_0).$$

(10)

In figure 7, this amounts to approximating $p_N(x_i, \tilde{\pi}_t)$ as the y-coordinate for the point

29 A more subtle version of it but essentially the same assumption is to recognize that workers reallocate away from the middle occupation but to impose that the extent of reallocation does not differ across observables $x_i$. In this case, equation (9) becomes

$$E(w_{i1}|x_i) - E(w_{i0}|x_i) = \Delta \pi_M + \text{const} + p_N(x_i, \tilde{\pi}_0)\Delta \tilde{\pi}.$$

Again the regression on $p_N(x_i, \tilde{\pi}_0)$ in section 4 directly identifies the price change. This is also not a good assumption as it does not allow for a differential reallocation effect across worker groups. In figure 7 this means that the arch connecting $A_1$ and $A_2$ is restricted to be the same no matter where we start off on the y-axis (even if we start off high, i.e. close to probability one).

31 In the case of three occupations, this would be multinomial probit with correlated errors or structural estimation of the three-sector Roy model.

32 In this case, without convincing exclusion restrictions or instrumental variables that affect only occupational choices but not wages, the identification of the parameter estimates would solely rely on the potentially incorrect functional form assumption for the skill distribution.
on the line $A_1A_2$ that corresponds to $\pi_t$ and by approximating $E(w_{i1}|x_i) - E(w_{i0}|x_i)$ as the trapezoid $a + b$. If the shape of $p_N(x_i, \tilde{\pi}_t)$ in $\tilde{\pi}_t \in (\tilde{\pi}_0, \tilde{\pi}_1)$ is not too convex or concave, the approximation should be reasonably close. Whether it is sufficiently accurate will be tested below.

Equation (9) now becomes

$$E(w_{i1}|x_i) - E(w_{i0}|x_i) = \Delta \pi_M + \frac{p_N(x_i, \tilde{\pi}_1) + p_N(x_i, \tilde{\pi}_0)}{2} \Delta (\pi_N - \pi_M).$$

This is one equation in two unknowns. However, as it holds for all values of $x_i$, I could simply run a pooled wage regression along the lines of section 4, which indentifies the changing return to $p_N(x_i, \tilde{\pi}_1) + p_N(x_i, \tilde{\pi}_0)$ as our price change estimate.

A more attractive way to obtain $\Delta (\pi_N - \pi_M)$ is to multiply both sides of equation (11) by $x_i$ and taking expectations. By the law of iterated expectations, this results in

$$cov(w_{i1}, x_i) - cov(w_{i0}, x_i) = \frac{cov(N_{i1}, x_i) + cov(N_{i0}, x_i)}{2} \Delta (\pi_N - \pi_M),$$

where $cov(w_{it}, x_i)$ is the coefficient from a linear wage regression of $w_{it}$ on $x_i$ and $cov(N_{it}, x_i)$ the coefficient from a linear allocation regression of occupational dummy $N_{it}$ on $x_i$—remember we assumed the variance of $x_i$ to be one in this simplified setup.

Condition (12) is in fact very intuitive. The return to talent $x_i$ should change by the extent to which it increases the probability to work in the nonmiddle occupations $cov(N_{i0}, x_i)$ and the extent to which this association changes $cov(N_{i1}, x_i) - cov(N_{i0}, x_i)$.

If I had just one talent as in this simple example, I could exactly solve equation (12). Yet, as I have $J$ different talents in my empirical implementation, prediction (12) has to hold for each single one of them so that I get $J$ different moment conditions from the model:

$$m_j(\Delta \tilde{\pi}) = \gamma_{j1} - \gamma_{j0} - \frac{\delta_{Nj0} + \delta_{Nj1}}{2} \Delta (\pi_N - \pi_M) = 0,$$

where $\delta_{Njt} \equiv \frac{cov(N_{it}, x_{jit})}{\text{var}(x_{jit})}$ and $\gamma_{jt} \equiv \frac{cov(w_{it}, x_{jit})}{\text{var}(x_{jit})}$. These parameters can be recovered...
from OLS allocation

\[ N_{it} = \delta_{N0} + \delta_{N1}x_{1it} + \delta_{N2}x_{2it} + \ldots + \delta_{NJ}x_{Jit} + v_{Nit} \]  

(14)

and wage regressions

\[ w_{it} = \gamma_{0} + \gamma_{1}x_{1it} + \gamma_{2}x_{2it} + \ldots + \gamma_{J}x_{Jit} + u_{it} \]  

(15)

Therefore, result (13) provides a simple procedure to assess polarization’s effect on the returns to detailed talents. I have data on individuals’ talents, their choices of entering the occupations, and their wages in the periods before \((t = 0)\) and after \((t = 1)\) polarization took place. First, I run allocation regressions (14) in both periods, which recover the partial correlations of the observed talents and occupational choices \(\delta_{Njt}\). Second, I run two wage regressions (15) for \(t = 0\) and \(t = 1\), which recover the partial correlations of the observed talents and wages \(\gamma_{jt}\) in each period. Then, according to condition (13), the change of a talent’s effect on the wage equals its average effect in the allocation regressions times the change in relative prices.\(^{34}\)

In order to estimate those prices in the second step, I plug the first step estimates \(\hat{\delta}_{Njt}\) and \(\hat{\gamma}_{jt}\) into (13) and stack these moment conditions in a \(J \times 1\) column vector \(m(\Delta \tilde{\pi})\). The minimum distance estimator for \(\Delta \tilde{\pi}\) then minimizes the quadratic form

\[ Q(\Delta \tilde{\pi}) = m(\Delta \tilde{\pi})' W m(\Delta \tilde{\pi}), \]  

(16)

where the weighting matrix \(W\) takes into account the variance-covariance relationship between the first-stage estimates \(\hat{\delta}_{Njt}\) and \(\hat{\gamma}_{jt}\). Thus, for the Optimal Minimum Distance (OMD) estimator \(W = [Var(m(\Delta \tilde{\pi}))]^{-1}\). This can be implemented by a feasible generalized least squares (FGLS) regression of the wage regression coefficients \(\gamma_{j1} - \gamma_{j0}\) onto the allocation regression coefficients \(\frac{\delta_{Nj0} + \delta_{Nj1}}{2}\).

Just as GLS the OMD is asymptotically optimal and it yields a consistent estimate

\(^{33}\)

To be exact, the allocation and wage regressions recover the covariance of \(N_{it}\) and \(w_{it}\) with the residual of regressing \(x_{jit}\) on the other observable talents. This is what I use in the following.

\(^{34}\)Note that the literature on SBTC has also run linear wage regressions on test scores (e.g. Murnane, Willett, and Levy (1995)). The difference here is that the drivers of returns changes are explicitly examined in the allocation regressions and that the results are interpreted within an explicit model of sorting and occupational demand.
of the relative price change. Moreover, the objective function \((16)\) in optimum can be shown to be asymptotically chi-squared distributed with \(J - 1\) degrees of freedom (\(J - 2\) when two relative prices are estimated):

\[
Q(\hat{\Delta \pi}) = m(\hat{\Delta \pi})'[\text{Var}(m(\hat{\Delta \pi}))]^{-1}m(\hat{\Delta \pi}) \sim \chi^2(J - 1)
\]

This provides me with a test statistic for the joint test of the hypothesis that the changing returns to talents are driven solely by polarization and my linear approximation of the reallocation adjustment path.

Before moving to the estimation results, two additional comments are in order. First, Altonji and Segal (1996) and Pischke (1995) present evidence for potential bias of the OMD in small samples and recommend using an Equally Weighted Minimum Distance (EWMD) and a Diagonally Weighted Minimum Distance (DWMD) estimator in addition, respectively. I thus report results for these two estimators as well.

Second, the implied absolute wage change in the middle-skill occupation \(\pi_M\) is not estimated in this procedure. But it can be bounded: under the initial prices, the original worker allocation has to (weakly) dominate the new allocation and vice versa under the new prices. I impose this for average wages.

Overall, the procedure of estimating the relative price changes described here has several advantages over the standard approach of estimating the Roy model under normality. It should give relative price estimates that are close to the actual prices and at the same time robust to different distributions of unobserved talents and to unobserved wage shocks. This is achieved by using an objective function that is chi-squared distributed in optimum, allowing for a joint test of the hypothesis that the changing returns to talents are driven solely by polarization and my linear approximation of the reallocation adjustment path.

35In the case of the EWMD, the weighting matrix is the identity matrix and it can be implemented using a simple OLS regression. The DWMD amounts to weighted least squares with \(W = \text{diag}(\text{Var}(m(\Delta \pi)))^{-1}\).

36Thus, in the three-occupation case of the empirics, \(\Delta \pi_M\) has to be such that

\[
\Delta E(w_{it}) \geq \Delta \pi_M + p_H(\pi_0)\Delta(\pi_H - \pi_M) + p_L(\pi_0)\Delta(\pi_L - \pi_M)
\]

since otherwise it would yield higher wages if workers had stayed in the old allocation and

\[
\Delta E(w_{it}) \leq \Delta \pi_M + p_H(\pi_1)\Delta(\pi_H - \pi_M) + p_L(\pi_1)\Delta(\pi_L - \pi_M)
\]

since otherwise it would have yielded higher average wages if workers had been in the new allocation from the outset. The sample statistics corresponding to \(\Delta E(w_{it}), p_H(\pi_t),\) and \(p_L(\pi_t)\) are the change in average wages and the fraction of workers in the high- and the low-skill occupations, respectively. I take the midpoint between the two bounds as my preferred point estimate for \(\Delta \pi_M\).
the functional form of how these combine into skills in occupations. Moreover, it is
transparent and easy to implement, and it directly provides a test of the model with
its restrictions.

6.2 Empirical Results

Web appendix C shows that the estimation and testing procedure of the previous
subsection carries over to the general three-occupation case: the first stage parameters
are estimated in four reduced form allocation regressions, i.e. for the high- and the
low-skill occupation in the NLSY79 and the NLSY97, and one wage regression in
each point in time. Then, a minimum distance estimation in the second stage yields
the relative prices $\Delta(\pi_H - \pi_M)$ and $\Delta(\pi_L - \pi_M)$.

Table 7 starts by reporting the reduced form allocation and wage regressions
similar to equations (14) and (15). In the first two columns, we see that math
talent is associated with the high-skill occupation, mechanical talent with the middle-
skill occupation, and verbal talent with the high-skill occupation to a lesser degree
than math. The illicit activities are associated with not working in the high-skill
occupation.

This is quite similar to the results from the MNL sorting regressions in table 5.
However, contrary to the MNL, the OLS coefficients for each occupation in table 7
are not interpreted with respect to an omitted base occupation but with respect to
the other two occupations taken together. Moreover, note that the R-squared for the
low-skill occupation allocation regressions is very low, i.e. little of the variation in
low-skill occupational choice is explained by the data. This will affect the precision
of my relative price change estimates for the low-skill occupation below.

The changes in returns to talents are reported in column three of table 7. The
returns to the highest math tercile increase significantly, the returns to mechanical
talents fall, and the returns to illicit activities fall as well. This is largely in line with
prediction (13). Thus, most of the returns changes to talents are in the direction
predicted by the model, apart from verbal talents whose returns decline. Yet, with
exception of the top math tercile and illicit activities, the changes are not statistically
significant by themselves.

Overall, thus, the results from table 7 are neither clearly in favor of nor against
the polarization model. The formal test of the restrictions implied by prediction (13) across all talents may therefore be quite informative. Table 8 reports the results from this test and the implied occupation-specific skill price change for the asymptotically optimal minimum distance estimator and the two alternatives suggested by Altonji and Segal (1996) and Pischke (1995). The EWMD, which amounts to OLS estimation, is also the first step of the feasible GLS procedure to implement the OMD.

In the OMD, the point estimates of $\Delta(\pi_H - \pi_M)$ and $\Delta(\pi_L - \pi_M)$ are of the expected sign and of substantial magnitude: the wage rates in the high- and the low-compared to the middle-skill occupation increase by 20.1 and 31.4 percent, respectively. The implied absolute wage rate in the middle-skill occupation itself decreases slightly by 2.4 percent. The p-value of the hypothesis test is at 10.7 percent and thus the model is narrowly not rejected at conventional significance levels. Furthermore, the estimates for $\Delta(\pi_H - \pi_M)$ are precise and do not change in the two alternative implementations of the minimum distance estimator. In contrast to that, at a standard error of 35.2, $\Delta(\pi_L - \pi_M)$ is imprecisely estimated and it actually drops to negative point estimates in the EWMD and the DWMD.

Under these caveats, the answer to the second question of this paper is therefore that the relative wage rate paid for a constant unit of skill in the middle-skill occupation does indeed seem to fall substantially since the end of the 1980s. In the next section, I use the wage rate estimates from the OMD to evaluate what share of the overall change in the wage distribution is due to them and to answer the third question.

7 Polarization’s Effect on the Wage Distribution

In this last section, I assess the effect that job polarization may have had on the change in overall wage inequality. I start by generating a counterfactual wage distribution that is solely due to the occupation-specific prices estimated in the last section. Then, I check whether the remaining difference with the actual change in the wage distribution may in principle be explained by the wage effects of workers’ reallocation out of the middle-skill occupation.

I conduct these exercises in the NLSY and in the CPS data from section [2]. To
use the CPS is now possible again because assigning the estimated skill prices only requires knowledge of workers’ occupations and not their talents anymore. Thus note that, conditional on having obtained the “correct” price estimates from the NLSY, the idiosyncracies of either dataset should not drive my conclusions in this section.

First, I use the price estimates from the optimal minimum distance estimation reported in table 8 and assign them to every worker in the initial cohort according to his occupation. The resulting counterfactual wage distribution together with the change in the actual wage distribution is displayed in figure 8. We can see that in both datasets the increase of wages at the top of the distribution is reproduced quite well by the estimated price changes alone. However, despite the high OMD estimate of $Δ(π_L - π_M) = 31.4\%$ and a slight increase, the surge in the bottom of the actual wage distribution can neither be matched in the NLSY nor in the CPS. In addition, figure 9 depicts the actual and counterfactual change of average wages in occupations. The counterfactual wage increase in the low-skill occupation is much higher than the actual in both datasets, while the increase in middle-skill occupations is somewhat lower.

Taken together, these pieces of evidence seem puzzling: why does a high relative price change and the resulting relative wage increase in the low-skill occupation not achieve much in terms of lifting the bottom of the counterfactual wage distribution? Figures 10 and 11 provide the answer by plotting the share of low-, middle-, and high-skill workers into the actual initial wage distribution and by depicting the counterfactual when workers’ positions in the distribution are fixed at the original quantile, respectively.

First, in figure 10 we see that in both the NLSY and the CPS the share of low-skill occupation workers monotonically declines with the quantiles of the wage distribution while the share of high-skill occupation workers increases. The share of middle-skill occupation workers features a hump shape—rising up to around the 30th percentile and then slowly declining. Thus, a drop in the wage rate for middle-skill occupations will strongest hit workers who already started out in the lower third of the wage distribution. Moreover, all three curves are relatively flat, which indicates 37The OMD estimates are the most favorable for polarization to generate a large share of the actual change in the wage distribution. More generally, this section constructs the “best-case” scenario of what polarization might be able to explain.
that the dispersion of wages within the three occupation groups is large and that
they are overlapping substantially. Hence, a decline in the wage rate for middle-skill
occupations will drag down the wages of many low earners and many middle-earners
as well. This smooths the impact of the price changes in the lower half of the wage
distribution.

Second, the estimated price changes are large enough such that overtaking of
middle-skill occupation workers by low-skill occupation workers becomes important
in the wage distribution. Consider figure II which depicts a smoothed version of
the counterfactual from figure VIII and an alternative counterfactual where workers’
quantiles in the original distribution of the 1980s are kept constant.\footnote{Note that
this alternative counterfactual is related to plots presented in other studies of wage
changes in occupations against their initial ranking in the 1980s (e.g. Autor and
Dorn 2013).}

In the figure we see that if we force all of the workers’ wage change impacts on
their original quantile, and thus shut down that they move up or down the wage
distribution depending on their occupations, the rise at the bottom of the counter-
factual is stronger while it is weaker at the top. Overtaking therefore flattens the
wage distribution in the bottom and steepens it in the top. This is an effect that is
not considered in the plots of occupational wages that are presented in the literature.

Thus, in order to generate the rise in the bottom of the wage distribution, one may
among other things want to shut down overtaking at the bottom. This is actually
an implication of the models with one-dimensional skills or with homogenous skills
in the low-skill occupation that are used in the literature. However, it is also a
strong assumption as we would in fact expect people to change position in the wage
distribution when polarization takes place and it does nothing to align the actual and
the counterfactual change of average wages in occupations in figure IX.

Another effect, within the more general setup of the Roy model, that may repro-
duce the lower part of the actual wage distribution and occupational wages is the
reallocation from workers out of the middle-skill occupation.\footnote{So far I have ignored
\footnote{The lines are smoothed because for the counterfactual under fixed quantiles the individuals who
correspond to these quantiles exclusively determine their change. This introduces large random
variation into the quantiles of the wage distribution.}
\footnote{Workers who optimally choose to leave their initial occupations have wage increases compared
to staying. This is even true when they “switch down” to low-skill occupations and in restricted

35
the potential effect of this on wages because, without further assumptions, the model and the empirical results in the previous sections are not informative about this effect. Therefore, my analysis in the following should be considered a calibration exercise that assesses whether one can in principle match the remaining differences between the actual and counterfactual wage distribution and occupational wages.

In the data, there is a net outflow from the middle- to the low-skill and to the high-skill occupation of three and 3.5 percent of the overall workforce, respectively. I assume that the lowest earners in the middle who make up three percent of the workforce switch into the low-skill occupation and assign them a fifteen percent wage increase, that is, about half of the maximum wage increase that they could possibly obtain from switching \((31.4\% - 2.4\%)\). Figure 12 plots the resulting counterfactual wage distribution which fits the actual quite well, especially in the CPS. Moreover, figure 13 displays the corresponding changes of average wages in occupations, which are now also closer to the actual than without reallocation.

Qualitatively, the reallocation effect at the bottom seems plausible. It not only matches better the unconditional wage distribution, but in addition brings occupational wages in the actual and the counterfactual closer together. Moreover, the low-earners in the middle-skill occupations may really have a strong incentive to switch jobs once the relative demand shock hits and it is also conceivable that they could do so gainfully: for example, given probably not too different skill requirements, someone who would have been a low-earning worker in a factory in the 1980s may instead relatively easily become a janitor today.

While qualitatively plausible, the assumptions made about reallocation in order to match the wage distribution in figure 12 are relatively strong. First, the concentrated switching of low-earners in the middle-skill occupation requires that the population distribution of potential wages in the low-skill occupation be condensed so that the low-earners are the first to find it profitable to “switch down”. This is close to the models, such as the one-dimensional skill model.

An additional one percent of low earners is assumed to move to the high-skill occupation with the same wage gain.

In fact, the fit may be even better than in figures 12 and ?? if the remaining difference between actual and counterfactual is due to small-sample variation for 27 year olds. For example, I have tried out assigning the same relative price estimates and making similar assumptions about reallocation to the larger group of 25-29 year olds in the CPS. This matches the actual changes almost perfectly. The same is the case if I do the exercise for prime age males aged 25-55.
assumption of homogenous skills in the low-skill occupation as in Autor, Katz, and Kearney (2006) and Autor and Dorn (2013), and it is hard to reconcile with the fact that the empirical wage distributions of the low- and the middle-skill occupation overlap substantially in both cross-sections. Second, the gains from switching that I need to assume are substantial.\footnote{A recent paper using panel data finds that low earners in the middle-skill occupations most strongly switch down due to job polarization Cortes (2012). However, the necessary assumptions for the reallocation effect are not strictly testable in my data. This is so because I do not know individual workers’ unobserved skills in the occupations that they have not chosen and thus I cannot estimate their overall gains from reallocation. The only assessment I can make is about the gains from reallocation for the observable components of skill. It turns out that according to observable skills there is no clear evidence in favor of the idea that the low earners have the highest gains from reallocation. To see this crudely, compare figure \[4\] again: contrary to what one would expect in the case of strong switching of low-earners out of the middle-skill occupation, the average talent measures in the middle-skill occupation do not improve visibly and they do not deteriorate in the low-skill occupation.}

These results therefore suggest that the simple implementation of polarization in this paper is in principle able to match the qualitative features of the change in the actual wage distribution—with the caveat that quantitatively it doesn’t seem to get all the way to explain the bottom. A further caveat is that the above calibration exercise is rather loosely connected to the strict empirical estimations of the previous sections. The reason is that in principle any vector of occupation-specific price changes could be applied to any occupational breakup in order to evaluate its effect on the wage distribution. However, this loose connection has also an advantage: I have played around with several occupational breakups into three groups and with different relative price changes in the two datasets. None of them could generate a counterfactual wage distribution that better matched the actual wage distribution than the low-, middle-, and high-skill occupational breakup used in this paper—not even a breakup into non- routine cognitive, routine, and non-routine manual task intensities as proposed in Autor, Levy, and Murnane (2003).\footnote{As above, all these groupings share the feature that the wage dispersion within them is substantial. However, in the case of tasks, one should note that the measurement that is available to me and much of the literature is far from perfect. This is because tasks that workers carry out are assigned on the three-digit occupation level (for details see the survey paper by Acemoglu and Autor 2011), which may capture only a relatively small share of the overall variation in workers’ actual tasks. Therefore, job groupings or task measures that correspond more closely to the tasks that technology and trade have replaced may help to better match the wage distribution in the bottom, since the dispersion of wages conditional on them may also be lower. The occupation groups and task measures that are used here explain only around five to ten percent of the variation in wages in the cross-section. Hence, if it were available for my application, individual-level data on tasks as}
Thus, simply using a different occupational breakup will not be sufficient to show that job polarization can also quantitatively match the bottom of the actual wage distribution. Future research might instead try to obtain an even more precise empirical measurement of the jobs or tasks for which demand has declined\textsuperscript{[44]} In addition, one may look for further evidence of a substantial reallocation effect at the bottom of the wage distribution. The answer to the third question of the paper is therefore ambiguous: job polarization can reproduce the increase of wages at the top of the actual distribution while it can match the increase in the bottom only under strong assumptions.

8 Conclusion

This article analyzed the relationship between job polarization and workers’ wages. To do this I exploited the match of labor demand across three occupation groups with labor supply along several dimensions of talents. That is, I compared before and after polarization the wages of individuals who are differentially likely to work in high-, middle-, and low-skill occupations according to their talents, and I levered this information with the help of the Roy model of occupational choice.

My results indicate that job polarization has had substantial effects on workers’ wages, the skill prices in occupations, and on inequality in the upper half of the wage distribution. Whether polarization can explain the decrease in inequality in the lower half of the wage distribution or whether this was due to other factors could not definitively be answered. Further work on the wage effects of job polarization should therefore focus on this open question.

References


\textsuperscript{[44]}Estimating a richer specification of the model that allows for changes in the production function of occupation-specific skills by itself is unlikely to suffice because this would just allow for more flexible returns to the relatively unimportant observables.


Figure 1: Percentage Growth of the Quantiles of the Wage Distribution

Notes: The figure depicts the change in log real wages along the quantiles of the wage distribution between the two cohorts for the NLSY and the comparable years and age group in the CPS.

Figure 2: Change in Employment Shares by Occupations

Notes: The figure depicts the percentage point change in employment in the low-, middle-, and high-skilled occupations in the NLSY and the comparable years and age group in the CPS. The high-skill occupations contain managerial, professional services, and technical occupations. The middle-skill occupations contain sales, office / administrative, production, and operator and laborer occupations. The low-skill occupation contain protective, food, cleaning and personal service occupations.
Figure 3: Wage Changes by Occupations

Notes: The figure depicts the change in average real wages in low-, middle-, and high-skilled occupations for the NLSY and the comparable years and age group in the CPS. The high-skill occupations contain managerial, professional services, and technical occupations. The middle-skill occupations contain sales, office/ administrative, production, and operator and laborer occupations. The low-skill occupations contain protective, food, cleaning and personal service occupations.

Figure 4: Average Talents in Occupations, NLSY 1979 and 1997

Notes: The figures display the average math, verbal, and mechanical test scores in the three occupation groups for the NLSY79 and the NLSY97.
Figure 5: Predicted Relative Skill Returns and their Changes

(a) Propensity High-Skill Occupation

(b) Propensity Low-Skill Occupation

(c) Propensity Middle-Skill Occupation

Notes: The figures plot the returns to propensities of entering the respective occupation in the NLSY79 and the NLSY97 together with the empirical density of these propensities in the NLSY79. The returns are estimated in regressions of log wages on a constant and the respective propensity together with an interaction term for the NLSY97.
Figure 6: Actual and Predicted Wage Distribution Change

(a) Returns to Propensities  (b) Returns to All Talents

Notes: The figures plot the actual and the predicted change in the wage distribution when workers in the NLSY79 are assigned the change in the returns to their observable characteristics between the two cohorts estimated in columns one and four of table 6.

Figure 7: The Estimation Problem

\[ p_N(x_i, \tilde{\pi}_t) \]

\[ \text{E}(w_{t1}|x_i) - \text{E}(w_{t0}|x_i) = \Delta \pi_M + a + b + c \]
Figure 8: Actual and Counterfactual Wage Distribution Change, NLSY and CPS

Notes: The figure plots the actual and the counterfactual change in the wage distribution when workers in the initial period are assigned the estimated price changes in their occupations from the optimal minimum distance estimator in table 8.

Figure 9: Actual and Counterfactual Occupational Wage Changes, NLSY and CPS

Notes: The figures plot the actual and the counterfactual change in occupational wages when workers in the initial period are assigned the estimated price changes in their occupations from the optimal minimum distance estimator in table 8.
Figure 10: Share of Occupations in the Wage Distribution, 1979 Cohort

(a) NLSY  
(b) CPS

Notes: The figure plots the smoothed employment share of the low-, middle-, and high-skill occupations within the quantiles of the initial wage distribution. Smoothing is done using the predicted values from a fourth order polynomial regression of the employment shares on the quantiles.

Figure 11: Smoothed Counterfactual Wage Distribution Change, Flexible and Fixed Quantiles

(a) NLSY  
(b) CPS

Notes: The solid line in this figure depicts the growth of the wage distribution along the quantiles under the counterfactual. The dashed line depicts the growth of the original quantiles under the counterfactual, i.e. individuals are fixed at their quantiles in the original wage distribution and the growth of these fixed quantiles is computed. The lines are smoothed because for the counterfactual under fixed quantiles the individuals who correspond to these quantiles exclusively determine their change. This introduces large random variation into the growth of the quantiles. Smoothing is done using the predicted values from a fourth order polynomial regression of average wage changes on the quantiles.
Figure 12: Actual and Counterfactual Wage Distribution Change with Reallocation, NLSY and CPS

(a) NLSY

(b) CPS

Notes: The figure plots the actual and the counterfactual change in the wage distribution when workers in the initial period are assigned the estimated price changes in their occupations plus a reallocation effect: the lowest-earning three percent are assumed to move out of the middle- to the low-skill occupation with a 15 percent relative wage increase and the next low-earning one percent is assumed to move to the high-skill occupation with the same relative wage gain.

Figure 13: Actual and Counterfactual Occupational Wage Change with Reallocation, NLSY and CPS

(a) NLSY

(b) CPS

Notes: The figures plot the actual and the counterfactual change in occupational wages when workers in the initial period are assigned the estimated price changes in their occupations plus a reallocation effect: the lowest-earning three percent are assumed to move out of the middle- to the low-skill occupation with a 15 percent relative wage increase and the next low-earning one percent is assumed to move to the high-skill occupation with the same relative wage gain.
Table 1: From the full NLSY to the analysis sample

<table>
<thead>
<tr>
<th>Reason for exclusion</th>
<th>NLSY79 (Birthyears 1956-1964)</th>
<th>NLSY97 (Birthyears 1980-1984)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total males</td>
<td>6,403</td>
<td>4,599</td>
</tr>
<tr>
<td>Excluded oversampled white and older arrivers in US than age 16</td>
<td>4,585</td>
<td>4,599</td>
</tr>
<tr>
<td>Birthyear &gt; 1982</td>
<td>4,585</td>
<td>2,754</td>
</tr>
<tr>
<td>Type of attrition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ought to be present with ASVAB at age 27</td>
<td>4,585</td>
<td>2,754</td>
</tr>
<tr>
<td>% No ASVAB excluded</td>
<td>4,299</td>
<td>2,081</td>
</tr>
<tr>
<td>Not present at age 27 excluded</td>
<td>3,939</td>
<td>1,737</td>
</tr>
<tr>
<td>% Conditioned on working</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excluded who report no or farm occupation, self-employed, and those with no wage income</td>
<td>3,054</td>
<td>1,207</td>
</tr>
</tbody>
</table>

Notes: The table reports how I get from the full NLSY 1979 and 1997 to my analysis sample and where observations are lost or need to be dropped.
Table 2: Labor Supply with Respect to Average Demographics, Early, and Contemporary Skill Determinants

<table>
<thead>
<tr>
<th></th>
<th>NLSY79</th>
<th>NLSY97</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nbr of observations</strong></td>
<td>3051</td>
<td>1210</td>
</tr>
<tr>
<td><strong>Percentage of observations</strong></td>
<td>71.60</td>
<td>28.40</td>
</tr>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>27.00</td>
<td>27.00</td>
</tr>
<tr>
<td>White</td>
<td>0.80</td>
<td>0.72</td>
</tr>
<tr>
<td>Black</td>
<td>0.13</td>
<td>0.14</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.06</td>
<td>0.14</td>
</tr>
<tr>
<td><strong>Early skill determinants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFQT</td>
<td>167.31</td>
<td>167.65</td>
</tr>
<tr>
<td>Low AFQT Tercile</td>
<td>0.34</td>
<td>0.33</td>
</tr>
<tr>
<td>Middle AFQT Tercile</td>
<td>0.33</td>
<td>0.34</td>
</tr>
<tr>
<td>High AFQT Tercile</td>
<td>0.33</td>
<td>0.32</td>
</tr>
<tr>
<td>Math Score (NCE)</td>
<td>50.45</td>
<td>50.73</td>
</tr>
<tr>
<td>Verbal Score (NCE)</td>
<td>50.26</td>
<td>50.49</td>
</tr>
<tr>
<td>Mechanical Score (NCE)</td>
<td>50.41</td>
<td>50.69</td>
</tr>
<tr>
<td>Illicit Activities (NCE, Measured 1980)</td>
<td>49.98</td>
<td>50.01</td>
</tr>
<tr>
<td>Precocious Sex (NCE, Measured 1983)</td>
<td>49.91</td>
<td>50.24</td>
</tr>
<tr>
<td>Mother’s Education (Years)</td>
<td>11.86</td>
<td>13.11</td>
</tr>
<tr>
<td>Father’s Education (Years)</td>
<td>10.83</td>
<td>13.09</td>
</tr>
<tr>
<td><strong>Contemporary skill determinants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School Dropout (HSD)</td>
<td>0.12</td>
<td>0.07</td>
</tr>
<tr>
<td>High School Graduate (HSG)</td>
<td>0.43</td>
<td>0.58</td>
</tr>
<tr>
<td>Some College (SC)</td>
<td>0.20</td>
<td>0.06</td>
</tr>
<tr>
<td>College Graduate (CG)</td>
<td>0.19</td>
<td>0.24</td>
</tr>
<tr>
<td>Advanced Degree (AD)</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>North East</td>
<td>0.22</td>
<td>0.17</td>
</tr>
<tr>
<td>North Central</td>
<td>0.29</td>
<td>0.25</td>
</tr>
<tr>
<td>South</td>
<td>0.32</td>
<td>0.35</td>
</tr>
<tr>
<td>West</td>
<td>0.17</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Notes: The table shows average demographics and skill proxies in the NLSY79 and NLSY97 for all individuals weighted by hours worked. NCE indicates variables in the population (including non-workers) are standardized to “normal curve equivalents” with mean 50 and standard deviation 21.06. This is done when absolute values of these variables cannot confidently compared over the two cohorts.
Table 3: Pairwise Correlations between Composite Test Scores

<table>
<thead>
<tr>
<th></th>
<th>NLSY79 AFQT (NCE)</th>
<th>Math Score (NCE)</th>
<th>Verbal Score (NCE)</th>
<th>Mechanical Score (NCE)</th>
<th>NLSY97 AFQT (NCE)</th>
<th>Math Score (NCE)</th>
<th>Verbal Score (NCE)</th>
<th>Mechanical Score (NCE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFQT (NCE)</td>
<td>1</td>
<td>0.82</td>
<td>0.93</td>
<td>0.63</td>
<td>1</td>
<td>0.83</td>
<td>0.92</td>
<td>0.63</td>
</tr>
<tr>
<td>Math Score (NCE)</td>
<td>0.82</td>
<td>1</td>
<td>0.71</td>
<td>0.53</td>
<td>1</td>
<td>0.75</td>
<td>0.75</td>
<td>0.54</td>
</tr>
<tr>
<td>Verbal Score (NCE)</td>
<td>0.93</td>
<td>0.71</td>
<td>1</td>
<td>0.61</td>
<td>0.92</td>
<td>0.75</td>
<td>0.92</td>
<td>0.54</td>
</tr>
<tr>
<td>Mechanical Score (NCE)</td>
<td>0.63</td>
<td>0.53</td>
<td>0.61</td>
<td>0.63</td>
<td>0.63</td>
<td>0.54</td>
<td>0.63</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Nbr Observations 2936 1210

Notes: The table shows the pairwise correlations between composite test scores after standardizing to normal curve equivalents with mean 50 and standard deviation 21.06.

Table 4: Occupational Propensities from the NLSY 1979 Sorting Regressions

<table>
<thead>
<tr>
<th></th>
<th>Prop High</th>
<th>Prop Middle</th>
<th>Prop Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>1979</td>
<td>26.6</td>
<td>63.9</td>
<td>9.5</td>
</tr>
<tr>
<td>Mean</td>
<td>19.1</td>
<td>17.3</td>
<td>5.1</td>
</tr>
<tr>
<td>St.Dev.</td>
<td>6.2</td>
<td>37.6</td>
<td>4.8</td>
</tr>
<tr>
<td>p10</td>
<td>20.9</td>
<td>69.5</td>
<td>7.9</td>
</tr>
<tr>
<td>p50</td>
<td>55.7</td>
<td>82.7</td>
<td>16.7</td>
</tr>
<tr>
<td>p90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>26.2</td>
<td>63.9</td>
<td>10.0</td>
</tr>
<tr>
<td>Mean</td>
<td>18.8</td>
<td>16.9</td>
<td>5.5</td>
</tr>
<tr>
<td>St.Dev.</td>
<td>6.0</td>
<td>39.4</td>
<td>4.8</td>
</tr>
<tr>
<td>p10</td>
<td>20.7</td>
<td>69.4</td>
<td>8.1</td>
</tr>
<tr>
<td>p50</td>
<td>54.3</td>
<td>82.4</td>
<td>18.1</td>
</tr>
<tr>
<td>p90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>4146</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table reports propensities to enter the high-, middle-, and low-skill occupations in the NLSY79 and NLSY97. The propensities are from the NLSY79 only and they are from multinomial logit regressions of occupational choice including mathematical, verbal, and mechanical talent terciles, illicit activities, precocious sex and dummies for respondents’ race.
Table 5: Sorting into Occupation Groups, Multinomial Logit Regressions

<table>
<thead>
<tr>
<th></th>
<th>(1) NLSY79</th>
<th>(2) NLSY79</th>
<th>(3) NLSY97</th>
<th>(4) NLSY97</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-4.024***</td>
<td>-1.710***</td>
<td>-3.176***</td>
<td>-1.384***</td>
</tr>
<tr>
<td>Black</td>
<td>0.235</td>
<td>0.159</td>
<td>-0.152</td>
<td>-0.106</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.03</td>
<td>-0.031</td>
<td>-0.472*</td>
<td>-0.456*</td>
</tr>
<tr>
<td>Math (NCE)</td>
<td>0.047***</td>
<td></td>
<td>0.034***</td>
<td></td>
</tr>
<tr>
<td>Verbal (NCE)</td>
<td>0.023***</td>
<td></td>
<td>0.032***</td>
<td></td>
</tr>
<tr>
<td>Mechanic (NCE)</td>
<td>-0.014***</td>
<td></td>
<td>-0.019***</td>
<td></td>
</tr>
<tr>
<td>Middle Math Tercile</td>
<td>1.144***</td>
<td></td>
<td></td>
<td>0.441*</td>
</tr>
<tr>
<td>High Math Tercile</td>
<td>2.315***</td>
<td></td>
<td></td>
<td>1.426**</td>
</tr>
<tr>
<td>Middle Verbal Tercile</td>
<td>0.207</td>
<td></td>
<td></td>
<td>0.670**</td>
</tr>
<tr>
<td>High Verbal Tercile</td>
<td>0.750***</td>
<td></td>
<td></td>
<td>1.445***</td>
</tr>
<tr>
<td>Middle Mechanic Tercile</td>
<td>-0.269</td>
<td></td>
<td></td>
<td>-0.258</td>
</tr>
<tr>
<td>High Mechanic Tercile</td>
<td>-0.552***</td>
<td></td>
<td></td>
<td>-0.618**</td>
</tr>
<tr>
<td>Illicit Activities (NCE)</td>
<td>-0.009***</td>
<td></td>
<td></td>
<td>-0.003</td>
</tr>
<tr>
<td>Precocious Sex (NCE)</td>
<td>-0.004</td>
<td></td>
<td></td>
<td>-0.006</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.689***</td>
<td>-1.608***</td>
<td>-1.339***</td>
<td>-2.053***</td>
</tr>
<tr>
<td>Black</td>
<td>0.636***</td>
<td>0.762***</td>
<td>0.473*</td>
<td>0.658**</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.201</td>
<td>0.243</td>
<td>-0.216</td>
<td>-0.114</td>
</tr>
<tr>
<td>Math (NCE)</td>
<td>-0.002</td>
<td></td>
<td>-0.009</td>
<td></td>
</tr>
<tr>
<td>Verbal (NCE)</td>
<td>0.018***</td>
<td></td>
<td>0.021**</td>
<td></td>
</tr>
<tr>
<td>Mechanic (NCE)</td>
<td>-0.023***</td>
<td></td>
<td>-0.017**</td>
<td></td>
</tr>
<tr>
<td>Middle Math Tercile</td>
<td>-0.381**</td>
<td></td>
<td></td>
<td>-0.07</td>
</tr>
<tr>
<td>High Math Tercile</td>
<td>0.128</td>
<td></td>
<td></td>
<td>-0.395</td>
</tr>
<tr>
<td>Middle Verbal Tercile</td>
<td>0.342</td>
<td></td>
<td></td>
<td>0.27</td>
</tr>
<tr>
<td>High Verbal Tercile</td>
<td>0.471*</td>
<td></td>
<td></td>
<td>0.790**</td>
</tr>
<tr>
<td>Middle Mechanic Tercile</td>
<td>-0.319</td>
<td></td>
<td></td>
<td>-0.281</td>
</tr>
<tr>
<td>High Mechanic Tercile</td>
<td>-0.908***</td>
<td></td>
<td></td>
<td>-0.608*</td>
</tr>
<tr>
<td>Illicit Activities (NCE)</td>
<td>-0.002</td>
<td></td>
<td></td>
<td>0.013*</td>
</tr>
<tr>
<td>Precocious Sex (NCE)</td>
<td>-0.003</td>
<td></td>
<td></td>
<td>-0.003</td>
</tr>
<tr>
<td>Pseudo R-Squared</td>
<td>0.132</td>
<td>0.123</td>
<td>0.114</td>
<td>0.112</td>
</tr>
<tr>
<td>N</td>
<td>2936</td>
<td>2936</td>
<td>1210</td>
<td>1210</td>
</tr>
</tbody>
</table>

Notes: Each columns presents the results from a multinomial logit regression of occupational choice on demographics and talent proxies. The omitted group is the middle occupation. The first column uses only linear test scores in the NLSY79. The second column, which is the specification to estimate occupational propensities in the following, uses terciles of test scores and adds measures of risky behavior. The last two columns repeat these estimations for the NLSY97. In order to save space, standard errors are not reported but statistical significance is indicated: * p<0.1, ** p<0.05, *** p<0.01.
Table 6: Returns to Occupational Propensities over the Two Cohorts

<table>
<thead>
<tr>
<th></th>
<th>(1) Log Wage</th>
<th>(2) Log Wage</th>
<th>(3) Log Wage</th>
<th>(4) Log Wage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>181.15***</td>
<td>185.17***</td>
<td>176.66***</td>
<td>183.21***</td>
</tr>
<tr>
<td></td>
<td>(3.10)</td>
<td>(3.11)</td>
<td>(3.76)</td>
<td>(21.61)</td>
</tr>
<tr>
<td>Const x NLSY97</td>
<td>-7.90</td>
<td>-10.27</td>
<td>-12.59</td>
<td>-43.37</td>
</tr>
<tr>
<td></td>
<td>(6.74)</td>
<td>(6.57)</td>
<td>(8.16)</td>
<td>(41.62)</td>
</tr>
<tr>
<td>Prop High Occup</td>
<td>0.31***</td>
<td>0.03</td>
<td>-0.06</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.57)</td>
</tr>
<tr>
<td>Prop H Occ x NLSY97</td>
<td>0.29***</td>
<td>0.25**</td>
<td>0.30**</td>
<td>1.41</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(1.03)</td>
</tr>
<tr>
<td>Prop Low Occup</td>
<td>-1.65***</td>
<td>-1.80***</td>
<td>-1.75***</td>
<td>-2.19**</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.17)</td>
<td>(0.17)</td>
<td>(0.97)</td>
</tr>
<tr>
<td>Prop L Occ x NLSY97</td>
<td>0.70*</td>
<td>0.86**</td>
<td>0.91**</td>
<td>2.26</td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
<td>(0.38)</td>
<td>(0.38)</td>
<td>(1.92)</td>
</tr>
<tr>
<td>College</td>
<td>19.23***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.92)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coll x NLSY97</td>
<td></td>
<td>4.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>4154</td>
<td>4149</td>
<td>4149</td>
<td>4154</td>
</tr>
<tr>
<td>R²</td>
<td>0.09</td>
<td>0.11</td>
<td>0.12</td>
<td>0.10</td>
</tr>
<tr>
<td>Degree dummies</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Talents directly</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: The table reports OLS wage regressions of 100 times the deflated log wage on propensities to enter occupations (predicted relative occupation-specific skills) and the change in the coefficient between the NLSY79 and the NLSY97. The propensities are from the NLSY79 only and they are from multinomial logit regressions of occupational choice including mathematical, verbal, and mechanical talent terciles, illicit activities, precocious sex and dummies for respondents’ race. “x NLSY97” stands for the interaction between the variable and an NLSY97 dummy, i.e. the change in the coefficient between the NLSY79 and the NLSY97. The specifications in columns two to four add dummies for college degree, detailed education (HS drop out, HS graduate, Some college, College and above), and the talents that were used in the estimation of the propensities directly. Standard errors are from bootstrapping the first (estimating the propensities) and second stage regressions together 500 times and they are reported below the coefficients. * p<0.1, ** p<0.05, *** p<0.01.
### Table 7: Talent Allocation and Returns Changes

<table>
<thead>
<tr>
<th></th>
<th>High-Skill Occup</th>
<th>Low-Skill Occup</th>
<th>Log Wage x NLSY97</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td>18.71***</td>
<td>11.87***</td>
<td>14.37**</td>
</tr>
<tr>
<td></td>
<td>(6.21)</td>
<td>(5.51)</td>
<td>(2.08)</td>
</tr>
<tr>
<td><strong>Black</strong></td>
<td>-0.762</td>
<td>9.292***</td>
<td>0.647</td>
</tr>
<tr>
<td></td>
<td>(-0.43)</td>
<td>(4.77)</td>
<td>(0.14)</td>
</tr>
<tr>
<td><strong>Hispanic</strong></td>
<td>-2.632</td>
<td>1.708</td>
<td>-1.586</td>
</tr>
<tr>
<td></td>
<td>(-1.34)</td>
<td>(1.12)</td>
<td>(-0.36)</td>
</tr>
<tr>
<td><strong>Middle Math Tercile</strong></td>
<td>10.83***</td>
<td>-4.466***</td>
<td>-2.615</td>
</tr>
<tr>
<td></td>
<td>(6.05)</td>
<td>(-2.85)</td>
<td>(-0.56)</td>
</tr>
<tr>
<td><strong>High Math Tercile</strong></td>
<td>34.90***</td>
<td>-5.997***</td>
<td>10.44*</td>
</tr>
<tr>
<td></td>
<td>(13.48)</td>
<td>(-3.10)</td>
<td>(1.68)</td>
</tr>
<tr>
<td><strong>Middle Mechanic Tercile</strong></td>
<td>-2.505</td>
<td>-2.332</td>
<td>-5.767</td>
</tr>
<tr>
<td></td>
<td>(-1.19)</td>
<td>(-1.52)</td>
<td>(-1.20)</td>
</tr>
<tr>
<td><strong>High Mechanic Tercile</strong></td>
<td>-7.043***</td>
<td>-4.827***</td>
<td>-1.740</td>
</tr>
<tr>
<td></td>
<td>(-2.81)</td>
<td>(-2.97)</td>
<td>(-0.30)</td>
</tr>
<tr>
<td><strong>Middle Verbal Tercile</strong></td>
<td>3.505*</td>
<td>2.429</td>
<td>-0.282</td>
</tr>
<tr>
<td></td>
<td>(1.84)</td>
<td>(1.48)</td>
<td>(-0.06)</td>
</tr>
<tr>
<td><strong>High Verbal Tercile</strong></td>
<td>15.34***</td>
<td>2.805</td>
<td>-4.535</td>
</tr>
<tr>
<td></td>
<td>(5.67)</td>
<td>(1.45)</td>
<td>(-0.65)</td>
</tr>
<tr>
<td><strong>Illicit Activities (NCE)</strong></td>
<td>-0.129***</td>
<td>0.0388</td>
<td>-0.183*</td>
</tr>
<tr>
<td></td>
<td>(-3.33)</td>
<td>(1.36)</td>
<td>(-1.89)</td>
</tr>
<tr>
<td><strong>Precocious Sex (NCE)</strong></td>
<td>-0.0612</td>
<td>-0.0120</td>
<td>0.0527</td>
</tr>
<tr>
<td></td>
<td>(-1.64)</td>
<td>(-0.41)</td>
<td>(0.62)</td>
</tr>
</tbody>
</table>

R-squared 0.182 0.0281 0.0933
N 4146 4146 4146

Notes: The first two columns present the coefficients from OLS allocation regressions of working in the high- and the low-skill occupation with pooled NLSY79 and NLSY97 data. The third column presents the change in the parameters between the two cohorts in an OLS wage regression. Coefficients represent 100 times the average partial increase in the probability of entering the occupation group and the log wage, respectively, for an additional unit of the regressor. “x NLSY97” stands for the interaction between the variable and an NLSY97 dummy, i.e. the change in the coefficient between the NLSY79 and the NLSY97. T-statistics below the coefficients. * p<0.1, ** p<0.05, *** p<0.01.
<table>
<thead>
<tr>
<th>Method</th>
<th>Estim. $\Delta (\pi_H - \pi_M)$ in % (s.e.)</th>
<th>Estim. $\Delta (\pi_L - \pi_M)$ in % (s.e.)</th>
<th>Implied $\Delta \pi_M$ in %</th>
<th>Test Statistic (p-value in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMD / Full GLS</td>
<td>20.1 (9.7)</td>
<td>31.4 (35.2)</td>
<td>-2.4 (10.7)</td>
<td>13.1 (10.7)</td>
</tr>
<tr>
<td>EWMD / OLS</td>
<td>19.4 (10.8)</td>
<td>-4.4 (32.0)</td>
<td>1.7 (10.5)</td>
<td>13.2 (10.5)</td>
</tr>
<tr>
<td>DWMD / WLS</td>
<td>22.0 (9.7)</td>
<td>-7.5 (35.1)</td>
<td>1.3 (19.1)</td>
<td>11.2 (19.1)</td>
</tr>
</tbody>
</table>

Notes: The table presents estimated relative wage rate changes in the high- and the low-skill occupation compared to the middle skill occupation, a point estimate for the absolute wage rate change in the middle, and the cross-equation restriction test of the polarization hypothesis. The characteristics used in the underlying allocation and wage regressions are my preferred specification, i.e. mathematical, verbal, and practical talent terciles, illicit activities, precocious sex, and dummies for respondents’ race. There are 8 degrees of freedom for the test (10 coefficients minus 2 parameters estimated on them). Implied prices and the test statistics are reported for the Optimal Minimum Distance (full feasible GLS) estimation and as alternatives for the Equally Weighted Minimum Distance (OLS regression of change in wage regression coefficients on allocation regression coefficients), and Diagonally Weighted Minimum Distance (WLS).
Appendix Intended for the Web

A Detailed Sample Construction

I use data from the National Longitudinal Survey of Youth (NLSY) cohort of 1979 and 1997. The individuals in these surveys are born between 1956 and 1964 and between 1980 and 1984, respectively. As is necessary for this paper, the NLSY studies provide detailed information about individuals’ background, education, and labor market outcomes. Moreover, the two cohorts are specifically designed to to be comparable to one another.

Consistent with many papers on the NLSY and in the literature on polarization and wage inequality, I restrict my attention to males (e.g. Firpo, Fortin, and Lemieux 2011, Cortes 2012). There are several reasons for doing this. Firstly, polarization seems to have had the most dire effect on males (Acemoglu and Autor 2011). Secondly, female hours worked and thus the type of selection of females into the labor market (see Mulligan and Rubinstein 2008) changed substantially over the two NLSYs. In addition, females made strides in educational attainment, their wages rose across the whole distribution, and attitudes towards them and discrimination against them in the labor market seem to have changed drastically. Thus, there are diverse changes in (the structure of) female labor supply and demand that are likely to work aside from the forces of polarization. Restricting the analysis to males provides a cleaner comparison of workers across the two decades between the NLSY79 and the NLSY97.

I evaluate individuals’ labor market outcomes at age 27. This is because, on the one hand, at younger ages the polarization facts that the paper sets out to analyze are not very pronounced in the NLSY as well as CPS data, which I use for comparison. On the other hand, at older ages than 27, I would lose too many observations from the NLSY97 as, at the time of the analysis, data was only available until the survey year of 2009. With the age 27 restriction, I already have to drop about two fifth of the NLSY97 sample (birth years 1983 and 1984 are dropped).

Table 1 summarizes how the sample restrictions, attrition, and labor market participation for males reduce my sample size from 6,403 to 3,054 and from 4,599 to 1,207 males in the NLSY79 and the NLSY97, respectively. I restrict the sample to
individuals who participated in the Armed Services Vocational Aptitude Battery of tests (ASVAB) in the first survey year. This restriction is necessary because ASVAB will provide me with measures of different dimensions of talent for each individual that are comparable over the two cohorts. Moreover, I argue that the subtests from ASVAB are proxies of individuals’ fundamental talents that do not react as elastically to changes in market returns as late skill determinants, such as education, which have been used in existing studies.

The participation in ASVAB is substantially lower in the NLSY97 than the NLSY79 where almost everyone participated. Moreover, sample attrition at age 27 is higher in the NLSY97 than the NLSY79 and overall only 63 percent of the NLSY79 participated in ASVAB and are also present at age 27. This problem is well known (e.g. Altonji, Bharadwaj, and Lange 2012, Aughinbaugh and Gardecki 2007). More generally, attrition rates in several panel surveys in the United States increased substantially during the 1990s (see also Fitzgerald, Gottschalk, and Moffitt 1998, MaCurdy, Mroz, and Gritz 1998). The attrition and non-test-participation rates in my data closely line up with those reported in the study by Altonji, Bharadwaj, and Lange (henceforth ABL). The only difference is that ABL consider outcomes at the younger age of 22 and thus have slightly lower attrition rates.

In their paper, ABL note that the higher attrition rate in the NLSY97 may be partly due to NLSY97 respondents being first interviewed at ages 12-16 versus ages 14-21 for the NLSY79 and thus had more time to attrit. ABL further extensively examine the potential non-randomness of attrition and non-test-participation and its likely impact in biasing important labor market outcomes. Aughinbaugh and Gardecki (2007) do a similar exercise but focus on social and educational outcomes. Both studies find evidence that attrition is not random with respect to youths’ outcomes and their backgrounds. However, Aughinbaugh and Gardecki (2007) conclude that attrition from the NLSY97 does not appear to affect inference when estimating the three outcomes at age 20 that they are considering and ABL decide that the differences between non-attritors and the whole sample are not forbidding.

Moreover, ABL carefully select the samples of NLSY79 and NLSY97 to make them comparable to one another and compute weights that adjust for attrition and non-test-participation on observable characteristics. I closely follow their procedures
for constructing my own sample. Thus, for even more information on the sample construction and statistics on the effects of attrition, please refer to ABL in addition to the description provided here.45 First, I follow ABL in excluding from the NLSY79 immigrants who arrived in the United States after age 16. This is done because the scope of the NLSY97 (age 12-16) also doesn’t include older than age 16 arrivals. Second, I exclude the economically disadvantaged whites and military supplemental samples from the NLSY79 because they were discontinued early on in the survey and thus don’t provide labor market outcomes at age 27 (or for ABL’s purposes). Table 1 reports that 1,818 observations are dropped by making these restrictions to the sample. For each individual I retain the observation that is closest to 27 years and 6 months of age and then measure labor market and final educational outcomes from this observation.

ABL use a probit model to adjust the NLSY79 and NLSY97 base year sample weights to account for attrition and non-test-participation according to several observable characteristics, such as parental education, parental presence at age 14, indicators by birth-year, urban and SMSA residence status, indicator variables for race and gender, and an interviewer coded variable describing the attitude of the respondent during the interview. I also employ a probit model to adjust weights for attrition and non-test-participation and use the same specification and variables as ABL apart from leaving out parental presence at age 14. Alternatively, a fully stratified set of indicators for birthyear, year, sex, and race, as employed by the Bureau of Labor Statistics for weighting, yields very similar results.46 As ABL do in their paper, I proceed from this point with the assumption that, after attrition weighting, my two NLSY samples are representative of the population of young Americans that they are supposed to cover. These samples have the size of 3,939 and 1,737 individuals in the NLSY79 and the NLSY97, respectively.

I follow Lemieux (2006), who uses CPS May Outgoing Rotation Group data, in how I compute wages and in defining the sample of working individuals (henceforth labor supply). First, I use hourly wages reported for the current main job instead of imputing hourly wages from last year’s income and total hours worked. Lemieux

45I am extremely grateful to Prashant Bharadwaj for providing me with their data and do-files. 46I thank Steve McClaskie and Jay Zagorsky for providing me with the official attrition-adjusted sample weighting program for the NLSY.
(2006) convincingly argues that the current main job measure is substantially more accurate because it better measures the wages of workers paid by the hour. Moreover, the reporting of weeks and hours per year worked in the NLSY seems somewhat inconsistent over the two cohorts. I normalize all wages to 1979 real values by adjusting with the PCE deflator provided by the St.Louis Federal Reserve Bank. While Lemieux (2006) removes outliers with 1979 real wages below $1 and above $100, I remove the high wages from $40 onward because my NLSY wage data is very inaccurate for values above this threshold.

Finally, in order to condition on the sample of working individuals, I keep all individuals who report not to be self-employed, and who are employed in a non-farm, non-fishing and non-forestry occupation according to the Census 1990 three-digit occupation classification. This leaves me with an analysis sample of 3,054 and 1,207 males in the NLSY79 and NLSY97, respectively (compare table 1 again). I weight all of those individuals by the number of hours that they work per week on top of the sample weights that are adjusted for test-participation and attrition. Lemieux (2006) argues that weighting by weekly hours can be viewed as a reasonable compromise between concentrating on full-time workers only and looking at all workers including part-time workers. An additional advantage from this is that I am not losing any more observations from a full-time work restriction.

B Generalization of the Theory to Three Occupations

In the following I derive predictions (8) and (9) for the three-occupation case. For ease of exposition, wages in occupations (4) are reproduced here:

\[ w_{Kit} = \pi_{Kit} + s_{Kit}. \]

\(^{47}\)Source: “Personal Consumption Expenditures: Chain-type Price Index (PCECTPI)”, accessed 2012-8-14, [http://research.stlouisfed.org/fred2/series/PCECTPI](http://research.stlouisfed.org/fred2/series/PCECTPI)
Note that $s_{Kit}$ can be a general function of observable talents $x_{it}$ and unobservables. From equation (3) and the wages in occupations we have:

$$w_{it} = \begin{cases} 
  w_{Hit} = \pi_{Ht} + s_{Hit} & \text{if } H_{it} = 1 \\
  w_{Mit} = \pi_{Mt} + s_{Mit} & \text{if } M_{it} = 1 \\
  w_{Lit} = \pi_{Lt} + s_{Lit} & \text{if } L_{it} = 1 
\end{cases}$$

When occupational wage rates change, by the envelope theorem, the marginal change in worker $i$’s wage becomes

$$dw_{it} = \begin{cases} 
  d\pi_{H} & \text{if } H_{it} = 1 \\
  d\pi_{M} & \text{if } M_{it} = 1 \\
  d\pi_{L} & \text{if } L_{it} = 1. 
\end{cases}$$

Thanks to its linearity, the change in the expectation can be written as

$$E(dw_{it}|x_{it}, \pi_{t}) = p_{H}(x_{it}, \pi_{t})d\pi_{H} + p_{M}(x_{it}, \pi_{t})d\pi_{M} + p_{L}(x_{it}, \pi_{t})d\pi_{L},$$

where $p_{K}(x_{it}, \pi_{t})$ is the probability for an individual of talent vector $x_{it}$ to enter occupation $K$ under prices $\pi_{t}$. Exploiting that the three probabilities sum to one gives the three-occupation analogue to equation (8):

$$dE(w_{it}|x_{it}, \pi_{t}) = d\pi_{Mt} + p_{H}(x_{it}, \tilde{\pi}_{HMt}, \tilde{\pi}_{LMt})d\tilde{\pi}_{HMt} + p_{L}(x_{it}, \tilde{\pi}_{HMt}, \tilde{\pi}_{LMt})d\tilde{\pi}_{LMt}, \tag{17}$$

where $\tilde{\pi}_{K Mt} \equiv \pi_{Kt} - \pi_{Mt}$ for $K \in \{H, L\}$,

$$p_{H}(x_{it}, \tilde{\pi}_{HMt}, \tilde{\pi}_{LMt}) = \Pr[s_{Hit} - s_{Mit} > -(\pi_{Ht} - \pi_{Mt})],$$

$$s_{Hit} - s_{Lit} > -(\pi_{Ht} - \pi_{Lt})],$$

and similarly for $p_{L}(x_{it}, \tilde{\pi}_{HMt}, \tilde{\pi}_{LMt})$.

To save on notation, omit the dependence on $x_{it}$ for now. Holding constant $\tilde{\pi}_{HMt}$
and $\tilde{\pi}_{LMt}$ at $t = 0$ and integrating equation (17) with respect to $\pi_{M1}$ we get

$$E(w_{it}|\pi_{M1}, \tilde{\pi}_{H0}, \tilde{\pi}_{L0}) - E(w_{it}|\pi_{M0}, \tilde{\pi}_{H0}, \tilde{\pi}_{L0}) = \Delta\pi_{M}.$$  

Similarly,

$$E(w_{it}|\pi_{M1}, \tilde{\pi}_{H1}, \tilde{\pi}_{L0}) - E(w_{it}|\pi_{M1}, \tilde{\pi}_{H0}, \tilde{\pi}_{L0}) = \int_{\tilde{\pi}_{H0}}^{\tilde{\pi}_{H1}} p_H(\tilde{\pi}_{H1}, \tilde{\pi}_{L0}) d\tilde{\pi}_{H1}$$

$$E(w_{it}|\pi_{M1}, \tilde{\pi}_{H1}, \tilde{\pi}_{L1}) - E(w_{it}|\pi_{M1}, \tilde{\pi}_{H0}, \tilde{\pi}_{L0}) = \int_{\tilde{\pi}_{L0}}^{\tilde{\pi}_{L1}} p_{L}(\tilde{\pi}_{H1}, \tilde{\pi}_{L1}) d\tilde{\pi}_{L1}.$$ 

Summing these three expressions gives the three-occupation analogue to equation (10):

$$E(w_{it}|\pi_{1}) - E(w_{it}|\pi_{0}) = \Delta\pi_{M} + \int_{\tilde{\pi}_{H0}}^{\tilde{\pi}_{H1}} p_H(\tilde{\pi}_{H1}, \tilde{\pi}_{L0}) d\tilde{\pi}_{H1} +$$

$$+ \int_{\tilde{\pi}_{L0}}^{\tilde{\pi}_{L1}} p_{L}(\tilde{\pi}_{H1}, \tilde{\pi}_{L1}) d\tilde{\pi}_{L1}$$

(18)

### C Estimating the Price Changes for Three Occupations

Analogous to the estimation of the relative price change using equation (9), I want to estimate the relative price changes $\Delta(\pi_{H} - \pi_{M})$ and $\Delta(\pi_{L} - \pi_{M})$ using equation (18) for the three-occupation case. Since I do not know the choice probabilities on the adjustment path, they have to be approximated analogously to equation (10) and figure 7.

$$p_{H}(\tilde{\pi}_{HM1}, \tilde{\pi}_{LM0}) \approx p_{H}(\tilde{\pi}_{HM0}, \tilde{\pi}_{LM0}) + \frac{p_{H}(\tilde{\pi}_{HM1}, \tilde{\pi}_{LM1}) - p_{H}(\tilde{\pi}_{HM0}, \tilde{\pi}_{LM0})}{\tilde{\pi}_{HM1} - \tilde{\pi}_{HM0}}(\tilde{\pi}_{HM1} - \tilde{\pi}_{HM0})$$

$$p_{L}(\tilde{\pi}_{HM1}, \tilde{\pi}_{LM1}) \approx p_{L}(\tilde{\pi}_{HM0}, \tilde{\pi}_{LM0}) + \frac{p_{L}(\tilde{\pi}_{HM1}, \tilde{\pi}_{LM1}) - p_{L}(\tilde{\pi}_{HM0}, \tilde{\pi}_{LM0})}{\tilde{\pi}_{LM1} - \tilde{\pi}_{LM0}}(\tilde{\pi}_{LM1} - \tilde{\pi}_{LM0}).$$

I plug these linear approximations into (18) and multiply the result on both sides by the residual from regressing the $j$th component of the talent vector $x_{it}$ onto the

---

48Note that one might prefer using $p_{H}(\tilde{\pi}_{HM1}, \tilde{\pi}_{LM0})$ instead of $p_{H}(\tilde{\pi}_{HM1}, \tilde{\pi}_{LM1})$ in the first approximation and $p_{L}(\tilde{\pi}_{HM1}, \tilde{\pi}_{LM0})$ instead of $p_{L}(\tilde{\pi}_{HM0}, \tilde{\pi}_{LM0})$ in the second, which are not observable in the data. Yet, $p_{H}(\tilde{\pi}_{HM1}, \tilde{\pi}_{LM0}) > p_{H}(\tilde{\pi}_{HM1}, \tilde{\pi}_{LM1})$ while $p_{L}(\tilde{\pi}_{HM1}, \tilde{\pi}_{LM0}) < p_{L}(\tilde{\pi}_{HM0}, \tilde{\pi}_{LM0})$, so this additional approximation error should not be too large.
other observable talents. For convenience, and although it is an abuse of notation, denote this residual by $x_{jit}$ as well. Then, I take expectations with respect to $x_{jit}$ on both sides of (18) and divide by the variance of $x_{jit}$. This yields $J$ different moment conditions in terms of first stage linear regression coefficients:

$$\triangle \gamma_j = \frac{\delta_{Hj0} + \delta_{Hj1}}{2} \triangle (\pi_H - \pi_M) + \frac{\delta_{Lj0} + \delta_{Lj1}}{2} \triangle (\pi_L - \pi_M),$$

(19)

where $\delta_{Kjt} = \frac{\text{cov}(K_{it}, x_{jit})}{\text{var}(x_{jit})}$ with $\delta_{Hjt} + \delta_{Mjt} + \delta_{Ljt} = 0$, $K_{it} \epsilon \{H_{it}, M_{it}, L_{it}\}$, and $\gamma_{jt} = \frac{\text{cov}(w_{it}, x_{jit})}{\text{var}(x_{jit})}$. These parameters can be recovered from OLS allocation

$$K_{it} = \delta_{K0t} + \delta_{K1t} x_{1it} + \delta_{K2t} x_{2it} + ... + \delta_{KJt} x_{Jit} + v_{Kit},$$

(20)

and wage regressions

$$w_{it} = \gamma_{0t} + \gamma_{1t} x_{1it} + \gamma_{2t} x_{2it} + ... + \gamma_{Jt} x_{Jit} + u_{it}.$$

(21)

Therefore, result (13) provides a straightforward procedure to assess polarization’s effect on the returns to detailed talents. First, I run seemingly unrelated wage and allocation regressions on the individual level in both points in time to obtain estimates $\hat{\delta}_{Hjt}$, $\hat{\delta}_{Ljt}$, and $\hat{\gamma}_{jt}$ as well as an estimate of their joint covariance matrix. Second, I combine the prices $\triangle \tilde{\pi}_H \equiv \triangle (\pi_H - \pi_M)$, $\triangle \tilde{\pi}_L \equiv \triangle (\pi_L - \pi_M)$, and $\triangle \tilde{\pi} = [\triangle \tilde{\pi}_H, \triangle \tilde{\pi}_L]$. I also define the $J \times 1$ vectors $\Delta \gamma$, $\hat{\delta}_K \equiv \frac{\hat{\delta}_{K0t} + \hat{\delta}_{K2t}}{2}$ for $K \epsilon \{H, L\}$, and $m(\triangle \tilde{\pi}) = \hat{\Delta} \gamma - \hat{\Delta} \gamma H \triangle \tilde{\pi}_H - \hat{\Delta} \gamma L \triangle \tilde{\pi}_L$.

The, the minimum distance estimator minimizes the quadratic form

$$Q(\triangle \tilde{\pi}) = m(\triangle \tilde{\pi})^t W m(\triangle \tilde{\pi}),$$

with $W$ a $J \times J$ weighting matrix. Depending on $W$, the minimizing wage rate changes can be the Equally Weighted Minimum Distance (EWMD) estimator if $W = I$, the Optimal Minimum Distance (OMD) estimator if $W = [\text{Var}(m(\triangle \tilde{\pi}))]^{-1}$, and the Diagonally Weighted Minimum Distance (DWMD) estimator if $W = [\text{diag}(\text{Var}(m(\triangle \tilde{\pi})))]^{-1}$.

\[49\] Also, note the oppressed dependence of the conditional expectation and the probabilities on $x_{it}$.

\[50\] The methods applied in the following can be found in the statistical appendix of Abowd and Card (1989) or chapter 6.7 of Cameron and Trivedi (2005). I explain them step by step.
The EWMD can be implemented by a simple OLS regression of $\hat{\Delta}\gamma$ on $\hat{\delta}_H$ and $\hat{\delta}_L$, the OMD by a (feasible) GLS regression, and the DWMD by weighted least squares. As the name suggests, the OMD is asymptotically optimal.

Under the polarization hypothesis, $Em(\Delta\tilde{\pi}) = 0$ and the variance of $m(\Delta\tilde{\pi})$ can be derived up to the parameter vector $\Delta\tilde{\pi}$ from the covariance matrix of the reduced form estimates:

$$
Var(m(\Delta\tilde{\pi})) = \hat{\text{Var}}(\hat{\Delta}\gamma) + \Delta\tilde{\pi}_H^2\hat{\text{Var}}(\hat{\delta}_H) + \Delta\tilde{\pi}_L^2\hat{\text{Var}}(\hat{\delta}_L) + 2\Delta\tilde{\pi}_H\Delta\tilde{\pi}_L\hat{\text{Cov}}(\hat{\delta}_H, \hat{\delta}_L) - 2\Delta\tilde{\pi}_H\hat{\text{Cov}}(\hat{\Delta}\gamma, \hat{\delta}_H) - 2\Delta\tilde{\pi}_L\hat{\text{Cov}}(\hat{\Delta}\gamma, \hat{\delta}_L)
$$

Since $\Delta\tilde{\pi}$ is unknown in $\hat{\text{Var}}(m(\Delta\tilde{\pi}))$, I run two step feasible GLS with the first stage being OLS using $W = I$ and plugging the resulting $\hat{\Delta}\tilde{\pi}_{OLS}$ into the weighting matrix $W = \hat{\text{Var}}(m(\Delta\tilde{\pi}))$ for the second step. The minimized value of the objective function can be shown to be chi-squared distributed asymptotically:

$$
m(\Delta\tilde{\pi}_{FGLS})'[\hat{\text{Var}}(m(\Delta\tilde{\pi}_{FGLS}))]^{-1}m(\Delta\tilde{\pi}_{FGLS}) \overset{\text{a}}{\sim} \chi^2(J - 2),
$$

which provides the specification test.

Since there are concerns about small sample bias of $\hat{\Delta}\tilde{\pi}_{FGLS}$ (in particular Altonji and Segal 1996, Pischke 1995), results for $\hat{\Delta}\tilde{\pi}_{OLS}$ and $\hat{\Delta}\tilde{\pi}_{WLS}$ are also reported. In this case, the test statistic for the model test has to be adjusted (see Abowd and Card 1989).

\footnote{Under some regularity conditions we can apply a central limit theorem to the OLS estimates $\hat{\delta}_H$, $\hat{\delta}_L$, and $\hat{\gamma}_t$ as well as to $m(\Delta\tilde{\pi})$:

$$
\sqrt{N}m(\Delta\tilde{\pi}) \overset{\text{d}}{\sim} \mathcal{N}(Em(\Delta\tilde{\pi}), N\text{Var}(m(\Delta\tilde{\pi}))).
$$}