The London School of Economics and Political Science

Three Essays on the Allocation of Talent

Michael Johannes Boehm

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy, London, April 2013

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 60,391 words.

Statement of conjoint work

I confirm that Chapter 2 was jointly co-authored with Dr Martin Watzinger and I contributed 50% of this work.

Abstract

In my thesis I investigate the causes and the effects of the allocation of workers into occupations, sectors, and locations. My analysis is substantially aided by the availability of new data on workers' talents (or skills). The first chapter of the thesis exploits the fact that workers choose occupations according to their talents in order to study the effects on wages of the declining demand for manufacturing and clerical occupations. This is done by relating the occupational choices and the wages associated with particular talents over two representative cohorts of young workers in the United States between the late 1980s and the late 2000s. The second chapter, which is conjoint work, analyses the effect of an inflow of talent on productivity and output in the academic sector. We exploit the countercyclical relative attractiveness of academia as an employer over the business cycle to study periods of high (*recessions*) and low (*booms*) inflow of talent into that sector. Finally, the third chapter shows that government policy in the form of commuting tax breaks has substantial effects on the allocation of workers into jobs and residences. In particular, I exploit two reductions of tax breaks for commuting in 2003/4 and 2006/7 in Germany to estimate commuting costs' effects on workers' decisions to change the location of their job and/or their house.

Contents

List of Figures

List of Tables

Chapter 1

Has Job Polarization Squeezed the Middle Class? Evidence from the Allocation of Talents

Has Job Polarization Squeezed the Middle Class? Evidence from the Allocation of Talents

Michael J. Boehm *

April 2013

Abstract Over the last two decades, earnings in the United States increased at the top and at the bottom of the wage distribution but not in the middle—the intensely debated middle class squeeze. At the same time there was a substantial decline of employment in middle-skill production and clerical occupations—so-called job polarization. I study whether job polarization has caused the middle class squeeze. So far little evidence exists about this because the endogenous selection of skills into occupations prevents credible identification of polarization's effect on wages. I solve the selection-bias problem by studying the changes in returns to occupation-specific skills instead of the changes in occupational wages using data over the two cohorts of the National Longitudinal Study of Youth (NLSY). This data features multidimensional and pre-determined test scores, which predict occupational sorting and thus measure relative occupation-specific skills. My estimation equations are derived from the Roy (1951) model over two cross-sections with job polarization amounting to a shift in the occupation-specific skill prices. In line with polarization, I find that a one percentage point higher propensity to enter high- (low-) as opposed to middle-skill occupations is associated with a .29 (.70) percent increase in expected wages over time. I then compute a counterfactual wage distribution using my estimates of the shifts in occupation-specific skill prices and show that it matches the increase at the top of the wage distribution but fails to explain the increase at the bottom. Thus, despite the strong association of job polarization with changes in the returns to occupation-specific skills, there remains room for alternative (e.g. policy related) explanations about the increase in the lower part of the wage distribution.

Keywords: Job Polarization; Wage Inequality; Talent Allocation; Roy Model
JEL Classification Numbers: J21, J23, J24, J31

---

* London School of Economics and Centre for Economic Performance, m.j.boehm@lse.ac.uk. 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 seminar participants at LSE and at the Centre for Economic Performance Annual Conference for comments. I also thank Joseph Altonji, Prashant Bharadwaj, and Fabian Lange for providing me with their data and code, and Steve McClaskie for support with the NLSY data.

1.1 Introduction

Over the last two decades, wages of middle class workers in the United States have been squeezed, in that earnings in the middle of the wage distribution have stagnated or even fallen while earnings at the top and at the bottom have increased. This has coincided with a decrease of employment in middle-skill production and clerical occupations, and an increase of employment in low-skill services and high-skill professional and managerial occupations—so-called job polarization. Many economists believe that job polarization and the middle class squeeze are two sides of the same coin. In particular, they think that a negative demand shock for the middle-skill occupations has simultaneously reduced middle-skill employment and middle-class wages. If this is true, the middle class squeeze is a consequence of market forces, and it will be difficult to design policies that reverse the trend and help the middle class without hampering the efficiency of the economy.[1]

However, there is little evidence so far which establishes a direct link between job polarization and the middle class squeeze. On the one hand, a large body of research in labor economics and international trade has found a drop in the demand for jobs that can be replaced by computers or off-shored and shown that many of these jobs are in middle-skill occupations. On the other hand, there are plenty of hypotheses about other factors which could have contributed to the U-shaped change in wage inequality that characterizes the middle class squeeze—including increases in the minimum wage, de-unionization, and the deregulation of financial and related professions. If such policy-related or institutional factors have caused the downward pressure on the middle of the wage distribution, policy makers may be called to action in order to support the middle class.

The goal of my paper is thus to answer the question: does job polarization explain the middle class squeeze? I do this by studying how the wages of workers who would have chosen the high-, middle-, or low-skill occupations in the 1980s have changed over time. To be exact, since the same workers cannot be observed both before and after polarization has taken place, I study the returns to talents that are associated with choosing the particular occupations over time. The Roy (1951)

[1]The struggles of the middle class are a major issue in the public and political debate. For example, this editorial in the International Herald Tribune from August 30, 2012 takes the market-based view: "The economic reality is that, thanks to smart machines and global trade, the well-paying, middle-class jobs that were the backbone of Western democracies are vanishing. Neither Mitt Romney's smaller state nor Barack Obama's larger one will bring them back."(Freeland 2012)

model of self-selection into sectors guides my empirics: in the model, workers' skills in occupations are made up of observable as well as unobservable components and the returns to talents that I estimate only reflect the observable part. However, using the sorting of- and the returns to observables, I can estimate the shifts in occupation-specific prices per unit of skill, which also apply to the unobservables, and examine how much of the middle class squeeze they explain. In addition, I assess the role that heterogeneous gains from switching occupations may play for the change of the wage distribution.

So far, the fundamental problem in linking job polarization to the wage distribution has been that one could not estimate the effect of occupational demand on workers' wages. Job choices are naturally dependent on the price movements so that the skill selection into occupations changes endogenously. Hence, a comparison of wages in high-, middle-, and low-skill occupations over time would confound the relative demand shifts with a changing composition of workers' skills in each occupation.[2] The problem is exemplified by the fact that average wages in the middle-skill occupations have not declined compared to average wages in the low-skill occupations in several datasets and samples (for example, Goos and Manning 2007, and the data used here).

The point of departure for my analysis is the regression equation formulation of the Roy model as in Heckman and Sedlacek (1985):[3] every worker possesses a vector of talents which combine into skills in each occupation and which are only partly observed in the data. The log wage offered to workers in a given occupation is then the sum of an occupation-specific log skill price, which is the regression intercept, an observable component of skill, which is the regressor, and an unobservable component of skill, which is the orthogonal regression error. In this framework, the relative demand shocks of polarization amount to a shift in the occupation-specific skill prices. This has the effect that the relative returns to workers' talents change, but also that workers switch occupations depending on their observed and unobserved skills. The switching due to the unobserved skills causes selection bias in

[2] To quote the well-known survey paper by Acemoglu and Autor (2010, p78): "[...] because the allocation of workers to tasks is endogenous, the wages paid to a set of workers previously performing a given task can fall even as the wages paid to the workers now performing that task rise. [...] a regression of wages on tasks currently performed, or their change over time, would be difficult to interpret."

[3] The only difference is in labels: I call talents what Heckman and Sedlacek (1985) call skills and I call skills what they call tasks.

occupational wages, since we do not know whether rising wages in an occupation are due to a rise in the price per unit of skill in this occupation or due to a better selection of workers with respect to the unobserved component of skill.

My paper solves the selection bias problem, and it circumvents the structural estimation of the Roy model, by shifting the analysis from occupational wages to the returns to occupation-specific skills. I estimate the changes in returns to the observable component of occupation-specific skills with a two-stage procedure. First, using workers' talents I predict their propensities to enter the high-, middle-, and low-skill occupations in the period before polarization took place. Second, I estimate the changes in the returns to these propensities.[4]

In order to implement this procedure, I need two cross-sections of data with consistent measures of workers' talents that predict occupational sorting but are not influenced by occupational choice and thus not endogenous to polarization. Such data has only recently become available in the form of the National Longitudinal Survey of Youth (NLSY):[5] the NLSY cohorts of 1979 and 1997 contain detailed and multidimensional measures of talents which are hardly malleable and determined well before a worker's entry into the labor market. The measures include mathematical, verbal, and mechanical test scores as well as risky behaviors and parental education. In addition, the data are available for two representative cross-sections of 27 year olds in the end of the 1980s and the end of the 2000s, and therefore well-timed for studying polarization and the middle class squeeze.[6]

The estimation results on the returns to observable occupation-specific skills indicate a strong impact of polarization on wages. I find that a one percentage point higher propensity to enter the high- as opposed to the middle-skill occupation is associated with a .29 percent increase in wages over time. A one percentage point higher propensity to enter the low as opposed to the middle-skill occupation is associated with a .70 percent increase in wages. Workers with a high propensity to enter the middle-skill occupations in the 1980s actually suffer an absolute decline

[4]Acemoglu and Autor recommend a similar procedure but lack the data to implement it. In their words: "[...] the approach here exploits the fact that task specialization in the cross section is informative about the comparative advantage of various skill groups, and it marries this source of information to a well-specified hypothesis about how the wages of skill groups that differ in their comparative advantage should respond [...]" (Acemoglu and Autor 2010, p78)

[5]Until recently the 1997 cohort of the NLSY was too young to warrant a reliable analysis of labor market outcomes.

[6]Moreover, the data from the NLSY79 and the NLSY97 were designed to be comparable to one another.

in their expected real wages. This finding is 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.

The effect identified in these estimations is a combination of the direct demand effect of polarization on talent returns as well as the potentially heterogeneous wage gains for workers of different talents from reallocating out of the middle-skill occupations. Moreover, at age 27, the workers in the NLSY97 are young enough to have chosen their occupations when most of polarization has already taken place. Thus, the effect on their wages is likely to be largely due to ex ante different talent endowments and not due to having acquired occupation-specific experience whose value has changed ex post. This indicates a long-lasting effect of polarization on relative wages that will not fade when the current generation of workers retires.[7]

The changing returns to propensities of entering high-, middle-, and low-skill occupations may in fact be driven by alternative factors which are correlated with occupations. I address this concern by exploiting the Roy model's prediction about specific talent returns under polarization: if only occupation-specific skill prices are shifting, the Roy model implies that the change in the return to each talent solely depends on how that talent is associated with occupational choice and how the association changes over time.

I use this prediction to estimate the change in relative occupation-specific skill prices and to test the null hypothesis that all changes in returns to talents—and equivalently all changes in returns to occupational propensities—were driven by polarization. In the data, I observe each talent's initial and final association with the three classes of occupations but not the adjustment path over time. I therefore linearly interpolate the adjustment path, which gives relative price estimates that are close to the actual prices and at the same time robust to different distributions of unobserved skills. Since the NLSY provides more talents—three test scores plus the risky behaviors and other demographics—than the two unknown relative prices, I obtain over-identifying restrictions on talent returns from the model which I use to test the polarization hypothesis and to estimate the relative occupation-specific skill price changes.

The over-identifying restrictions test does not reject the polarization hypothesis

[7] It thus implies the need for long-term policy responses, e.g. long-term changes in education or tax policy instead of income support for the current generation of workers.

in the data. Moreover, the relative skill price increase in the high compared to the middle-skill occupation is precisely estimated at 20 percent, while the relative skill price increase in the low-skill occupation is imprecisely estimated with a point estimate of 31 percent. The relative skill price estimates are crucial to assess the overall impact of polarization on the wage distribution. This is because the returns to observable talents or occupational propensities alone can only explain a small part of the change in the wage distribution—just as they can only explain a small part of the variation of wages in the cross-section. In contrast, the relative occupation-specific skill price estimates change the return to the observable as well as the unobservable components of skill in each occupation and thus allow me to assess the full effect of polarization on the wage distribution.

Therefore, I compute a counterfactual wage distribution which is due to the relative occupation-specific skill price effect of polarization and compare it to the actual distribution. I do this by assigning the estimated relative skill price changes to each worker in the NLSY79 according to his occupation. It turns out that the counterfactual distribution closely matches the increase of wages at the top of the actual distribution compared to the middle. However, it fails to match the increase of wages at the bottom of the actual distribution compared to the middle. The reason is that the wage rate estimates and the dispersion of wages within occupations is so high that also many middle-earners' wages are lifted by the price changes and that some low-skill occupation workers become middle-earners themselves.

Finally, if polarization is to be the main driver of the middle class squeeze, the remaining difference at the bottom between the counterfactual and the actual change in the wage distribution must be due to the heterogeneous effect of optimal occupational switching in response to polarization on different parts of the wage distribution—a reallocation effect. Since there are no clear predictions from the Roy model about this effect, I conduct rule-of-thumb experiments to assess whether the reallocation effect may in principle explain the remainder: I assign the lowest earning workers in the middle-skill occupation in the initial period gains that they could obtain from switching to the low-skill occupation due to polarization and examine the effect that this has on the change in the lower part of the wage distribution. Experiments with a substantial gain from switching can relatively well match the wage distribution in the bottom as well as average wages in occupations. However, the assumptions that I need to make for this are strong and they are not supported by the reallocation (effect) of observable skills, which I can measure in the data.

It thus seems that, despite its strong effect on relative wages, polarization can account well for only part of the evolution of the wage distribution over the past two decades. The results therefore suggest that market-based forces may not be responsible for all of the changes in the lower half of the wage distribution. This opens the door for policy-related and institutional factors—such as de-unionization and the minimum wage—that other studies have found to have an impact on earnings at the bottom of the wage distribution over this period (Machin and Van Reenen 2008, Autor, Manning, and Smith 2010, Firpo, Fortin, and Lemieux 2011).

The findings in this study are qualitatively similar when implementing alternative definitions of occupations or tasks in occupations that have been used in the literature on polarization. These include grouping occupations according to initial median wages or average education, splitting up the large middle-skill group into blue collar and white collar occupations, and employing continuous measures of routine and nonroutine (analytical and manual) task content in occupations.

The paper continues as follows. The remainder of this section discusses the relation to the existing literature. 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), and it shows that workers sort themselves systematically into occupations according to the talent measures available in the NLSY. The Roy model and its empirical predictions are analyzed in section 3. Section 4 presents the empirical results on the returns to occupation-specific skills. Section 5 estimates the occupation-specific skill prices and tests the model, while section 6 assesses whether the resulting counterfactual wage distribution may match the actual. Section 7 concludes.

1.1.1 Related Literature

There are other studies that have tried to link job polarization to changes in the wage distribution. The most explicit effort is a recent paper by Firpo, Fortin, and Lemieux (2011) who use a Roy-style model to study the effect of shifts in the demand for tasks on occupational wages. They also carry out a decomposition to assess the effect of different factors such as occupational demands, skill supply, unionization, and minimum wages on the change in the wage distribution. Neither of these exercises control for the endogenous selection of workers with respect to unobservable skills. This limitation of Firpo, Fortin, and Lemieux (2011)'s analysis is noted by Acemoglu and Autor (2010), whose comparative advantage model predicts a changing self-selection of workers into occupations due to movements in wages rates across occupations or tasks. Exercises similar to Firpo, Fortin, and Lemieux (2011)'s that feature as part of broader papers may thus be regarded as mostly descriptive (e.g. Goos and Manning 2007, Autor, Katz, and Kearney 2008).

An alternative method to deal with endogenous selection is to employ panel data and worker fixed effects. Cortes (2012) uses data from the Panel Study of Income Dynamics to analyze the transition from middle- to high- and low-skill occupations due to polarization and its associated wage changes. Liu and Trefler (2011) similarly estimate the impact of trade in services with China and India on US workers using matched data from the Current Population Survey (see also Ebenstein, Harrison, McMillan, and Phillips 2011). Cortes finds a substantial impact of polarization on workers' wages while Liu and Trefler (2011) find a rather small impact of trade. A general difficulty with the panel data approach is the need to make an appropriate assumption about—or to control for—workers' counterfactual experience profiles of wages and occupations in the absence of polarization. Moreover, contrary to this paper and the one by Firpo, Fortin, and Lemieux (2011), these studies do not link their estimated earnings impacts of polarization to the change in the aggregate wage distribution.

The large literature on the causes of job polarization provides the hypothesis on occupational demands analyzed in my paper. During the last decade, many studies in labor economics and international trade have examined rapidly changing information and communication technology (ICT) and the off-shoring of goods and services production as causes of polarization. For example, papers that consider technological change include Autor, Levy, and Murnane (2003), Goos and Manning (2007), Michaels, Natraj, and Van Reenen (2010), Acemoglu and Autor (2010), and Autor and Dorn (2012). Papers that consider trade and offshoring include Blinder (2009), Becker, Ekholm, and Muendler (2009), Crinò (2010), Ottaviano, Peri, and Wright (2010), and Autor, Dorn, and Hanson (2012). Many of these studies find that it is largely occupations in the middle of the skill distribution that are affected by technology or trade.[8]

[8] Jaimovich and Siu (2012) find that job polarization and jobless recoveries after recessions are

My approach to linking job polarization with changes in the wage distribution relies critically on Roy (1951)'s model of occupational choice and the development of his ideas by Gronau (1974) and Heckman (e.g. Heckman 1974, Heckman and Sedlacek 1985). In particular, the mathematical specification of how occupational skills are composed of observable and unobservable worker characteristics is identical to that of Heckman and Sedlacek (1985). Gould (2002) and Mulligan and Rubinstein (2008) are the first papers to explicitly link the Roy model to increases in wage inequality and skill-biased technological change (see also Yamaguchi 2012). Compared to these papers I study the Roy model in relation to job polarization and the U-shaped change of wage inequality.

Finally, there exists a large and diverse body of literature that analyzes hypotheses about drivers of wage inequality other than polarization. The most important of those is skill-biased technological change (SBTC), which is detached from demand for specific occupations (e.g. Bound and Johnson 1992, Autor, Katz, and Krueger 1998). Hypotheses complementing that of SBTC in the top of the wage distribution have emphasized firm size and organization as well as pay increases in financial services and other professions (e.g. Garicano and Rossi-Hansberg 2004, Gabaix and Landier 2008, Tervio 2008, Philippon and Reshef 2009).[9] In terms of the developments specific to the lower part of the wage distribution, changes in policy variables and labor market institutions such as minimum wages and unionization have been prominent in the discussion (see Machin and Van Reenen 2008, Autor, Manning, and Smith 2010, Firpo, Fortin, and Lemieux 2011).

Both a falling and a rising supply of skills have been analyzed as different sets of explanations for the change in inequality. Card and Lemieux (2001) and Goldin and Katz (2008) consider a slowdown in the rate of supply of college graduates, while Lemieux (2006, p461) argues that a large part of the changes in the wage distribution that we observe is due to "composition effects linked to the secular increase in experience and education". My study is most closely related to the papers that analyze the supply of, and returns to, ability test scores (e.g. Murnane, Willett, and Levy 1995, Blau and Kahn 2005, Altonji, Bharadwaj, and Lange 2008). After

related to one another.

[9]For example, Garicano and Rossi-Hansberg (2004) show that improvements in communication technology lead to lower inequality at the bottom and higher inequality at the top of the wage distribution, and thus a squeezed middle, in a hierarchy model of endogenous firm size and organization.

Altonji, Bharadwaj, and Lange (2012, ABL), this is also the first study to analyze labor market outcomes across the two cohorts of the NLSY. While ABL 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.

1.2 Data and Empirical Facts

This section establishes the stylized facts of job polarization and the u-shape change in wage inequality in my data. Median real wages for 27 year old males rise only very little, so the other characteristic of the middle class squeeze—stagnating incomes—is also present in my data. Moreover, the section shows how workers systematically sort themselves into the occupations affected by polarization depending on their talent endowments.

1.2.1 Job Polarization and the U-Curve of Wages

I use data from the National Longitudinal Survey of Youth (NLSY) cohorts of 1979 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 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.[10] 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 (2008). Labor supply by hours worked and hourly wages are defined as in Lemieux (2006). The details of the sample construction can be found in Appendix A.1. Table 1.1 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-shape, i.e. wages

[10] At the time of writing, NLSY97 data was available up to 2009. The periods that I compare are thus 1983-1991 and 2007-2009.

increase substantially at the top of the distribution and somewhat less at the bottom but hardly at all in the middle (the middle class squeeze). 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 (2010).

I start with the stylized fact about the wage distribution in my data. Figure 1.1 graphs the empirical cumulative log wage distribution in the NLSY79 and the NLSY97 in the top two sub-figures and the change in wages by distribution quantile compared to the CPS in the bottom sub-figure. We see that the wage distribution levels and, more importantly, 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]

The second important fact is job polarization. The literature has measured high-, middle-, and low-skill occupations in different ways and arrived at the same results. It has ranked them by initial median wages or average education (e.g. Autor, Katz, and Kearney 2006, Goos and Manning 2007). Alternatively, it has grouped managerial, professional, and technical occupations as high-skill; sales, office and administrative, production, and operator and laborer occupations as middle-skill; and protective, food, cleaning and personal service occupations as low-skill (e.g. Acemoglu and Autor 2010, Cortes 2012, Jaimovich and Siu 2012).

I use the latter approach of grouping occupations in figure 1.2 and in the paper more generally for two reasons: it is becoming a standard in the literature and it explicitly delineates occupations by the extent of abstract (high-skill), routine (middle-skill), and manual (low-skill) tasks that they require (see Acemoglu and Autor 2010, Jaimovich and Siu 2012). The upper two sub-figures graph the employment shares in the three occupation groups for the NLSY79 and NLSY97 and compared to the CPS. The share of employment in the middle-skill occupations is declining while the share of employment in the high- and the low-skill occupations is rising. This can be seen more clearly in the lower sub-figure, which plots the changes in employment shares. These facts establish job polarization in the NLSY, which is very similar to what can be found for 27 year olds in the CPS. The findings are

[11] 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 2010). 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, are likely to be larger at older ages.

the same if I use the alternative approaches of grouping occupations as low, middle, and high-skilled.

Before moving on, figure 1.3 shows average 1979 real wages in high-, middle-, and low-skill occupations and how they have changed over the two cohorts in the NLSY and, for comparison again, the CPS. Unsurprisingly, average wages in high-skill occupations are higher than in middle-skill occupations, which in turn are higher than average wages in low-skill occupations. The changes are more interesting. 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.[12] 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.[13] Appropriately adjusting for this effect is the main contribution of my paper.

1.2.2 Talent Sorting into Occupations

Workers do not choose to work in the high-, middle-, and low-skill occupations at random. This section uses choice regressions to establish and quantify systematic occupational sorting in the data.

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 choice and wages, that may approximate different dimensions of skill, and that can be compared over the two cohorts.[14]

[12]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 NLSY79 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.

[13]Also other studies find a further decrease in low-skill compared to middle-skill wages (Goos and Manning 2007). Autor and Dorn (2012) find that relative wages in clerical occupations rise while quantities fall.

[14]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.

Table 1.2 reports labor force averages of NLSY variables that fulfill the four criteria ("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.[15]

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. 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 1.2 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 1.3 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.

[15]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.

Sorting into Occupations

Figure 1.4 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 1.4 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 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. 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 1.3.

To test the idea of sorting according to comparative advantage I run multinomial choice regressions. Let $\{K_{it}\}$ be a set of indicator variables that take the value of 1 when individual i works in occupation $K \epsilon \{L, M, H\}$ 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. For now, I model the conditional choice probabilities as multinomial logit (MNL):[16]

$$p(K_{it} = 1|x_{it}, \pi_t) = \frac{exp(b_{K0t} + b_{K1t}x_{1it} + ... + b_{KJt}x_{Jit})}{\sum_{G=H,M,L} exp(b_{G0t} + b_{G1t}x_{1it} + ... + b_{GJt}x_{Jit})}. \tag{1.1}$$

Maximum likelihood estimation of equation (1.1) yields the coefficients of this model and it provides conditional probabilities ("propensities") to enter each occupation based on the observable talents. As I show in the next section, 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 1.4 reports the results from the multinomial choice regressions. These extract the marginal effect of another 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. On the other hand, 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

[16] This is a commonly made modeling decision because the MNL is convenient to work with. For example, the relative risk of choosing occupation K rather than the base category M becomes

$$log\left[\frac{p(K_{it} = 1)}{p(M_{it} = 1)}\right] = (b_{K0t} - b_{M0t}) + (b_{K1t} - b_{M1t})x_{1it} + ... + (b_{KJt} - b_{MJt})x_{Jit}.$$

Using a multinomial probit (MNP) model with uncorrelated disturbances across options instead of the MNL would have been a natural choice, too. Although more difficult to interpret, the MNP has the attraction of being motivated by a latent normal random vector. Empirically, there is often little difference between the predicted probabilities from probit and logit models (see Cameron and Trivedi 2005, p489ff) and in particular my results are robust to using the MNP. Both, the MNL and the MNP, invoke an Independence of Irrelevant Alternatives (IIA) assumption (i.e. uncorrelated errors) which is too restrictive if one wants to interpret the regression coefficients as structural parameters of an economic model.

middle, and those workers with low math and mechanical skills moving into the low-skill occupations. 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 1.4 and in column three of table 1.4.

Finally, the regressions in columns two and four of table 1.4 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-shape in the change in demand for skill levels. Moreover, normalized measures of illicit activities and engagement in precocious sex are added. The regressions omit parental education because is 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.

1.3 Theory and Econometric Methods

On the one hand, as explained in the introduction, the large body of research on job polarization indicates that the drop of employment in the middle-skill occupations is due to a decrease in demand. On the other hand, the empirical analysis of occupational choice shows that there is systematic sorting with respect to talents in the NLSY data. This naturally motivates a Roy model of occupational choice in order to analyze the effect of demand changes on the supply side.

In this model, a given worker i chooses the occupation that offers him the highest log wage:

$$w_{it} = max\{w_{Hit}, w_{Mit}, w_{Lit}\}, \tag{1.2}$$

where $\{H, M, L\}$ indexes the high-, middle-, and low-skill occupation, respectively. 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. The w_{Kit}s with $K \epsilon \{H, M, L\}$ can more generally be utility levels in each occupation.

As seen above, the NLSY provides a multidimensional array of relevant talent proxies for each respondent. Thus, the log occupational wages can be written as a sum of log prices and quantities of occupation-specific skills in the following way:

$$w_{Kit} = \pi_{Kt} + s_{Kit} = \pi_{Kt} + \beta_{K0} + \beta_{K1}x_{1it} + ... + \beta_{KJ}x_{Jit} + u_{Kit}, \tag{1.3}$$

where π_{Kt} is the price per unit of skill in occupation K, s_{Kit} individual i's specific skill in occupation K, $x_{it} = [x_{1it}, ..., x_{jit}, ..., x_{Jit}]'$ are the observed talents, the β_{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. This linear factor formulation is adopted from Heckman and Sedlacek (1985).[17]

The demand side hypothesis about job polarization in terms of this model is therefore

$$\triangle(\pi_H - \pi_M) > 0 \text{ and } \triangle(\pi_L - \pi_M) > 0, \tag{1.4}$$

i.e. the relative occupation-specific skill price in the middle falls compared to the high- and the low-skill occupation. The polarization hypothesis examined in the following has two components: first, that equation (1.4) is true, and second, that it is the reason for the U-shape change in the wage distribution.

The assumption that it is the occupation-specific skill prices that are changing under polarization is crucial. This is in fact the same as in much of the existing literature on job polarization, which models the effect of shifting demand for tasks or occupations on labor supply via changing wage rates. For example, the driving force on the labor market in the original papers of Autor, Levy, and Murnane (2003) and Autor, Katz, and Kearney (2006) is a drop in the relative wage rate for the routine task due to computerization. In Acemoglu and Autor (2010), similar to this paper, the authors analyze how technological change and offshoring alter wages and worker sorting via the relative price of the tasks corresponding to low-, middle-, and high-skill occupations.[18]

In the following I analyze whether the simple assumption of shifting occupation-specific skill prices may get us all the way to explaining the change in the wage distribution over the last decades. Since the theoretical argument and explanation of empirical methods is rather involved and the general case requires complex notation,

[17] Contrary to Roy (1951) or Heckman and Sedlacek (1985) I will not make a distributional assumption on the unobserved component of skill in the following. Moreover, the primary interest is not in the sectoral distribution of skills and wages, but in changes in returns to occupation-specific skills.

[18] Other papers that make essentially the same assumption include Cortes (2012) and Liu and Trefler (2011).

I use a maximally simplified version of the model for the rest of this section. The results can be extended to the general case for the empirical analysis.

1.3.1 A Simplified Model

In order to strip the model of equations (1.2)-(1.4) to its essence, assume there are only two occupations, middle M and nonmiddle N, with $\triangle(\pi_N - \pi_M) > 0$ according to the polarization hypothesis. Moreover, there is only one observable talent x_i with mean zero ($E(x_i) = 0$) and variance one ($Var(x_i) = 1$), and β_{K0} is zero. I indicate the difference between N and M sector 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 π_{Nt} and π_{Mt} and their functions. Wages in occupations $K \epsilon \{N, M\}$ become:

$$w_{Kit} = \pi_{Kt} + s_{Ki} = \pi_{Kt} + \beta_K x_i + u_{Ki} \tag{1.5}$$

For intuition, we can think of x_i as math talent where a high value is associated with the non-middle occupation and higher wages in the initial period.

How do the workers who have a comparative advantage in the middle occupation fare 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

$$p_N(x_i, \tilde{\pi}_t) \equiv p(N_{it} = 1|x_i) = Pr(\tilde{u}_i > -(\tilde{\pi}_t + \tilde{\beta} x_i))$$

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

$$E(w_{Kit}|x_i, K_{it}) = \pi_{Kt} + \beta_K x_i + E(u_{Ki}|x_i, K_{it} = 1)$$

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

$$\begin{aligned} \triangle E(w_{it}|x_i) = {} & \triangle \pi_M + p_N(x_i, \tilde{\pi}_0) \triangle(\pi_N - \pi_M) + \\ & + \triangle p_N(x_i, \tilde{\pi}_t) \left[E(w_{Ni0}|x_i, N_{it} = 1) - E(w_{Mi0}|x_i, N_{it} = 0) \right] + \\ & + p_N(x_i, \tilde{\pi}_0) \triangle E(s_{Ni}|x_i, N_{it} = 1) + p_M(x_i, \tilde{\pi}_0) \triangle E(s_{Mi}|x_i, N_{it} = 0) \end{aligned} \tag{1.6}$$

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 $\triangle p_N(x_i, \tilde{\pi}_t) \geq 0$), 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 an assumption 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.[19] One way to evaluate the average effect of polarization on workers of different observable talents would thus be to assume the normal distribution and structurally estimate the Roy model in the NLSY79 and the NLSY97 cross-section, respectively. However, 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

[19]Even with a distributional assumption, say normality, $E(w_{Kit}|x_{it}, y_{Kit} = 1)$ 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 (2012) use an extreme value distribution to solve the problem, but this comes at the cost of the very strong assumption that individuals' skills are uncorrelated across occupations.

form assumption for the skill distribution.

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 π_{Kt}s and then applying it beyond the margin. Consider the change in worker i's wages for a marginal shift in prices:

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

where d denotes a marginal change. Thus, due to the optimality of workers' occupational choice and the envelope theorem, the effect on wages of a marginal change in π_{Kt}s is only the direct price effect

$$dE(w_{it}|x_i) = d\pi_M + p_N(x_i, \tilde{\pi}_t)d(\pi_N - \pi_M). \tag{1.7}$$

According to prediction (1.7), under the polarization hypothesis, workers who are ceteris paribus more likely to enter the nonmiddle occupation are expected to see their relative wages increase. For example, randomly picking two workers from the population, the worker with lower math talent (call him $\bar{m}$ with $x_{\bar{m}} = \bar{m}$) will *on expectation* have a lower wage increase under polarization than the worker with higher math talent (call him m with $x_m = m$) because $p_N(x_{\bar{m}}, \tilde{\pi}_t) < p_N(x_m, \tilde{\pi}_t)$ and $d(\pi_N - \pi_M) > 0$. The nice feature about this result on the margin is that it is solely in terms of variables that I can straightforwardly estimate from the information on wages, occupational choice and my observables, i.e. $E(w_{it}|x_i)$ and $p_N(x_i, \tilde{\pi}_t)$, and parameters that I have hypotheses about or that I want to estimate, i.e. $d(\pi_N - \pi_M) = d\tilde{\pi}_t$.

Prediction (1.7) also holds qualitatively beyond the margin, i.e. 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 π_0 to π_1. Hence, we can integrate prediction (1.7) from $t = 0$ to $t = 1$ to obtain:

$$E(w_{i1}|x_i) - E(w_{i0}|x_i) = \triangle\pi_M + \int_{\tilde{\pi}_0}^{\tilde{\pi}_1} p_N(x_i, \tilde{\pi}_t)d\tilde{\pi}_t, \tag{1.8}$$

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. In terms of the example, if $p_N(x_{\bar{m}}, \tilde{\pi}_0) < p_N(x_m, \tilde{\pi}_0)$ then $p_N(x_{\bar{m}}, \tilde{\pi}_t) \leq p_N(x_m, \tilde{\pi}_t)$ for all $t\epsilon(0, 1]$. Therefore, we expect a higher increase in wages for worker m than for worker $\bar{m}$.[20]

In section 1.4 I estimate the change in wages associated with $p_N(x_i, \tilde{\pi}_0)$ between the NLSY79 and NLSY97. Because of prediction (1.7), I expect the return per unit of $p_N(x_i, \tilde{\pi}_0)$ to increase over time. Note, though, that this return change includes the direct price effect and the reallocation effect discussed in equation (1.6). In terms of the example, the expected wage increase for worker m versus worker $\bar{m}$ between $t = 0$ and $t = 1$ includes the initial difference in propensities $p_N(x_m, \tilde{\pi}_0) - p_N(x_{\bar{m}}, \tilde{\pi}_0)$ and the change in this difference along the adjustment path. The identification assumption in my data is that the distribution of unobservable skill components conditional on x_i is the same across the NLSY79 and the NLSY97, i.e. that a given value of math talent measures on average the same person in both cohorts. Section 1.4 explains the details.

1.3.2 Identifying the Change in the Occupation-Specific Skill Prices

The second and more difficult question is to identify the actual changes in relative prices $\triangle(\pi_{Nt} - \pi_{Mt}) = \triangle\tilde{\pi}_t$. One way or another I will have to make an additional assumption for this and I argue that my approach of choice is particularly attractive for several reasons.

The overall change in worker i's expected wage is the sum over his marginal expected wage changes along the adjustment path from π_0 to π_1 as shown in equation

[20] Another way of deriving equation (1.8) is illustrative: Concentrate on a specific worker i first and note again that $\tilde{\pi}_t \equiv \pi_{Nt} - \pi_{Mt}$, $\triangle\tilde{\pi}_t > 0$, 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 $\tilde{\pi}^i_t \equiv -\tilde{s}_i = -(s_{Ni} - s_{Mi})$, we get:

$$\begin{aligned} w_{i1} - w_{i0} &= \triangle\pi_M + N_{i1}(w_{Ni1} - w_{Mi1}) - N_{i0}(w_{Ni0} - w_{Mi0}) \\ &= \triangle\pi_M + \begin{cases} \triangle\pi_N - \triangle\pi_M = \tilde{\pi}_1 - \tilde{\pi}_0 & \text{if } N_{i0} = 1, N_{i1} = 1 \\ \tilde{\pi}_1 + \tilde{s}_i = \tilde{\pi}_1 - \tilde{\pi}^i_t & \text{if } N_{i0} = 0, N_{i1} = 1 \\ 0 & \text{if } N_{i0} = 0, N_{i1} = 0 \end{cases} \\ &= \triangle\pi_M + \int_{\tilde{\pi}_0}^{\tilde{\pi}_1} N_{it} d\tilde{\pi}_t. \end{aligned}$$

Taking expectations w.r.t. $\tilde{u}_i$ conditional on x_i on the top left and bottom of this equation gives result (1.8). Hence, since within occupations the wage gain is constant, the overall gain for a specific worker depends solely on the "distance" of the adjustment that the worker is still in the middle $(\pi^i_N - \pi_{N0})$ and already in the nonmiddle $(\pi_{N1} - \pi^i_N)$ occupation. This principle is the same for expected wages and probabilities of being in the nonmiddle occupation.

(1.8). In this equation, I want to estimate $\triangle\tilde{\pi}_t$ and possibly $\triangle\pi_M$. I know $E(w_{t1}|x_i)$ 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 nicely illustrated in a graph. In figure 1.5, 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.

The first assumption that comes to mind is to simply assume 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 (1.8). 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 (1.7) holding exactly for the discrete price change as well and the regression in section 1.4 on the propensity $p_N(x_i, \tilde{\pi}_0)$ directly identifying the price change. Of course, this is not a good assumption.

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 terms of the example it is to assume that the probability change for the high math worker m is the same as for the low math worker $\bar{m}$, i.e. $\triangle p_N(x_{\bar{m}}, \tilde{\pi}_t) = \triangle p_N(x_m, \tilde{\pi}_t)$. In this case, equation (1.8) becomes

$$E(w_{i1}|x_i) - E(w_{i0}|x_i) = \triangle\pi_M + const + p_N(x_i, \tilde{\pi}_0)\triangle\tilde{\pi}_t.^{21}$$

Again the regression on $p_N(x_i, \tilde{\pi}_0)$ in section 1.4 directly identifies the price change. This is also not a good assumption as it does not allow for a differential reallocation

[21]Suppose $\frac{\partial p_N(x_i,\tilde{\pi}_t)}{\partial\tilde{\pi}_t} = F'(\tilde{\pi}_t) \geq 0$. Then $p_N(x_i, \tilde{\pi}_t) = p_N(x_i, \tilde{\pi}_0) + F(\tilde{\pi}_t) - F(\tilde{\pi}_0)$ and

$$const = \int_{\tilde{\pi}_0}^{\tilde{\pi}_1} [F(\tilde{\pi}_t) - F(\tilde{\pi}_0)]d\tilde{\pi}_t.$$

effect across worker groups, e.g. that the low math worker $\bar{m}$ may be able to reallocate out of the middle to a larger extent than the high math worker m because the latter is more likely in the nonmiddle to start with. In figure 1.5 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).

A seemingly attractive alternative would be to assume that $\tilde{u}_i$ is normally distributed (for simplicity assume $\tilde{\sigma} = 1$), which modifies (1.8) to

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

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.[22] But then I am estimating the price change by relying on a distributional assumption in (1.8) 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

$$p_N(x_i, \tilde{\pi}_t) \approx p_N(x_i, \tilde{\pi}_0) + \frac{p_N(x_i, \tilde{\pi}_1) - p_N(x_i, \tilde{\pi}_0)}{\tilde{\pi}_1 - \tilde{\pi}_0}(\tilde{\pi}_t - \tilde{\pi}_0). \tag{1.9}$$

In figure 1.5, this amounts to approximating $p_N(x_i, \tilde{\pi}_t)$ as the y-coordinate for the point on the line $\overline{A_1A_2}$ that corresponds to $\tilde{\pi}_t$ 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 $\tilde{\pi}_t \epsilon (\tilde{\pi}_0, \tilde{\pi}_1)$ is not too convex or concave, the approximation should be reasonably close. Whether it is sufficiently accurate will be tested below.

Equation (1.8) now becomes

$$E(w_{i1}|x_i) - E(w_{i0}|x_i) = \triangle\pi_M + \frac{p_N(x_i, \tilde{\pi}_1) + p_N(x_i, \tilde{\pi}_0)}{2}\triangle(\pi_N - \pi_M). \tag{1.10}$$

This is one equation in two unknowns. However, as it holds for all x_i, I could for example identify $\triangle(\pi_N - \pi_M)$ and $\triangle\pi_M$ by imposing it for workers with high and

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

low math talent m and $\bar{m}$, respectively.

A more attractive way to estimate $\triangle(\pi_N - \pi_M)$ is to multiply both sides of equation (1.10) by x_i and taking expectations. By the law of iterated expectations, this results in

$$cov(w_{i1}, x_i) - cov(w_{i0}, x_i) = \frac{cov(N_{i1}, x_i) + cov(N_{i0}, x_i))}{2} \triangle(\pi_N - \pi_M), \qquad (1.11)$$

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.

If I had just one talent as in this simple example, I could exactly solve equation (1.11). Yet, as I have J different talents in my empirical implementation, prediction (1.11) has to hold for each single one of them so that I get J different moment conditions

$$m_j(\triangle\tilde{\pi}_t) = cov(w_{i1}, x_i) - cov(w_{i0}, x_i) - \frac{cov(N_{i1}, x_i) + cov(N_{i0}, x_i)}{2} \triangle(\pi_N - \pi_M) = 0$$

from the model. I can stack those moment conditions in a column vector and apply the minimum distance estimator for $\triangle\tilde{\pi}_t$ which minimizes the quadratic form:

$$m(\triangle\tilde{\pi}_t)' W m(\triangle\tilde{\pi}_t), \qquad (1.12)$$

where the asymptotically optimal W takes into account the variance-covariance matrix of the first-stage estimates of $cov(w_{it}, x_i)$ and $cov(N_{it}, x_i)$. The objective function (1.12) in optimum also provides a test statistic for the joint test of the polarization hypothesis and my linear approximation of the reallocation adjustment path. Section 1.5 details and implements this estimation and testing procedure in the more general case of three occupations in my data.

Overall, the procedure of estimating the relative price changes described here has two 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 be robust to different distributions of unobserved skills, and it is transparent and easy to implement.

1.4 Polarization's Effect on Observable Skills

How do the workers who have a comparative advantage in the high-, middle-, and low-skill occupation fare over time? This section analyzes the effect of polarization on the returns to propensities to enter occupations and on absolute skill measures.

1.4.1 Prediction

Predictions (1.7) and (1.8) generalize to the three-occupation case (for detailed derivation see Appendix A.2):

$$dE(w_{it}|x_{it}) = d\pi_{Mt} + p_H(x_{it},\pi_t)d(\pi_{Ht}-\pi_{Mt}) + p_L(x_{it},\pi_t)d(\pi_{Lt}-\pi_{Mt}), \tag{1.13}$$

and

$$\begin{aligned} E(w_{i1}|x_{i1}) - E(w_{i0}|x_{i0}) = \triangle\pi_M + \int_{\pi_{H0}-\pi_{M0}}^{\pi_{H1}-\pi_{M1}} p_H(x_{it},\pi_t)d(\pi_{Ht}-\pi_{Mt}) + \\ + \int_{\pi_{L0}-\pi_{M0}}^{\pi_{L1}-\pi_{M1}} p_L(x_{it},\pi_t)d(\pi_{Lt}-\pi_{Mt}). \end{aligned} \tag{1.14}$$

where $p_K(x_{it},\pi_t)$ is the probability of working in occupation $K\epsilon\{H,M,L\}$ under the price vector π_t. Moreover, I now give a time subscript to the observable characteristics to indicate which dataset they are from.

Hence, under the polarization hypothesis (1.4), workers who are ceteris paribus more likely to enter the high- and the low-skill occupation are expected to see their relative wages increase. In order to evaluate this, I estimate ordinary least squares (OLS) regressions for pooled data of the form

$$\begin{aligned} w_{it} = \alpha_0 + \alpha_1 p_H(x_{it},\pi_0) + \alpha_2 p_L(x_{it},\pi_0) + \alpha_3 \times NLSY97 + \\ + \alpha_4 p_H(x_{it},\pi_0) \times NLSY97 + \alpha_5 p_L(x_{it},\pi_0) \times NLSY97 + \varepsilon_{it}, \end{aligned} \tag{1.15}$$

where NLSY97 is a dummy for whether a particular observation is from the NLSY97 and $p_H(x_{it},\pi_0)$ and $p_L(x_{it},\pi_0)$ are the probabilities to choose the high- and the low-skill occupation in the NLSY79, i.e. under the old prices. Hence, the approach is to hold groups of workers constant over time in terms of their predicted occupation-specific skills (the probabilities) and study their average wages over time. According to prediction (1.13), I expect the parameter estimates for α_4 and α_5 to be positive.[23] There are no predictions from the theory on α_1, α_1 and α_3, although one would think that a higher probability to enter the H and the L occupation is associated with higher and lower wages, respectively.

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. The parameter estimates are then used to predict $p_H(x_{it}, \pi_0)$ and $p_L(x_{it}, \pi_0)$ for each individual in the NLSY79 and the NLSY97. This makes the estimation of (1.15) a two-step procedure. In fact, I am using two-step estimation procedures throughout this paper since my empirical strategy exploits measuring comparative advantage in occupations with respect to observable talents and then relating this comparative advantage to changes in the returns to talents:

> "[...] the approach here exploits the fact that task specialization in the cross section is informative about the comparative advantage of various skill groups, and it marries this source of information to a well-specified hypothesis about how the wages of skill groups that differ in their comparative advantage should respond [...]"

These are the words of Acemoglu and Autor (2010, p78) who suggest the same procedure in their well-known survey paper but lack the data that I have to implement it satisfactorily.

In terms of the two-step procedure used here, two clarifications are in order. First, different functional form assumptions can be used to specify $p_K(x_{it}, \pi_0)$. 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 and below zero, i.e. 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 1.4 in the following but my results do not change if I use the other options to specify $p_K(x_{it}, \pi_0)$.

Second, the standard errors in the second stage regression (1.15) have to reflect the fact that $p_H(x_{it}, \pi_0)$ and $p_L(x_{it}, \pi_0)$ are estimates and thus possess sampling

[23] In general, regression (1.15) provides the best linear predictor of

$$\triangle E\left(w_{it}|p_H(x_{it}, \pi_0), p_L(x_{it}, \pi_0)\right) = \alpha_3 + \alpha_4 p_H(x_{it}, \pi_0) + \alpha_5 p_L(x_{it}, \pi_0).$$

variation. Among others, Murphy and Topel (1985) provide a procedure to do this, which is however somewhat tedious.[24] Therefore I report bootstrapped standard errors instead, which are also asymptotically consistent.

Note that, although they identify the average relative wage changes for workers of different observables x_{it} due to polarization, the parameter estimates for α_4 and α_5 do not identify the structural relative price changes $\triangle(\pi_H - \pi_M)$ and $\triangle(\pi_L - \pi_M)$. This is because, as we have seen in equation (1.6), the conditional wage changes for different x_{it} consist of a combination of the direct price effect and a reallocation effect. As discussed at length in section 1.3.2, the latter may differ across worker groups, while interpreting α_4 and α_5 as the relative price changes would impose that it is the same across x_{it}.

1.4.2 Results

Table 1.5 reports the results from wage regressions a la (1.15) on the propensities to enter the high- and the low-skill occupation in the NLSY79 and the NLSY97. As expected, 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.

The prediction from polarization in equation (1.13) is however about changes in 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 -1.65 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 effect by the respective standard deviations of the propensities, 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

[24] 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.

compared to a 5.9 percent higher and 8.4 percent lower wage in the NLSY79.[25]

For illustration of the effect of different propensities to enter the three occupations, figure 1.6 plots the predictions from linear wage regressions on each propensity at a time together with their probability densities.[26] 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 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 very 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 (1.15) is based on the assumption that for a given vector of talents x_{it} workers are in expectation the same in terms of their relative labor market productivities over the two cohorts. Tables 1.2 and 1.3 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. Consequently, unreported descriptive statistics show 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 my observable measures has not changed over the two cohorts. Combined, these pieces of evidence lend substantial support to my identification assumption.[27]

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 workers in the

[25]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 19.0 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.

[26]The coefficients and standard errors from these wage regressions on each propensity separately are not reported in a table for saving space.

[27]The racial distribution does however change over the cohorts. Therefore, I control for race in all my analyses.

NLSY97 entered the labor market only recently when the bulk of the occupational demand change had likely already taken place. Hence, polarization's effect on their relative wages mostly reflects returns changes to ex ante relative talent differences and not to skills that they acquired in a specific occupation. Identifying that there exists a substantial ex ante effect is relevant for policy makers as it implies that the relative earnings effects of polarization will not fade over time and that temporary policy responses are therefore not sufficient.

The result in column one of table 1.5 does not exclude the possible influence of other factors than polarization on wages of workers with comparative advantage in the high- or the low-skill occupation. In particular, skill-biased technological change 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, comparative advantage in occupations is not important because returns to skills change across the board. The SBTC amounts to $d\beta_{Kj} = d\beta_j$ in my framework and it is easily incorporated in prediction (1.13) in addition to polarization:[28]

$$dE(w_{it}|x_{it}, \pi_t) = d\pi_M + p_H(x_{it}, \pi_t)d(\pi_H - \pi_M) + p_L(x_{it}, \pi_t)d(\pi_L - \pi_M) + \\ + d\beta_0 + d\beta_1 x_{1it} + ... + d\beta_J x_{Jit}$$

When allowing for SBTC with all the talents included on top of polarization, the identification will have to rely on the functional form of $p_H(x_{it}, \pi_t)$ and $p_L(x_{it}, \pi_t)$, 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 1.5 assess the potential importance of the SBTC hypothesis versus polarization. Column two adds a dummy of whether the individual completed a four-year college or more to the regression. 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,

[28] Actually, SBTC may predict that the return to $p_L(x_{it}, \pi_t)$ falls instead of rises.

some college, and at least four year college) in column three.[29] 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 very 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.

How much of the U-shape change in the wage distribution can the changing returns to observable skills explain? Figure 1.7 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 1.5 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 are large share of the change in the wage distribution. This is not surprising since the observables also only explain a relatively small share of variation in wages in the cross-section. The remainder should thus be explained by changes in returns to unobservable occupational skills (u_{Kit} in the notation of the model).

To sum up, I conclude that the results reported in this section indicate a substantial longterm decline in relative wages of workers with comparative advantage in the middle-skill occupation. Moreover, the driver of this decline is more likely to be relative demand changes for occupations as implied by the polarization hypothesis than increases in absolute returns to skills that are detached from comparative advantage (SBTC). Nonetheless, the analysis so far remains unsatisfactory in two dimensions: it does not formally exclude other drivers of skill returns than polarization and, because a substantial part of skill is unobserved, the changing returns to observable talents can naturally only hope to match part of the wage distribution.

[29] 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.

The next sections tackle these two shortcomings.

1.5 Estimating the Change in Occupation-Specific Skill Prices

The last section provided convincing evidence for polarization to have driven workers' skill returns over the past two decades. In this section, I formally test whether the polarization model can explain the whole variation in observable skill returns in the NLSY via a test of over-identifying restrictions. The procedure yields an estimate of the implied change in occupation-specific skill prices. In the next section, I use this estimate to explore how much of the U-shape change of wage inequality can be explained by relative price changes across occupations and the potential role of reallocation to explain the rest.

1.5.1 Methodology

A more detailed assessment of the effect of polarization looks at each talent in turn. I use the fact that I observe $x_{it} = [x_{1it}, ..., x_{Jit}]'$ and that individuals have comparative advantages in occupations varying with each x_{jit} in order to over-identifying restrictions from the polarization hypothesis. The intuition is that the return to a talent should change depending on which occupational choice it predicts and how that changes.

Linearly approximating the probabilities under the integral in prediction (1.14) as discussed in relation to figure 1.5 (see also Appendix A.2), and writing in terms of regression coefficients gives:

$$\triangle\gamma_j = \frac{\delta_{Hj0} + \delta_{Hj1}}{2}\triangle(\pi_H - \pi_M) + \frac{\delta_{Lj0} + \delta_{Lj1}}{2}\triangle(\pi_L - \pi_M), \tag{1.16}$$

where $\delta_{Kjt} = \frac{cov(K_{it}, x_{jit})}{var(x_{jit})}$ with $\delta_{Hjt} + \delta_{Mjt} + \delta_{Ljt} = 0$, K_{it} is an indicator for working in occupation K, and $\gamma_{jt} = \frac{cov(w_{it}, x_{jit})}{var(x_{jit})}$. These parameters can be recovered from OLS allocation

$$K_{it} = \delta_{K0t} + \delta_{K1t}x_{1it} + \delta_{K2t}x_{2it} + ... + \delta_{KJt}x_{Jit} + v_{Kit} \tag{1.17}$$

and wage regressions

$$w_{it} = \gamma_{0t} + \gamma_1 x_{1it} + \gamma_2 x_{2it} + ... + \gamma_J x_{Jit} + u_{it}.^{30} \tag{1.18}$$

Therefore, result (1.16) provides a simple to implement procedure to assess polarization's effect on the returns to detailed talents. I have data on individuals' talents, their choices of entering high, middle, or low-skill occupations, and their wages in the periods before ($t = 0$) and after ($t = 1$) polarization took place. First, I run four allocation regressions (1.17) for $K = H$ and $K = L$ in $t = 0$ and $t = 1$, which recover the partial correlations of the observed talents and occupational choices δ_{Kjt}. Second, I run two wage regressions (1.18) for $t = 0$ and $t = 1$, which recover the partial correlations of the observed talents and wages γ_{jt} in each period. Then, according to condition (1.16), the change of a talent's effect on the wage equals its effect in the allocation regressions times the change in relative prices.[31]

Condition (1.16) is in fact very intuitive. The return to a talent x_{jit} should change by the extent to which, conditional on the other talents, it increases the probability to work in occupations H and L, i.e. δ_{Hj0} and δ_{Lj0}, and the extent to which this association changes, i.e. (δ_{Hj1}-δ_{Hj0}) and (δ_{Lj1}-δ_{Lj0}).

In order to assess the validity of the polarization hypothesis in the data, one could thus simply check whether the returns changes to individual talents line up with what their allocation coefficients imply. However, a more encompassing test of the model recognizes that condition (1.16) has to hold for all J talents at the same time. Thus, as long as there are more talents than the two unknown model parameters $\triangle(\pi_H - \pi_M)$ and$\triangle(\pi_L - \pi_M)$, I can use the over-identifying restrictions implied in (1.16) to devise an overall test of the model.

The first step in such a test is to implement a minimum distance estimator for the implied relative wage rate changes. Define $\bar{\delta}_{Kj} \equiv \frac{\delta_{Kj0}+\delta_{Kj1}}{2}$, and stack $\triangle\gamma_j$ and $\bar{\delta}_{Kj}$ into $J \times 1$ vectors. Then, using the first stage estimates $\hat{\triangle\gamma}$ and $\hat{\bar{\delta}}_K$ and defining the $J \times 1$ vector $m(\triangle\pi) = \hat{\triangle\gamma} - \hat{\bar{\delta}}_H \triangle(\pi_H - \pi_M) - \hat{\bar{\delta}}_L \triangle(\pi_L - \pi_M)$, this estimator

[30] To be exact, the allocation and wage regressions in fact recover the covariance of K_{it} and w_{it} with the residual of regressing x_{jit} on the other observable talents. This is what I use in the following.

[31] 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.

minimizes

$$Q(\triangle\pi) = m(\triangle\pi)'Wm(\triangle\pi) \tag{1.19}$$

with respect to $\triangle(\pi_H - \pi_M)$ and $\triangle(\pi_L - \pi_M)$. Depending on the weighting matrix 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 $W = [Var(m(\triangle\pi))]^{-1}$, and the Diagonally Weighted Minimum Distance (DWMD) estimator if $W = [diag(Var(m(\triangle\pi)))]^{-1}$. The EWMD can be implemented by a simple OLS regression of $\hat{\triangle\gamma}$ on $\hat{\delta}_{Ht}$ and $\hat{\delta}_{Lt}$, the OMD by a (feasible) GLS regression, and the DWMD by weighted least squares.

Just as GLS the OMD is asymptotically optimal and it yields consistent estimates of the relative price changes $\triangle(\pi_H - \pi_M)$ and $\triangle(\pi_L - \pi_M)$. Moreover, the objective function (1.19) in optimum can be shown to be asymptotically chi-squared distributed with $J - 2$ degrees of freedom:

$$Q(\hat{\triangle\pi}) = m(\hat{\triangle\pi})'[Var(m(\hat{\triangle\pi}))]^{-1}m(\hat{\triangle\pi}) \overset{a}{\sim} \chi^2(J-2)$$

This provides me with an overall test of the cross-equation restrictions implied by the model.

Finally, Altonji and Segal (1996) and Pischke (1995) present evidence for potential bias of the OMD in small samples and recommend using the EWMD and the DWMD in addition, respectively. I thus report results for these two estimators as well. For more details of how I implement the minimum distance estimation and test, please refer to Appendix A.3.

Given optimal worker reallocation, the implied absolute wage change in the middle-skill occupation π_M can be bounded: under the initial prices, the initial worker allocation has to (weakly) dominate the new allocation and vice versa under the new prices. A natural approach is to impose this for average wages. Thus, $\triangle\pi_M$ has to be such that

$$\triangle E(w_{it}) \geq \triangle\pi_M + p_H(\pi_0)\triangle(\pi_H - \pi_M) + p_L(\pi_0)\triangle(\pi_L - \pi_M)$$

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

$$\triangle E(w_{it}) \leq \triangle\pi_M + p_H(\pi_1)\triangle(\pi_H - \pi_M) + p_L(\pi_1)\triangle(\pi_L - \pi_M)$$

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 $\triangle E(w_{it})$, $p_H(\pi_t)$, 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 $\triangle\pi_M$.

Finally, by assigning the estimated price changes to the workers in the NLSY79 and comparing the resulting change in the counterfactual wage distribution to the actual one, I can assess what the contribution of changes in occupational prices is to the overall change in the wage distribution. I also assess what share of the remainder may be due to reallocation.[32]

1.5.2 Empirical Results

Table 1.6 reports the reduced form allocation and wage regressions according to equations (1.17) and (1.18). 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 1.4. However, contrary to the MNL, the OLS coefficients for each occupation in table 1.6 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 occupation 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 1.6. 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 (1.16). 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.

[32] In fact, the model does allow for a change in the population supply of talents to play a part. As we saw in table 1.2, this is however minuscule in the data.

Overall, thus, the results from table 1.6 are neither clearly in favor of- nor against the polarization hypothesis. The formal test of the restrictions implied by prediction (1.16) across all talents may therefore be quite informative. Table 1.7 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 $\triangle(\pi_H - \pi_M)$ and $\triangle(\pi_L - \pi_M)$ 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 at 2.4 percent. The p-value of the hypothesis test is at 10.7 percent and thus the model is not rejected at conventional significance levels. Furthermore, the estimates for $\triangle(\pi_H - \pi_M)$ 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, $\triangle(\pi_L - \pi_M)$ is imprecisely estimated and it actually drops to negative point estimates in the EWMD and the DWMD.

With this caveat in mind, I use the price estimates from the OMD to evaluate what share of the overall change in the wage distribution is due to the occupation-specific skill prices in the next section.

1.6 Matching the Change in the Wage Distribution

In this last section, I assess whether the polarization model can in principle account for the change in the overall wage distribution.

First, I use the price estimates from the OMD to evaluate what share of the change in inequality is due to the occupation-specific skill prices. I obtain the skill price effect by assigning the price changes to the workers in the initial period. According to the model, the remaining differences between the actual and the counterfactual wage distribution should then be due to the reallocation effect. I conduct this exercise in the NLSY and in the CPS data from section 1.2.1. 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.

Figure 1.8 displays the effect of the occupation-specific skill prices. We can see that in both datasets the increase in wages at the top of the distribution is quite well explained by the estimated price changes alone. The increase at the bottom is however hardly explained at all, despite the high point estimate of $\triangle(\pi_L - \pi_M) = 31.4\%$. This appears somewhat as a puzzle, since I would have expected that at least part of the increase in the bottom of the wage distribution should be due to higher relative prices in the low-skill occupation.

There are two interrelated reasons for the lack of an increase in the bottom of the counterfactual wage distribution compared to the middle. First, the dispersion of earnings within occupation groups is large, such that the respective occupational wage distributions overlap substantially and that an increase in the price per unit of skill in the low-skill occupation lifts wages of some middle-earners as well. Second, the estimated price changes are large enough such that an "overtaking effect" becomes empirically relevant, whereby some low-wage earners in the low-skill occupations become middle-wage earners and vice versa for some middle-wage earners in middle-skill occupations.[33] Together, these two factors prevent a strong increase of relative wages in the bottom of the counterfactual wage distribution despite the high point estimate for the relative price changes.

The results about the difference between actual and counterfactual wages are similar when I use the alternative definitions of occupation groups that have been used in the literature. These include grouping occupations according to initial median wages or average education, splitting up the large middle-skill group into blue collar and white collar occupations, and employing continuous measures of routine and nonroutine (analytical and manual) task content in occupations. As above, all these groupings share the feature that the wage dispersion within them is substantial.[34] However, in the case of tasks, one should note that measurement 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 2010), 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

[33] The corresponding statistics are not reported for the sake of brevity.

[34] Again, the results on alternative occupational groupings are not reported in detail in order to save space but available from the author upon request.

to the tasks that technology and trade have replaced may help to better match the wage distribution, since the dispersion of wages conditional on them may also be lower.[35]

In addition to the change in the overall wage distribution, figure 1.9 depicts the change in average wages in high-, middle-, and low-skill occupations for the NLSY and CPS. The counterfactual wage increase in the low-skill occupation is much higher than the actual in both datasets, while the increase in middle- and high-skill occupations is lower. Again, this is similar when I use alternative occupational groupings. Overall, hence, it seems that the estimated relative price changes across occupations alone cannot match the empirical facts about wages in the data.

What remains as an explanation, according to the model, is therefore the effect of reallocation on different parts of the wage distribution. 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, i.e. about half of the maximum wage increase that they could possibly obtain (31.4% − 2.4%).[36] Figure 1.10 plots the resulting counterfactual wage distribution which fits the actual quite well, especially in the CPS. Moreover, figure 1.11 displays the corresponding changes of average wages in occupations, which are now also closer to the actual than without reallocation.[37]

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 to so gainfully: for example, given probably not too different skill require-

[35]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 employed by Autor and Handel (2012) or by Spitz-Oener (2006) for Germany might improve the precision of measurement and the variation in wages that it captures substantially.

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

[37]In fact, the fit may be even better than in figures 1.10 and 1.11 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.

ments, 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 1.10 are quite strong. Firstly, the concentrated switching of low-earners in the middle-skill occupation requires that the population distribution of skills in the low-skill occupation be very condensed so that the low-earners are the first to find it profitable to "switch down". This 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. Secondly, the gains from switching that I need to assume seem high.

Moreover, the assumptions are not strictly testable. 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 1.4 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. Moreover, unreported regressions of the gains from reallocation for observables on workers' wages in the 1980s yield no clear relationship. Finally, I obtain essentially the same results about reallocation when I use the alternative definitions of occupation groups or task measures discussed above.

Therefore, I conclude that, as it is currently implemented in the literature, polarization seems to explain much but not all of the changes in the wage distribution that have occurred over the last decades. Within the polarization story, the most promising avenues for matching the whole wage distribution are to provide evidence for a large reallocation effect at the bottom and to search for more precise empirical measurements of the jobs or the tasks for which demand has declined.[38] However,

[38] The strong role for reallocation, if it was substantiated in further research, would be conceptually and economically important. First, workers in fact gain from switching down into an on average lower-paying occupation because they find a better match there. This is a conceptually important point that only models of relative—rather than absolute—advantage can make. It would thus emphasize the fact that there is no one mapping from occupations to the wage distribution. Second, contrary to some existing studies which find strong wage losses from workers switching

simply having some more occupation groups or tasks alone will not help much unless the increase in the variation in wages that these finer groups explain is large.

1.7 Conclusion

This article is the first to study the effect of job polarization on the wage distribution accounting for the endogenous sorting of skills. I do this by employing newly available data from the National Longitudinal Survey of Youth (NLSY) which provides detailed, multidimensional, and pre-determined measures of workers' talents (i.e. test scores) in order to hold different kinds of workers fixed and analyze the returns to occupation-specific skills over time. The estimation equations are derived from a Roy model over two cross-sections with job polarization amounting to a change in the occupation-specific skill prices. In this case, I show that predictions about wage changes depend exclusively on relative occupation-specific skills, which can be measured via the allocation of talents.

My results indicate that a one percentage point higher propensity to work in high- (low-) as opposed to the middle-skill occupations in the base period is associated with a .29 (.70) percent increase in wages over time, and therefore workers with comparative advantage in the middle-skill occupations lose out substantially over time. Furthermore, the effect of job polarization on workers' wages does well to match the changes at the top of the wage distribution but appears unable to wholly explain the changes at the bottom. Thus, occupational demand seems to have been the driving force of a substantial part but not all of the changes in the wage distribution over the past two decades.

These findings suggest that the dismal trend in middle class wages over the last couple of decades may not be fully explained by the changes in technology and globalization that coincided with it. In particular, (relative) incomes in the bottom and the middle of the distribution could have been affected by policy variables and labor market institutions such as the minimum wage and de-unionization. Thus, policies that encourage union formation or other measures that increase workers' bargaining power may be effective in raising middle class wages.

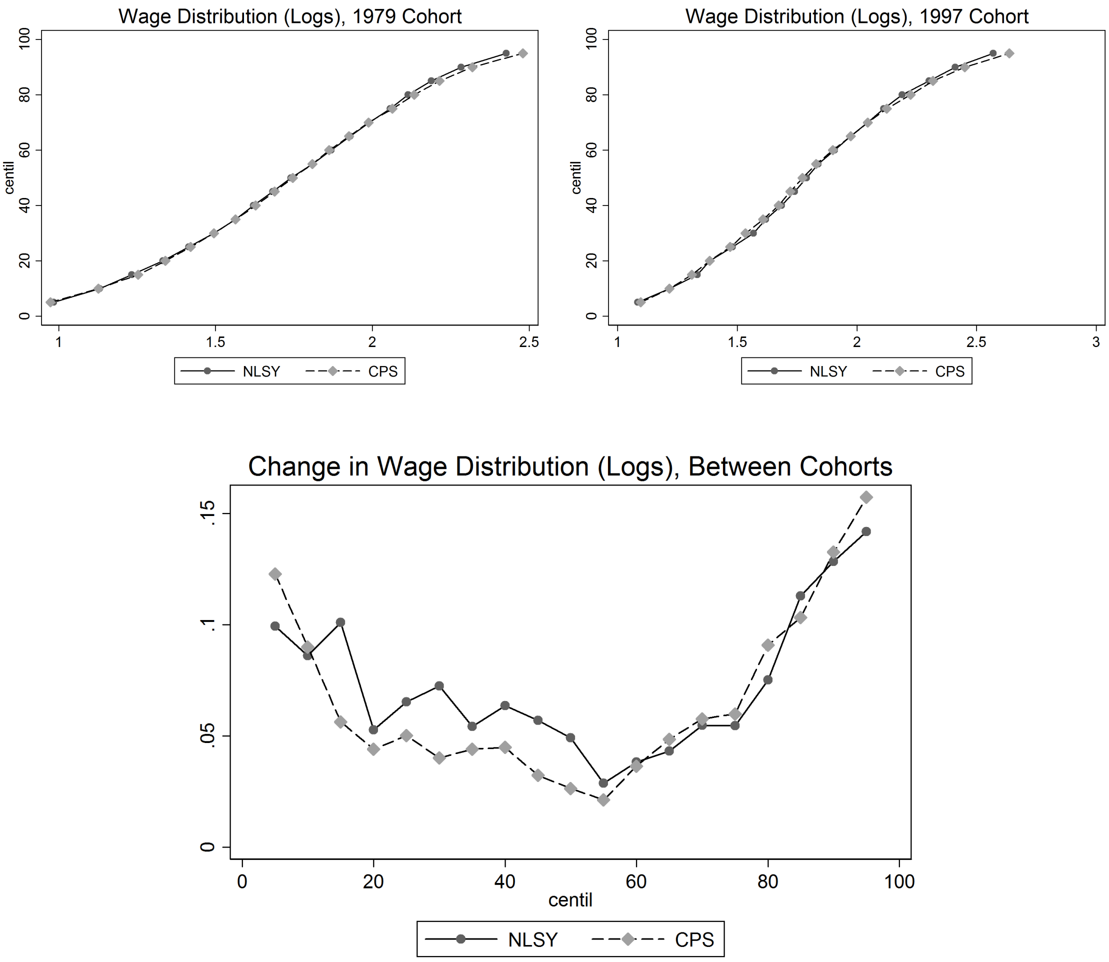
In future research it will be important to examine whether the result that the

down in panel data (e.g. Cortes 2012, Ebenstein, Harrison, McMillan, and Phillips 2011, Liu and Trefler 2011), it would suggest that switching may be an important channel to cushion the negative impact of polarization on the lowest earners.

wage distribution at the bottom cannot be fully explained by demand shocks is robust in other datasets that may become available and for a more precise measurement of the jobs and tasks that may have declined. In addition, similar analyses for European countries, with their different labor market institutions, would help to disentangle the effect of policy instruments on the change in the wage distribution and their interaction with the undoubtedly existent demand shocks.

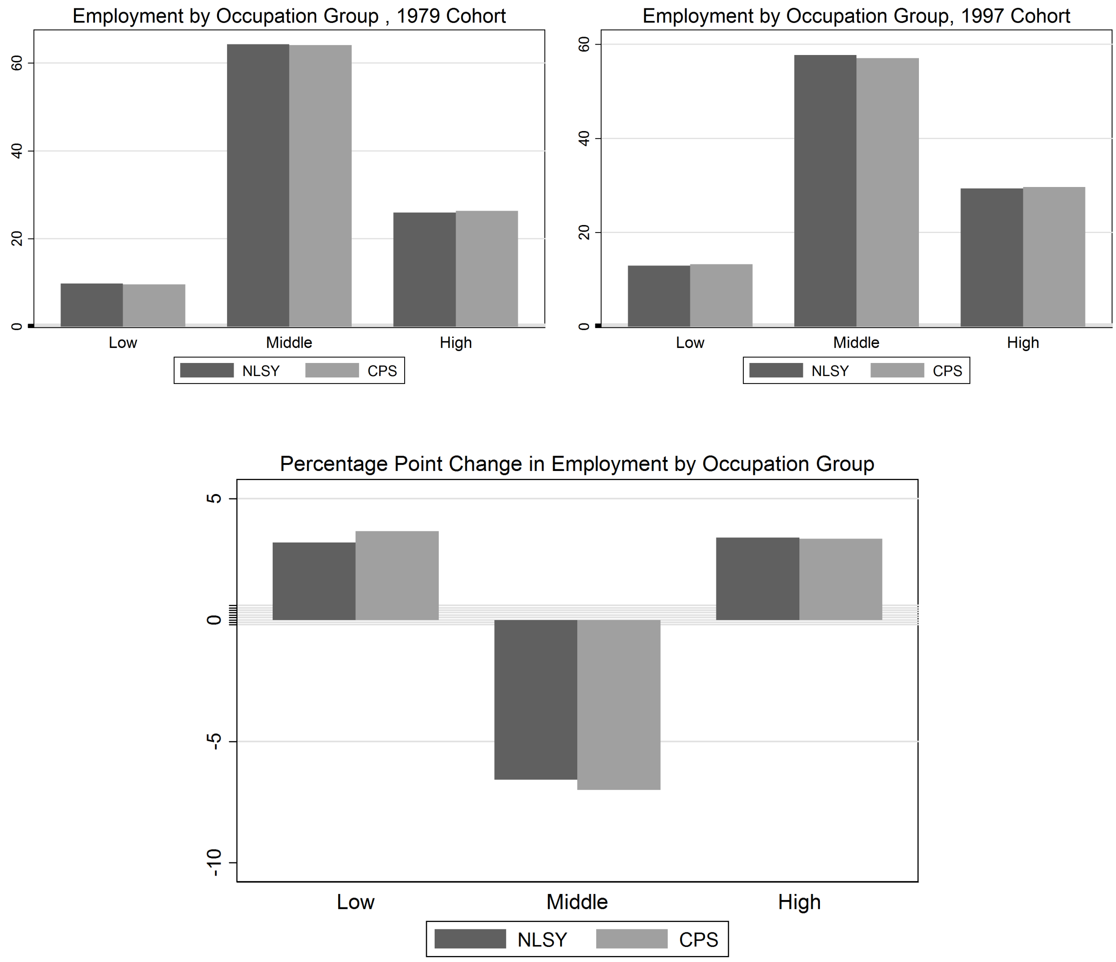
Finally, the methods developed in this paper can be applied more generally to study the effect of other important demand shocks on the labor market. For example, there are ongoing debates about a long-term increase in the demand for talent in the financial- and related sectors, and about the effect of the Great Recession on the wage distribution. These debates may be vitally informed by the "allocation of talents" perspective.

Figure 1.1: The Distribution of Log Wages and its Change



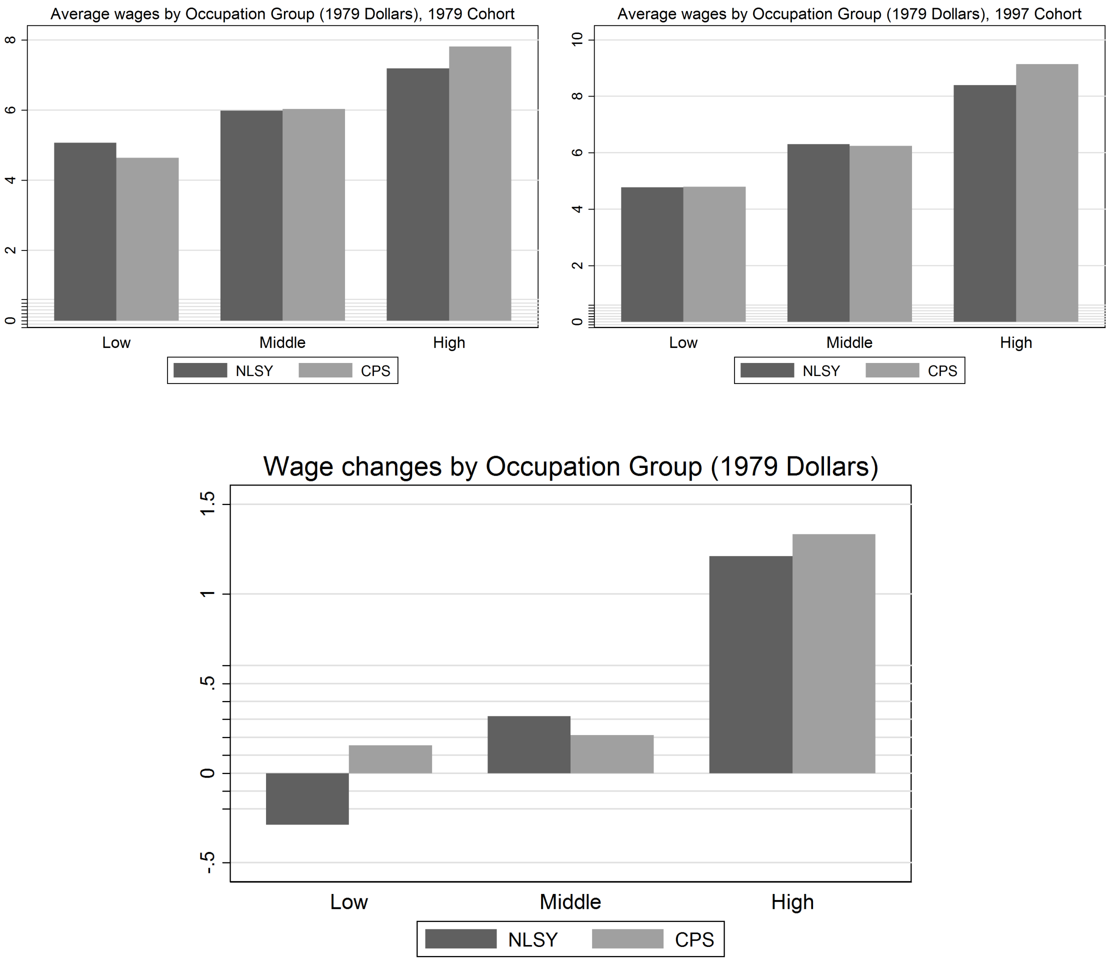
Notes: The subfigure on the top left depicts the empirical cumulative distribution of log real wages for 27 year olds in the NLSY79 cohort and for the comparable years and age group in the CPS. The subfigure on the top right does the same for the NLSY97. The subfigure at the bottom compares the changes in log real wages along the quantiles of the wage distribution over the two cohorts.

Figure 1.2: Employment Shares by Broad Occupation Group and their Changes



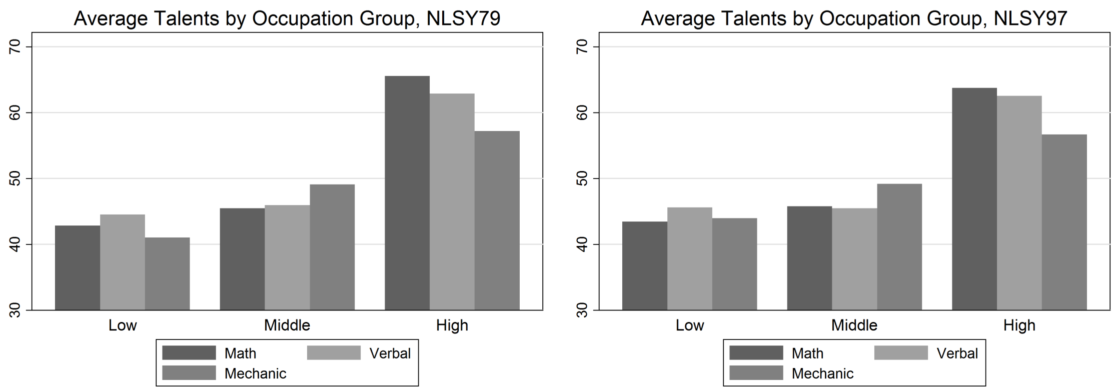
Notes: The subfigure on the top left depicts the employment shares of low-, middle-, and high-skilled occupations for the NLSY79 cohort and the comparable years and age group in the CPS. The subfigure on the top right does the same for the NLSY97. The subfigure at the bottom depicts the percentage point change in employment in the three occupation groups and again the CPS in comparison. The high skill occupation group contains managerial, professional services, and technical occupations. The middle skill occupation group contains sales, office / administrative, production, and operator and laborer occupations. The low skill occupation group contains protective, food, cleaning and personal service occupations.

Figure 1.3: Real (1979) Wages by Broad Occupation Group and their Changes



Notes: The subfigure on the top left depicts the average real wages of low-, middle-, and high-skilled occupations for the NLSY79 cohort and the comparable years and age group in the CPS. The subfigure on the top right does the same for the NLSY97. The subfigure at the bottom depicts the change in real wages in the three occupation groups and again the CPS in comparison. The high skill occupation group contains managerial, professional services, and technical occupations. The middle skill occupation group contains sales, office / administrative, production, and operator and laborer occupations. The low skill occupation group contains protective, food, cleaning and personal service occupations.

Figure 1.4: Average Talents in Occupation Groups, 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 1.5: The Estimation Problem

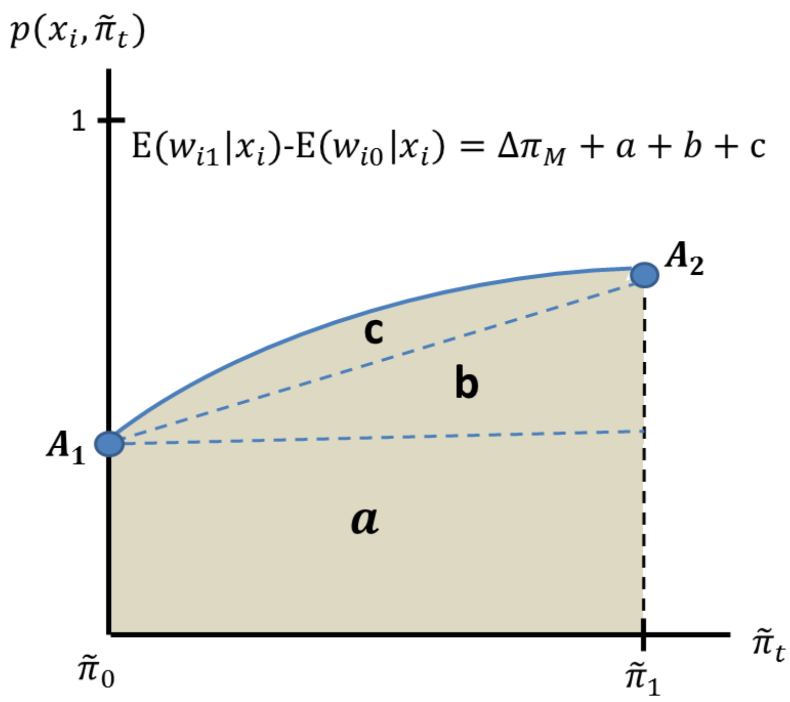
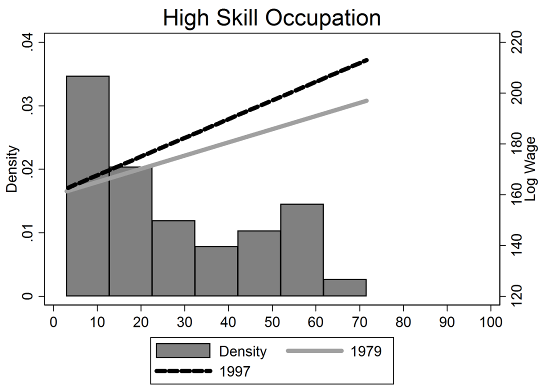
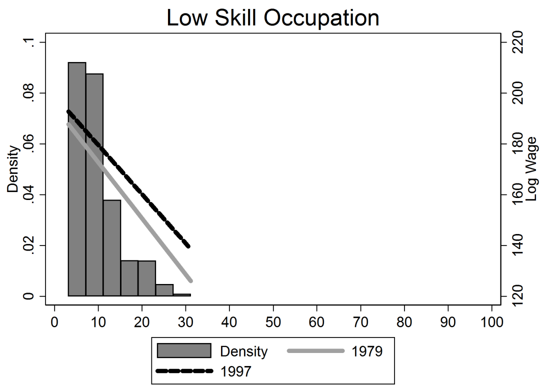


Figure 1.6: Predicted Relative Skill Returns and their Changes

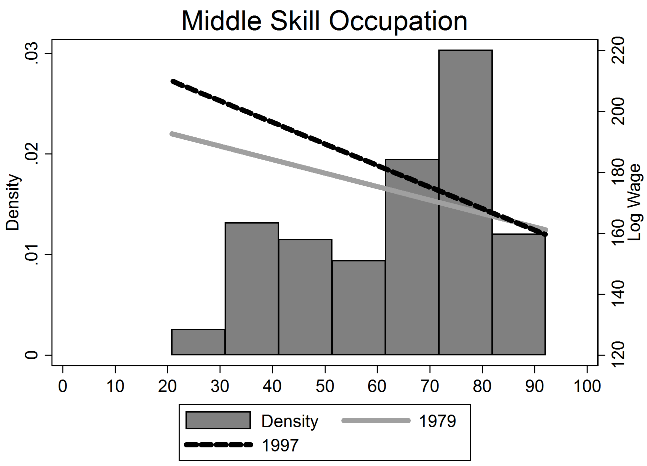
(a) Propensity High Occupation



(b) Propensity Low Occupation



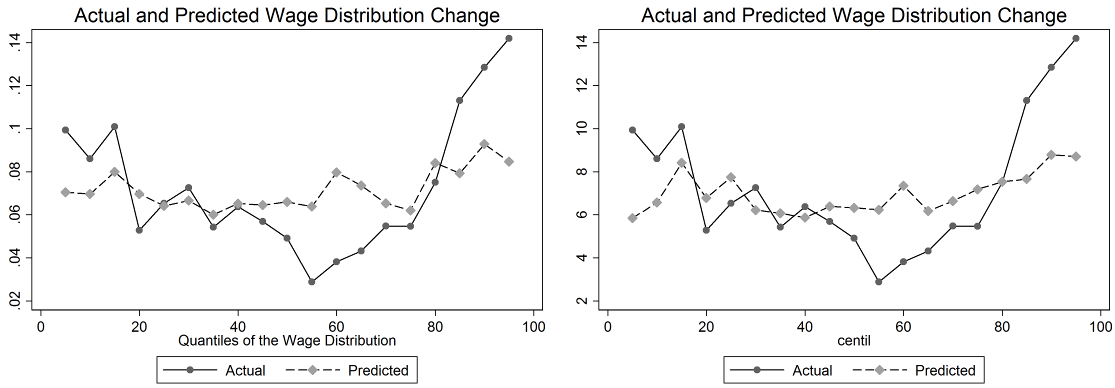
(c) Propensity Middle Occupation



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 1.7: Actual and Predicted Wage Distribution Change

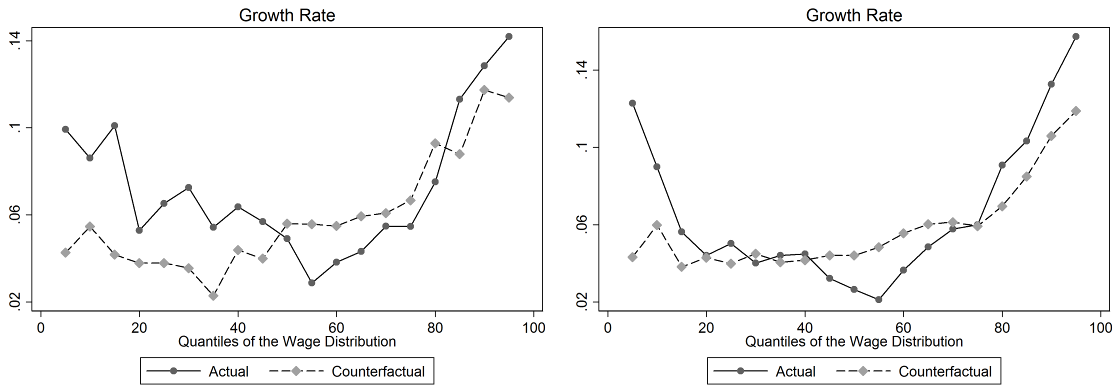
(a) Returns to Propensities (b) Returns to All Talents



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 1.5.

Figure 1.8: Actual and Counterfactual Wage Distribution Change, NLSY and CPS

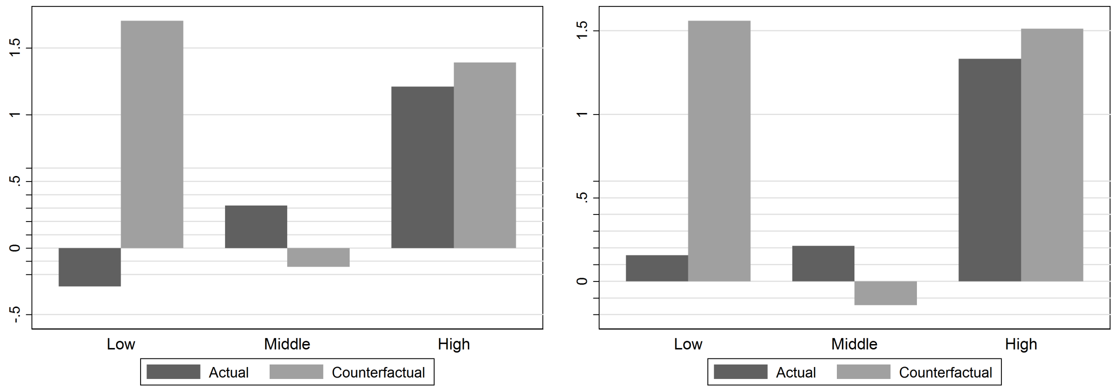
(a) NLSY (b) CPS



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 from the optimal minimum distance estimator in table 1.7.

Figure 1.9: Actual and Counterfactual Occupational Wage Changes, NLSY and CPS

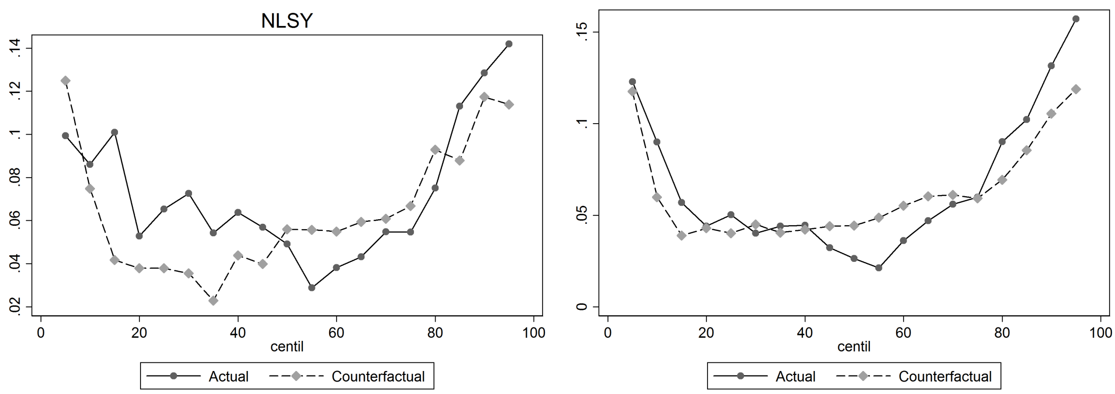
(a) NLSY (b) CPS



Notes: The figures plot the actual and the counterfactual change in occupational wages in the NLSY and CPS when workers in the initial period are assigned the estimated price changes in their occupations from the optimal minimum distance estimator in table 1.7.

Figure 1.10: Actual and Counterfactual Wage Distribution Change with Reallocation, NLSY and CPS

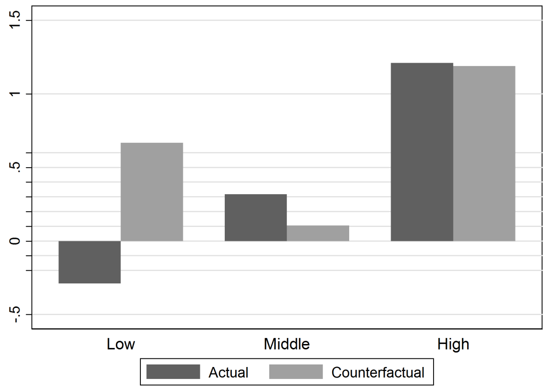
(a) NLSY (b) CPS



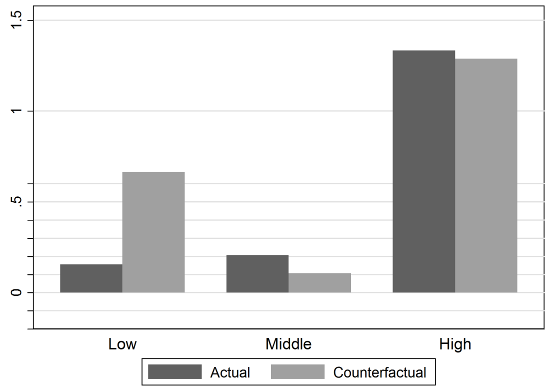
Notes: The figures plot the actual and the counterfactual change in occupational wages in the NLSY and CPS when workers in the initial period are assigned the estimated price changes in their occupations plus a reallocation effect: the lowest-earning three percent are assumed to move out of the middle- 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 1.11: 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 in the NLSY and CPS 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 1.1: From the full NLSY to the analysis sample

	NLSY79 (Birthyears 1956-1964)	NLSY97 (Birthyears 1980-1984)
Reason for exclusion		
Total males	6,403	4,599
Excluded oversampled white and older arrivers in US than age 16	4,585	4,599
Birthyear > 1982	4,585	2,754
Type of attrition		
Ought to be present with ASVAB at age 27	4,585	2,754
No ASVAB excluded	4,299	2,081
%	94	76
Not present at age 27 excluded	3,939	1,737
%	86	63
Conditioned on working		
Excluded who report no or farm occupation, self-employed, and those with no wage income	3,054	1,207

Note: 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 1.2: Labor Supply with Respect to Average Demographics, Early, and Contemporary Skill Determinants

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

Note: 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 1.3: Pairwise Correlations between Composite Test Scores

	NLSY79 AFQT	Math	Verbal	NLSY97 AFQT	Math	Verbal
AFQT (NCE)	1			1		
Math Score (NCE)	0.82	1		0.83	1	
Verbal Score (NCE)	0.93	0.71	1	0.92	0.75	1
Mechanical Score (NCE)	0.63	0.53	0.61	0.63	0.54	0.63
Nbr Observations	2936			1210		

Note: 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 1.4: Sorting into Occupation Groups, Multinomial Logit Regressions

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

Note: 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 1.5: Returns to Occupational Propensities over the Two Cohorts

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

Note: The table reports OLS wage regressions of 100 times the deflated log wage on propensities to enter occupation groups (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. 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. "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. 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 1.6: Talent Allocation and Returns Changes

	High Skill Occup	**Low Skill Occup**	**Log Wage x NLSY97**
Constant	18.71***	11.87***	14.37**
	(6.21)	(5.51)	(2.08)
Black	-0.762	9.292***	0.647
	(-0.43)	(4.77)	(0.14)
Hispanic	-2.632	1.708	-1.586
	(-1.34)	(1.12)	(-0.36)
Middle Math Tercile	10.83***	-4.466***	-2.615
	(6.05)	(-2.85)	(-0.56)
High Math Tercile	34.90***	-5.997***	10.44*
	(13.48)	(-3.10)	(1.68)
Middle Mechanic Tercile	-2.505	-2.332	-5.767
	(-1.19)	(-1.52)	(-1.20)
High Mechanic Tercile	-7.043***	-4.827***	-1.740
	(-2.81)	(-2.97)	(-0.30)
Middle Verbal Tercile	3.505*	2.429	-0.282
	(1.84)	(1.48)	(-0.06)
High Verbal Tercile	15.34***	2.805	-4.535
	(5.67)	(1.45)	(-0.65)
Illicit Activities (NCE)	-0.129***	0.0388	-0.183*
	(-3.33)	(1.36)	(-1.89)
Precocious Sex (NCE)	-0.0612	-0.0120	0.0527
	(-1.64)	(-0.41)	(0.62)
R-squared	0.182	0.0281	0.0933
N	4146	4146	4146

Note: The first two columns present the coefficients from OLS allocation regressions of working in the low and high skill occupation with pooled NLSY79 and NLSY97 data. The third colum 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 1.7: Implied Wage Rate Changes and Cross-Equation Restriction Test

	Estim. $\triangle(\pi_H - \pi_M)$ in % (s.e.)	**Estim. $\triangle(\pi_L - \pi_M)$ in % (s.e.)**	**Implied $\triangle\pi_M$ in %**	**Test Statistic (p-value in %)**
OMD / Full GLS	20.1	31.4	-2.4	13.1
	(9.7)	(35.2)		(10.7)
EWMD / OLS	19.4	-4.4	1.7	13.2
	(10.8)	(32.0)		(10.5)
DWMD / WLS	22.0	-7.5	1.3	11.2
	(9.7)	(35.1)		(19.1)

Note: 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 prefered 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 feasable 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).

Chapter 2

The Allocation of Talent over the Business Cycle and its Long-Term Effect on Sectoral Productivity

The Allocation of Talent over the Business Cycle and its Long-Term Effect on Sectoral Productivity

Michael Boehm and Martin Watzinger*

February 2013

Abstract

It is well documented that graduates enter different occupations in recessions than in booms. In our article, we examine the impact of this reallocation for long-term productivity and output in a sector. We develop a model in which talent flows to stable sectors in recessions and to cyclical sectors in booms. We find evidence for the predicted change in productivity caused by the business cycle in a setting where output can be readily measured: economists starting or graduating from their PhD in a recession are significantly more productive over the long term than economists starting or graduating in a boom.

Keywords: Talent Allocation, Sectoral Productivity, Business Cycle, Roy Model, PhD Economists

JEL Classification Numbers : J24, E32, I23, J22, J23

* Boehm: London School of Economics. Watzinger: University of Munich. This is a substantially revised version of a paper previously circulated as “The Allocation of Talent: Evidence from the Market of Economists”. We thank Oriana Bandiera, Ziad Daoud, Luis Garicano, Dirk Jenter, Daniel Metzger, Paul Oyer, Joern-Steffen Pischke, Andrea Prat, and Monika Schnitzer for helpful suggestions. Furthermore we thank the SFB-TR 15 of the German Research Foundation for financial support and JSTOR for providing their Data for Research interface.

2.1 Introduction

An extensive recent literature has documented a strong and persistent impact of initial labor market conditions on individuals' earnings.[1] Many of these studies have identified a change in first jobs or occupations as the main cause for this effect. For example, graduating MBAs are less likely to get a job in investment banking if there is a shock to financial markets (Oyer 2008). Since starting on Wall Street upon graduation makes a person more likely to work there later, temporary shocks can have large impacts on MBAs' lifetime earnings. The effects on initial jobs are not limited to the very top of the earnings distribution: for example Oreopoulos, von Wachter, and Heisz (2012) find that college graduates entering the labor market during recessions take up jobs with lower paying employers and then gradually—but not fully—recover by switching to higher paying employers over time.[2]

Despite showing that temporary shocks change workers' initial jobs and in principle recognizing that this "can lead to persistent changes in the allocation of workers" (Oreopoulos, von Wachter, and Heisz 2012), none of these studies has analysed the aggregate implications of such a reallocation. More concretely, because of the above findings, the business cycle can affect the size and the composition of the workforce across sectors and thus have long-term effects on sectoral productivity and output. This is particularly important if sectors benefiting from an inflow of talent during recessions have a higher (or lower) social value than the sectors where the talent is drawn from. For example, if in a recession talented individuals choose to work in entrepreneurship or research instead of rent-seeking sectors, this may have some social benefits that subtract from the immense adverse effects of downturns. And, even if this is not the case, the observation that hiring in a downturn is a cost-effective way to attract and retain talent may be a compelling rationale for anti-cyclical recruitment policies of businesses and the public sector alike.

This is the first study to explore the impact of the business cycle on long-term productivity and output in a sector via the allocation of talent. To do this, we focus on the productivity of new hires over the business cycle in an occupation where

[1]See, for example, Oreopoulos, von Wachter, and Heisz (2012), Devereux (2002a), Raaum and Røed (2006), Sullivan and von Wachter (2009), Kahn (2010), Genda, Kondo, and Ohta (2010), Oyer (2006), Oyer (2008).

[2]Other studies have shown that college enrolment rates rise during recessions (e.g. Gustman and Steinmeier 1981, Black and Sufi 2002). This suggests that at least part of the effect on initial occupations is a deliberate choice by the affected individuals.

output is well measurable. We proceed in two steps: first, we develop a Roy-style model of occupational choice that speaks to our empirical setting below. Yet it makes two general points, informing us about the expected composition of talent if all individuals can freely choose their occupation and if only a certain number of jobs are available. In the first case workers are allocated according to relative advantage, while in the second case individuals must have an absolute advantage to be able to enter the restricted sector. Second, we collect data from an empirical setting that is well suited for our study of the reallocation of talent and sectoral productivity: the career choices and publication records of economics PhDs who graduated from the top 30 US universities.

We are examining this particular occupation because academic publication records provide us with a direct measure of productivity that does not suffer from two fundamental flaws: endogeneity to the business cycle and imprecision. First, wages or firm output are not only affected by the productivity of workers but also by product market demand, which falls in recession. Therefore it is hard to learn from such measures about a change in worker productivity caused by recessions. Second, since a firm's production is normally the result of a collaborative effort of many individuals, it is difficult to infer from a change in output the value of a specific worker's or cohort's contribution. Contrary to that, academic publications can be attributed to particular individuals, we can quite well assess their quality, and journals' demand for articles does not vary over the business cycle.[3] In addition, the education of PhD economists provides two different and well-defined career decision points, the application to and the graduation from graduate school, which conform relatively well to the absolute and the comparative advantage cases mentioned above.

For our empirical analysis we construct a new dataset of economists' career choices and publication output from publicly available sources. The dataset consists of graduation years and the degree granting universities of 13,624 PhDs from 1955 to 1994 from the top 30 American institutions. We match each person with all their publications in JStor during the first ten years after graduation and with an indicator for becoming a faculty member or a member of the American Economic Association (AEA) after the PhD. Thus, we can calculate the propensity to stay in academia and the long-term publication output for each economist. Finally, we aggregate each

[3] While there may be fiercer competition for fixed journal spaces, this should be the same for recession and boom cohorts.

cohort according to university and graduation year, and match different business cycle indicators (recession dummies, GDP growth, and unemployment rates and their changes) at time of application to and at time of graduation from a PhD program. We quantify the influence of the business cycle indicators at both points in time on economists' propensity to decide in favor of academic employment and on their productivity.

Our empirical results support the predictions of the Roy Model and the more general idea that the business cycle influences sectoral productivity and output via a reallocation of talent. The model predicts that a recession during the application period makes entry into graduate school more competitive, because application rates rise and the number of available spaces is more or less fixed. Consequently, the ability of admitted PhD students should increase during recessions. Indeed, in our data, cohorts who entered during a recession publish more on average, i.e. are more productive, than boom cohorts. The model also predicts that at graduation from the PhD, when top-jobs in academia are hard to get but there is potentially some flexibility in the number of lower-ranked academic jobs, the number and / or the quality of individuals staying in academia rises during recessions. In our data, we find that individuals who graduated during a recession are more likely to become academics and those who do, publish more on average. Overall output in the academic sector thus rises. Finally, the model predicts that individuals who entered a PhD program during recession are less likely to stay in academia after graduation. The reason is that some of them have relatively strong non-academic skills and find it profitable to leave academia after the economy recovered. We also find empirical evidence in support of this idea.

We quantify the long-term effects of the business cycle on productivity and output because our measure takes into account all publications authored in the first ten years after graduation. The effects accumulate during this time span indicating a persistent difference between boom and recession cohorts. Moreover, they are of economically substantial magnitude: we expect assistant professors from a cohort who applied to the PhD during a typical recession (a rise in unemployment rates by 2.5 percentage points) to be 17 percent more productive than assistant professors who applied in an average year (0 percentage points unemployment change). Furthermore, three percent more PhD graduates stay in academia in a typical recession and they produce on average 14 percent more publications than economists graduating in an average year.

Our results are robust over a wide variety of alternative measures of output and occupational choice, different control variables as well as in different subsamples. For example, the change in productivity is most pronounced for the graduates of the top Tier 1 universities but also holds over the entire skill distribution. Unemployment change is our preferred measure of the business cycle, but NBER recession indicators or GDP Growth as explanatory variables deliver similar results. The number of publications in the top 5 journals as a measure of productivity works almost as well as impact weighted-publications. Controlling for a time trend or academic subfields does not change our results. It does not matter qualitatively if we use listing in a faculty database, the propensity to publish, or membership in the American Economic Association as indicator for being a member of the academic sector.

This paper informs at least three ongoing debates. Foremost, in the literature on long-term effects of recessions, it identifies and demonstrates the quantitative importance of a hitherto ignored implication of its findings—that recessions change the composition of talent across sectors and thus their long-term productivity and output. Prior literature recognized that initial jobs shape long-term careers of individuals and that they differ for recession and boom cohorts (e.g. Oreopoulos, von Wachter, and Heisz 2012, Devereux 2002a, Kahn 2010, Oyer 2006, Oyer 2008).[4] Furthermore, it documents that less advantaged workers seem to be most affected (in particular Oreopoulos, von Wachter, and Heisz 2012), although Oyer (2008) also finds large effects for MBAs. Our study adds to these findings that the recession-induced reallocation of workers changes the long-term productivity and output of a sector and that these effects are quantitatively important for the top of the skill distribution. A concurrent paper corroborates our results by showing that MIT students who graduate during recessions are producing more patents over the long term and that this is likely to stem from initial occupational affiliations (Shu 2012).

Our findings also add to the literature on the cyclical upgrading of labor (e.g. Okun 1973, Vroman and Wachter 1977, McLaughlin and Bils 2001, Devereux 2002b). This literature shows that workers move to higher paying employers or occupations during booms and to lower paying ones during recession. The reason is that the labor demand of high-wage employers increases in upturns, reducing competition

[4] These studies also discuss the sources of long-term effects, considering hypotheses about human capital accumulation, employer learning, stigma, search, and others.

for their jobs. In recession, when competition is correspondingly higher, the average education level for new hires increases (Devereux 2002b). Naturally, this indicates a reallocation of talent over the business cycle. In particular, it appears that in booms workers with a lower skill endowment are able to enter jobs which they would not enter in a recession. Our findings quantify the resulting effect on productivity and output for one particular occupation.

Another debate that our paper contributes to is concerned with the impact of science funding on research productivity. Funding increases, like recessions in our context, raise the attractiveness of the academic sector compared to the private sector. Goolsbee (1998) shows that up to 50% of a government spending increase goes into higher salaries for scientists and engineers. Suggesting that the supply of such knowledge workers is relatively inelastic, he argues that a large fraction of governmental research funding may in fact be ineffective and may only constitute a windfall gain for scientists. To the contrary, our results imply that the quantity and / or quality of scientists should strongly and persistently increase with more funding.[5]

The remainder of this paper proceeds as follows. We derive our theoretical predictions from a modified version of the Roy Model in the next section. Then we describe how we assembled our novel dataset of PhD economists' career choice and publication success. Section 3.4 presents and interprets the empirical results, while the conclusion discusses to what extent our results may generalize to other segments of the labor market. The appendices contain robustness checks that seem important to us but would disturb the flow of the argument in the main text.

2.2 Theory

We are interested in how the selection of skills into academia and business varies with the state of the business cycle. This section modifies a standard Roy (1951) model for the problem at hand. The Roy Model analyzes the self-selection of individuals with heterogeneous skills into sectors according to their highest expected earnings. In the following, we model two sectors—academia and business—into which individuals can self-select. Every individual has distinct skills (and therefore different wages)

[5] Along these lines, Freeman and van Reenen (2009) assert that, at least in the long run, not only the number of scientists but also the selection of talent into science will increase due to higher salaries.

in each sector but can choose only one occupation. The main departure from the original Roy framework is that compensation in business and academia vary with the business cycle and that the number of open positions in academia is assumed to be fixed.

2.2.1 Assumptions

Suppose that individuals are endowed with two skills, an academic skill α and a business skill β. There are two sectors, academia (A) and business (B), which produce outputs utilizing the respective skills. Individuals maximize their expected lifetime compensation by applying for jobs in academia or business. This compensation implicitly consists of a pecuniary and a non-pecuniary component, where the non-pecuniary component might be particularly important in the academic sector (see Stern 2004).

The business sector is assumed to hire anyone offering a compensation w_t. The compensation depends linearly on the skill level β of the employee and the state of business cycle $\tilde{y}_t$:

$$w^B(\beta) = \beta + \tilde{y}_t.$$

An employee's lifetime compensation in the business sector is higher in a boom (high $\tilde{y}_t$) and lower in a recession (low $\tilde{y}_t$). In academia, total compensation also varies with the business cycle but is less cyclical than in the business sector:

$$w^A(\alpha) = \alpha + a\tilde{y}_t$$

with $a < 1$.

Two sources may contribute to the variability of compensation over the business cycle: First, in a recession, lower immediate wages can lead to a lower lifetime compensation in both sectors. Second, during recessions employees enter inferior career paths in business or start at a lower ranked institution in academia, which could hurt lifetime income and non-pecuniary benefits. This is consistent with recent findings (e.g. Oyer 2008, Oreopoulos, von Wachter, and Heisz 2012). Importantly, we assume that the academic sector is less cyclical than the business sector and we provide empirical evidence supporting this assumption in appendix B.2, where we show that academic job offers for economists are less cyclical than non-academic ones and we argue that the non-pecuniary benefits from academia should make its even less cyclical.

In order to become an academic, an individual must decide for academia twice: first by applying to a PhD program (at time of application $t = app$) and a second time by pursuing an assistant professorship after the PhD (at graduation $t = grad$). At time of application, we assume that PhD programs admit the best N applicants according to academic skill and that there are always more applicants than available spaces.[6] Thus, the entry into the doctoral program is competitive.[7]

At graduation, we do not know if only a fixed number of academic jobs are available or if graduates can freely choose to stay in academia: Obtaining an assistant professorship at a ranked university is very competitive, indicating that only a limited number of spaces are offered. However, conditioned on graduating from one of the top 30 US economics departments, it also seems unlikely that a student cannot secure an academic job at a lower ranked institution, a teaching college, a university outside the United States, or a postdoc position even in times of recession. Probably, the truth lies somewhere in between these two extremes, so we derive predictions for both cases.

When taking his decision to apply for a PhD program, the applicant should also take into account the option value of having another choice about his career path after graduation. To simplify our problem, we assume that this option value is a constant, i.e. that it does not vary with the state of the macroeconomy at the time of application.[8] Thus, we can subsume this constant in the individual's non-varying compensation component, the academic skill level α.

Given these assumptions, an individual compares the expected compensation from academia $\alpha + a\tilde{y}_t$ and business $\beta + \tilde{y}_t$ at time of application and at graduation. He decides to apply for the academic sector (the PhD program or the assistant

[6] PhD entry cohort sizes as measured by the number of full-time, first-time graduate students, are not related to the business cycle in our data (see Appendix B.3).

[7] The allocation is therefore governed by absolute advantage for those individuals who prefer academia (see Sattinger 1993).

[8] In effect, this assumption amounts to imposing that the business cycle at time of application has no predictive power for the business cycle at graduation. We think that this is defendable as it takes on average six years to complete a PhD and we show in Appendix B.3 that there is no correlation between the business cycle at time of application and graduation in our data. In general, we expect that our results should also hold in all of the cases where there is a reversal in the business cycle during that time frame, i.e., $Pr(\tilde{y}_{grad}^{Boom}|\tilde{y}_{app}^{Rec}) > Pr(\tilde{y}_{grad}^{Boom}|\tilde{y}_{app}^{Boom})$ and $Pr(\tilde{y}_{grad}^{Rec}|\tilde{y}_{app}^{Boom}) > Pr(\tilde{y}_{grad}^{Rec}|\tilde{y}_{app}^{Rec})$, and in a lot of cases where there is sufficiently strong mean reversion.

professorship) whenever

$$\alpha > \beta + y_t. \tag{2.1}$$

where $t \in \{app, grad\}$ and $y_t \equiv (1-a)\tilde{y}_t$. y_t is the relative attractiveness of the business sector that is due to the business cycle.[9]

2.2.2 Predictions

We are interested in how the selection of skills into academia and business varies with the state of the business cycle. To ease the exposition, we compare a generic boom cohort versus a generic recession cohort, i.e. $y^{Boom} > y^{Rec}$. All proofs are relegated to Appendix B.1.

Proposition 2.2.1 *For PhD applicants, the joint distribution of academic and business skills selected into the academic sector during a recession first order stochastically dominates (FSD) the corresponding boom distribution.*[10]

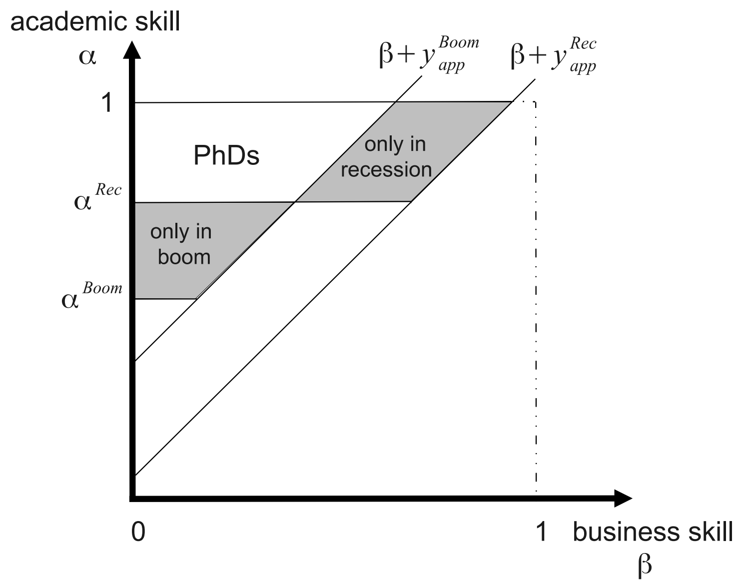
Figure 2.1 illustrates Proposition 2.2.1 when academic and business skills are distributed uniformly in the unit interval. Given our assumptions, an individual's career choice is governed by a "one-shot" decision, with those individuals for whom $\alpha > \beta + y_{app}$ preferring academia. During a boom (a high y^{Boom}_{app}), fewer individuals apply for academia than during a recession (a low y^{Rec}_{app}), which is depicted by a higher cutoff line for the former than for the latter. Academic employers always hire a fixed number, N, of graduates (PhDs & "only in boom" in boom, PhDs & "only in recession" in recessions) and therefore the distribution of skills for the recession cohort lies to the "North-East" of the corresponding distribution for the boom cohort.

However, Proposition 2.2.2 shows that fewer of the PhDs who were admitted in a recession remain in academia and become assistant professors after the PhD.

[9] We could have added to the model that a PhD constitutes an investment into academic (and business) skills. This is clearly an important feature of obtaining a graduate education and we did this in an earlier version of this section. However, as long as the skill update and the uncertainty about it can be assumed to be independent of the state of business cycle, it does not change the predictions of the model other than by adding noise. Hence, we refrain from defining different (updated) αs, βs, and y_ts at PhD application and graduation.

[10] On the flipside, this implies that the joint distribution of skills selected into business during a boom first order stochastically dominates its recession counterpart. Note that in contrast to the well known result of the general Roy model (e.g. see Heckman and Honore 1990), we can make a definitive statement about the stochastic dominance for a general distribution of skills here. This is due to the assumption of binding quantity constraints and the resulting competitiveness of the admission into the academic sector.

Figure 2.1: Selection with a U(0,1) distribution of both skills at application



Proposition 2.2.2 *For every realization of the state of the economy at graduation* y_{grad}*, a (weakly) higher fraction of the members of a "recession at time of application" cohort do not remain in academia after the PhD.*

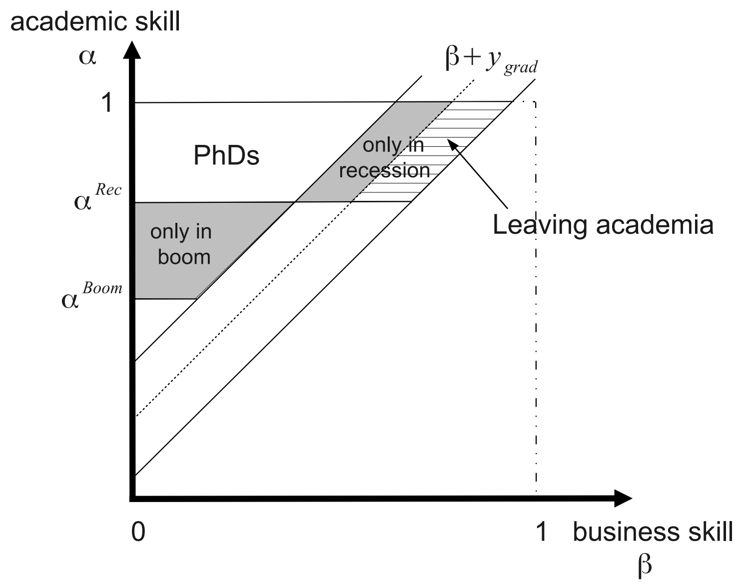
The proposition implies that, on average, cohorts of PhD graduates more often leave academia if they experienced a recession at the time of application. Figure 2.2 provides some intuition for the proposition. The academic skill cutoff, above which individuals will prefer academic employment after the PhD, "on average" moves down to the dashed line in the figure for a boom cohort and up for a recession cohort. Thus, in the figure, some individuals of the recession cohort exit academia and enter business after the PhD when the economy is out of recession, while everyone in the boom cohort stays in academia. The recession graduates who leave academia here are the marginal ones who applied for the PhD "because of" the recession in the first place.

Proposition 2.2.3 *For any given realization of the business cycle at graduation* y_{grad}*, the (partial) distribution of academic skills of the members of a "recession at application" cohort who remain in academia after the PhD first order stochastically dominates the distribution of skills of the corresponding members of the "boom at application" cohort.*[11]

Proposition 2.2.3 implies that, no matter how many more recession students than boom students leave academia after the PhD, the recession students who remain in

[11]However, the stochastic dominance of the joint distribution of business and academic skills does not feed through in general.

Figure 2.2: Selection with a U(0,1) distribution of both skills at graduation



academia are still better in each quantile of their (academic) skill distribution. In our specific example in Figure 2.2 we see that, although some mass of the recession cohort is cut off, the recession distribution of skills in academia still lies to the "North-East" of the boom distribution.

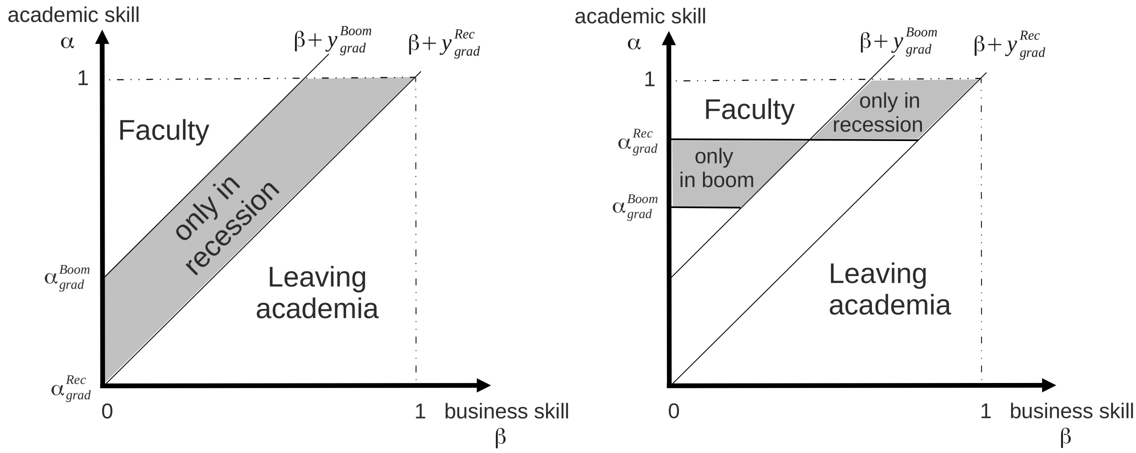
We now turn to the effect of the business cycle at graduation (y_{grad}). In a recession, relatively more graduates want to take up academic employment than in a boom. If all of these graduates can take up an academic occupation, more academics come out of a recession-at-graduation cohort than out of a boom-at-graduation cohort. Depending on the underlying skill distribution, these additional academics might be less or more able than the ones always staying in academia. If only a fixed number can take up an academic position (independently whether there is a boom or bust), a recession leads to (on average) better academics. Proposition 2.2.4 states this observation and its implications. Figure 2.3 provides a graphical representation in the special case of PhD graduates with academic and business skills distributed uniformly in the unit square.

Proposition 2.2.4 *A higher fraction of PhD economists wants to stay in the academic sector if they experience a recession at graduation. Depending on whether a fixed number of academic positions are available or not, the quality and / or the quantity of academics from recession cohorts increases.*

As mentioned above, the number of spaces in academia at graduation is probably neither completely flexible nor completely fixed, therefore we expect a combination

Figure 2.3: Selection at graduation

(a) with a flexible number of spaces. (b) with a fixed number of spaces.



of these two effects. In addition, this proposition implies that the average PhD cohort who graduated during a recession publishes weakly more, because weakly more PhDs choose to enter academia (if spaces at graduation are flexible) or better PhDs enter academia (if the number of jobs are fixed).

Finally, we can reformulate the four propositions of the model into empirical predictions for our data:

1. Fewer of the economists who experienced a recession at the time of application to the PhD end up in academia (from Proposition 2.2.2).

2. However, those who remain in academia are better researchers, both on average and in each quantile of their publication distribution (from Proposition 2.2.3).

3. More and / or better economists who experienced a recession at graduation stay in academia, increasing the publication output for the full sample of PhD graduates (from Proposition 2.2.4).

2.3 Data

We have collected a new dataset of career choices and individual productivity for a large sample of economists in the United States from 1955 to 2004. We aggregate the individuals into university year cohorts and match these with measures of the business cycle in the year of application and the year of graduation. The data sources are described in Table 2.1.[12]

[12] All further details of the data collection procedure and the employed programs are available from the authors on request.

Table 2.1: Data Sources

Variable	Description	Source
Personal information of graduates	Name, university and graduation year	AEA "List of Doctoral Dissertations in Economics" from 1955 to 1994
Faculty membership	Faculty directory of (mainly American) Economics, Business and Finance departments by James R. Hasselback	"Faculty Directories," James R. Hasselback, accessed 2011-02-07, `http://www.facultydirectories.com/`
Membership in the AEA	Membership data of the American Economic Association in 1970, 1974, 1981, 1985, 1989, 1993, 1997, 2003 and 2007	Supplement to the Papers and Proceedings Issue in the respective year digitalized by JSTOR
University ranking	Tier of a university according to the National Research Council	"The American Economic Association Graduate Study in Economics Web Pages," accessed 2011-02-08, `http://www.vanderbilt.edu/AEA/gradstudents/`
Publication records	Publications in 74 journals listed in the JSTOR online repository, from 1955 to 2004	"JSTOR Data for Research," last accessed 2011-02-07, `http://dfr.jstor.org/`.
Journal rankings	Citation ranking of journals in Economics, Business and Finance from 1950 to 2000	Laband and Piette (1994), Kalaitzidakis, Mamuneas, and Stengos (2003), Kim, Morse, and Zingales (2006) and "IDEAS/RePEc Recursive Discounted Impact Factors for Journals," last accessed 2011-02-07, `ideas.repec.org/`
Measure of the business cycle	Seasonally adjusted change in unemployment, unemployment levels and GDP growth from 1949 to 1994	Thomson Reuters Datastream
Recession Indicators	NBER recesssion indicators from 1949 to 1994	"The NBER's Business Cycle Dating Committee," last accessed 2011-08-09 `http://www.nber.org/cycles/recessions.html`
Duration of the PhD	Median years between registration and graduation from the PhD for 1977, 1986, 1996, 1997, 2001	National Science Foundation, Stock and Siegfried (2006), Hansen (1991)
Number of Graduates (NSF list)	Number and graduating PhDs according to the "NSF Survey of Earned Doctorates/Doctorate Records File"	"WebCASPAR Integrated Science and Engineering Resource Data System", last accessed 2012-03-16, `https://webcaspar.nsf.gov/`
Number of First-Time, Full-Time Graduate Students	Number of full-time, first-time graduate students according "NSF-NIH Survey of Graduate Students & Postdoctorates in Science and Engineering"	"WebCASPAR Integrated Science and Engineering Resource Data System", last accessed 2012-03-16, `https://webcaspar.nsf.gov/`

Table 2.2: The National Research Council Ranking of 1993

Tier	Universities
Tier 1 (ranked 1–6):	Chicago, Harvard, MIT, Princeton, Stanford, and Yale
Tier 2 (ranked 7–15):	Columbia, Michigan, Minnesota, Northwestern, Pennsylvania, Rochester, California-Berkeley, California-Los Angeles, and Wisconsin-Madison
Tier 3 (ranked 16–30):	Illinois-Urbana, Boston University, Brown, Cornell, Duke, Iowa, Maryland, Michigan State, New York University, North Carolina, Texas-Austin, Virginia, California-San Diego, University of Washington, and Washington University-St. Louis

Source: "The American Economic Association Graduate Study in Economics Web Pages", accessed 2011-02-08, http://www.vanderbilt.edu/AEA/gradstudents/

2.3.1 Economist Sample Selection

The bases of our dataset are the names, graduation years and PhD granting institutions of 13,624 economists who graduated from the top 30 US universities from 1955 to 1994. This data is obtained from the American Economic Association's (AEA) yearly "List of Doctoral Dissertations in Economics", which was published in the Papers and Proceedings issue of the "American Economic Review" until 1986 and in the "Journal of Economic Literature" thereafter. We supplement this information with the tier of the degree granting university according to the ranking of the National Research Council.

2.3.2 Career Choice and Productivity Measures

We add an "academic" indicator which takes the value one if the economist was a faculty member of a US economics, business or finance department in 2001 or listed as a member of the American Economic Association, and zero otherwise.

The US faculty directories are compiled by James R. Hasselback and made available on his webpage.[13] AEA Membership data is obtained from the American Economic Association Directory of Members in 1970, 1974, 1981, 1985, 1989, 1993, 1997, 2003 or 2007.[14] AEA membership serves as a proxy for faculty membership

[13] We only have access to the faculty listing of 2001. Therefore it is unlikely that graduates from before 1965 are included because they are retired by 2001. This biases our estimates if the retirement age is systematically higher or lower for recession cohorts compared to boom cohorts, which seems unlikely.

[14] An individual is classified as a member of the AEA if he appears in any of the membership lists from the year of his PhD graduation onward. It was pointed out to us that many PhD candidates become AEA members when they go on the job market although they eventually do not enter the academic sector and never renew their membership. Our results are robust to this concern. For

Table 2.3: Ranking of Journals in Different Decades.

Rank	Journal (ordered by 2000 rank)	1960	1970	1980	1990	2000
1	The Quarterly Journal of Economics	65.6	16.2	41.6	58.1	100
2	Econometrica	46.6	31.6	78.4	96.8	68.7
3	Journal of Economic Literature	-	100	100	18.8	63.5
4	The Review of Economic Studies	100	30.7	40.7	45.2	54.3
5	Brookings Papers on Economic Activity	-	96.9	15.9	0.7	51.5
6	The Journal of Political Economy	63.5	59.1	63	65.2	49.8
7	Economic Policy	-	-	-	-	45.7
8	Journal of Labor Economics	-	-	15.4	12.8	45.5
9	The American Economic Review	93.3	34.5	40.2	100	39.9
10	The Journal of Economic Perspectives	-	-	23.3	34.3	39.8
11	The Review of Financial Studies	-	-	-	-	39.2
12	Journal of the European Economic Association	-	-	-	-	38.6
13	The RAND Journal of Economics (Bell Journal of Economics)	-	39.5	40.2	11.4	38.2
14	The Journal of Finance	37.8	14.6	34.1	34.1	31.1
15	The Review of Economics and Statistics	59.8	12.4	6.5	28	21.7
16	Journal of Business & Economic Statistics	-	-	7.9	38.4	20.8
17	The Economic Journal	47.5	28	23.9	20.7	20.5
18	Journal of Applied Econometrics	-	-	-	16.6	19.1
19	Journal of Money, Credit and Banking		18.5	22.1	18.6	18.6
20	The World Bank Economic Review	-	-	-	5.7	18.5
21	International Economic Review	35.1	19	12.3	23	18.4
22	IMF Staff Papers	-	-	-	5.1	18.3
23	Journal of Law, Economics, & Organization	-	-	-	4.1	16.1
24	Journal of Law and Economics	51.8	43.3	33.1	3.9	14.1
25	The Journal of Human Resources	-	13.6	4.6	21.3	13.4
26	Journal of Population Economics	-	-	-	2.41	10.6
27	The Scandinavian Journal of Economics	2.5	7.1	2.1	10.7	9.2
28	The Journal of Business		18.5	37.4	8.7	8.7
29	The Journal of Industrial Economics	14.9	16.4	16	3.85	8.7
30	The World Bank Research Observer	-	-	-	0.9	8.5

NOTE.—These are the first 30 out of 74 journals. The rankings for the 1960s, 1970s and 1980s are taken from Laband and Piette (1994) and the ranking for the 1990s is from Kalaitzidakis, Mamuneas, and Stengos (2003). For the 2000s, we normalize the current discounted recursive impact factors ranking from the IDEAS RePEc website (`http://ideas.repec.org/top/top.journals.rdiscount.html`, last accessed 2011-02-07) to make it comparable to the other rankings.

outside of the United States, because Hasselback's faculty directories strongly focus on US colleges and feature only very few foreign institutions.

In order to compare the oeuvres of different economists over time we calculate a consistent measure of publication productivity. For all economists in our sample, we collect the publication records in the first ten years after their graduation, multiply each publication of an author by its weight ("publication points") according to a dynamic journal ranking, and divide it by the number of coauthors of the paper. We then sum up all these contributions within the ten years after graduation to obtain a productivity measure for every individual in our sample.

More specifically, we match the PhD graduates with their publications (including journal title, number of pages and the number and identity of co-authors) in 74 journals listed in JSTOR, a leading online archive of academic journals. We select all journals contained in JSTOR for which a ranking was available. Thus we include all major publications in economics and finance except the journals published by Elsevier, most notably the "Journal of Monetary Economics" and the "Journal of Econometrics".[15] To ensure comparability among researchers, we restrict our attention to the first ten years after graduation. JSTOR currently only provides full publication data up to the year 2004. With the ten year requirement we can thus rightfully analyze the sample from 1955 to 1994 without placing younger researchers at a disadvantage.

Comparing the value of the collected publication records for different researchers over the decades is difficult because the relative impact of economics journals has changed substantially over time (Kim, Morse, and Zingales 2006). Therefore, we construct a dynamic journal ranking with decade specific publication points for each journal from 1950 onwards. For the period from 1960 to the 1989, we use the ranking from Laband and Piette (1994), for the 1990s the equivalent ranking published in Kalaitzidakis, Mamuneas, and Stengos (2003), and for the 2000s the recursive discounted ranking available on the "ideas" webpage. For the 1950s we were not able to find a journal ranking and thus decided to extrapolate a ranking for articles published in the 1950s from our 1960s ranking. Table 2.3 lists 30 out of the total 72 journals with their associated publication points over time.

In Appendix B.4.1, we show that our results are robust to the use of various

example, they are very similar if we measure AEA membership from three years after the PhD.

[15]Because we do not believe that either recession or boom cohorts systematically prefer or dislike Elsevier journals, this should be of no consequence.

other productivity measures.

2.3.3 Macro Data and PhD Entry Date

The main aim of our study is to relate the career decisions and the publication success of economists to a proxy for the state of the macroeconomy at the times of application to and graduation from their PhD program. As our data contains only person-specific graduation dates, we infer the application date by subtracting the median duration of a PhD of 6 years from the graduation date.[16]

Using a fixed duration of the PhD—both in boom and in recession—to infer the application date, implies that students do not systematically time their graduation depending on the business cycle. This is the same assumption as in Oyer (2006) and we find (in line with his results) that the number of graduating PhDs is not correlated with the business cycle in publicly available NSF data. We discuss what would happen to our results if this assumption is violated in Appendix B.3.[17]

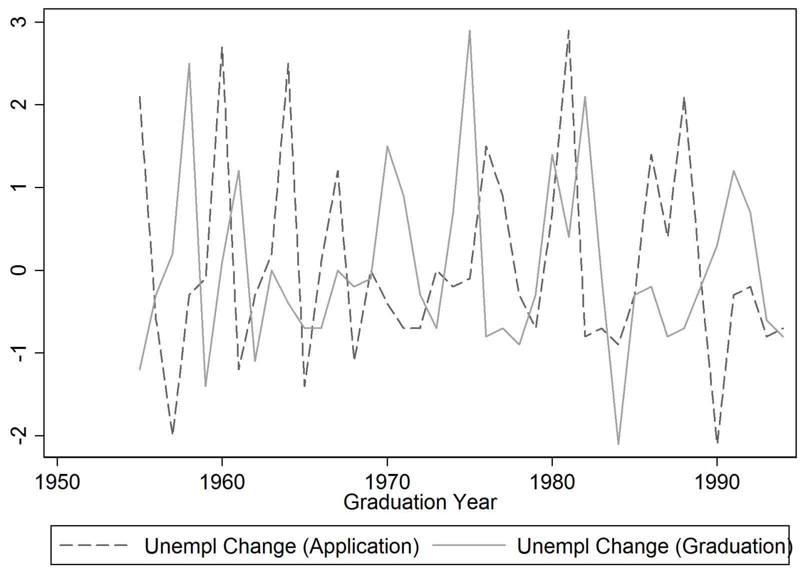
Our preferred proxy for the state of the business cycle is the change in the rate of unemployment from June of the preceding year to June of the considered year. The National Bureau of Economic Research (NBER) recession indicators are arguably the most convincing measures of recessions. However, binary indicators cannot carry information about the state of the economy as fine as continuous measures. Unemployment change is such a continuous measure and—out of several candidate variables that are available for the entire sample period—it is the most strongly correlated with the NBER recession indicators. For example, Figure 2.4 shows that recessions go hand in hand with a large change in unemployment. Unemployment levels are high only after a recession. To demonstrate the robustness of our conclusions, we also estimate all our specifications using unemployment levels and GDP growth as explanatory variables.[18]

[16]The median duration of a PhD stayed almost constant at from five to six years since the 1970s (see Table B.7 in Appendix B.4.3).

[17]Furthermore using six years for graduates is a potential problem for the precision of our estimates because the variation in completion times across PhDs is substantial. Section B.4.3 in the appendix reruns our main regressions using the distribution of completion times for the 1997 graduating cohort. The results become stronger, which suggests that measurement error in the business cycle at application potentially biases our estimates.

[18]We refrain from using some more business sector- or economist-specific measures of the state of the business cycle because they are generally not available for the entire study period. For example, Job Openings for Economists (JOE), a listing of open positions for economists published by the American Economic Assocation, is only available from 1976 onwards. Since our study period ends in 1994, using the JOE listings would reduce the length of our time series to 18 data points (minus six if we used job openings at application to the PhD as well).

Figure 2.4: Unemployment change



2.3.4 Aggregation to University-Year Level

Finally, we group our graduates' publication performances and the indicator for being an academic or not into university-graduation year averages. Thus, we reduce the number of our observations from 13,624 individuals who graduated from institutions in tiers one, two, and three between 1955 and 1994, to 1068 cohort means. Because we do not use any explanatory or control variables that vary below the university-year level, this grouping entails no loss of information.

2.3.5 Descriptive Statistics

Table 2.4 provides summary statistics for the PhD cohorts' average productivity, the average probability to become an academic, and the macroeconomic variation.

The average ten-year productivity of a university-year cohort is about 31.49 publication points. The average probability to become an academic is about 60% and is slightly falling over time as we can see in Figure 2.5a. Conditioned on being an academic, the average ten-year cohort productivity totals 48.14 publication points. This is about 50% of an article in the AER in the 1990s.[19]

Figure 2.5b depicts the average productivity of the PhD cohorts for every year

[19] In order to translate these publication points in terms of articles in a certain journal, one has to take into account that the importance of journals changes over time. For example, an article in the American Economic Review (AER) in the 1990s was worth 100 publication points while it was "only" worth 40.2 points in the 1980s (see Table 2.3). Therefore, the average ten-year productivity of a member of a university-year cohort in the full sample is about the equivalent of one-third of an AER article in the 1990s.

in our analysis, distinguishing between the average productivity of all graduates and graduates that became an academic. As expected, we see that the performance measures move together to a substantial degree.

The change in the unemployment rate, our preferred independent variable, has a mean value of approximately zero. The 10% quantile is -0.9 percentage points and the 90% quantile is 1.5 percentage points for the change in the rate of unemployment. The average unemployment level is 6.1% and the average GDP growth is 3.4%. From 1955 to 1994 the US was in recession 17% of all years. As an example, Figure 2.5c plots the change in the unemployment rate and in GDP growth together with indicators for recessions from 1955 to 1994.

Table 2.4: Summary Statistics

	mean	sd	min	max	p10	p90
Productivity	31.41	84.78	0.00	1738.10	0.00	93.60
Productivity (Academic)	47.96	103.66	0.00	1738.10	0.00	144.70
Academic	0.60	0.49	0.00	1.00	0.00	1.00
Unempl Change	0.02	1.03	-2.10	2.90	-0.90	1.50
Unemployment	6.11	1.50	3.50	9.70	3.80	7.70
GDP Growth	3.37	2.29	-1.94	7.20	-0.23	6.42
Recession	0.17	0.37	0.00	1.00	0.00	1.00
Observations	13651					

2.4 Results

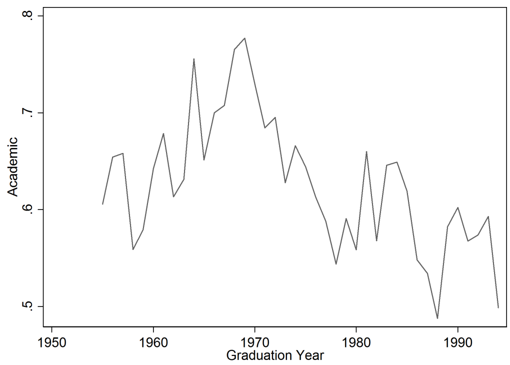
In this section we examine the empirical predictions derived from the modified Roy model. To do this, we estimate the following model in three different specifications:

$$q_{i,t} = \beta \cdot y_{app,t} + \gamma \cdot y_{grad,t} + \delta \cdot \text{controls} + \epsilon_{i,t} \tag{2.2}$$

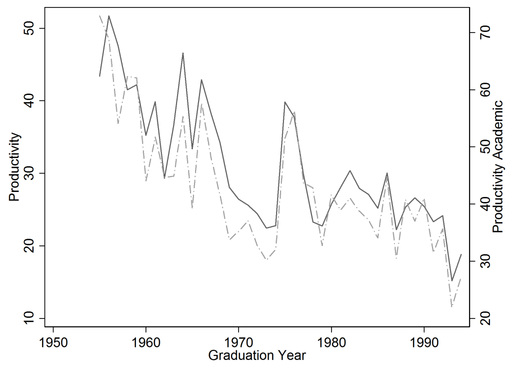
In the first specification, the outcome variable $q_{i,t}$ is the average publication output of a cohort of graduates from university i in year t. In the second specification, it is the average propensity to decide in favor of an academic career after the PhD, and in the third specification, $q_{i,t}$ is the average productivity of those who stayed in academia after the PhD. The unit of observation in all three cases is the average of a given university in a given year, weighted by the number of underlying individual observations. Moreover, the standard errors are clustered on the graduation year level, in order to allow for contemporaneous correlation between the outcome

Figure 2.5: Dependent and independent variables over time

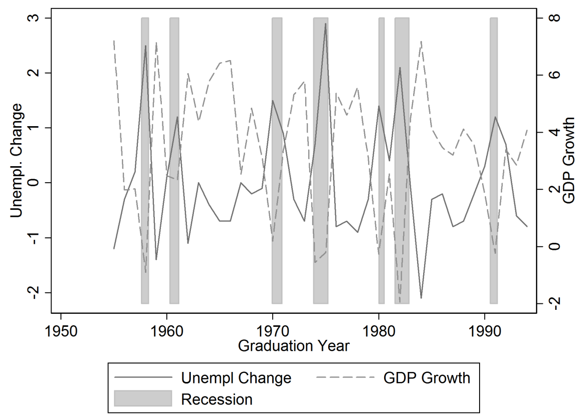
(a) Probability of becoming an academic



(b) Research productivity



(c) Unemployment change and GDP growth



variables in the presence of regressors that do not vary within a given year.

The regressors $y_{app,t}$ and $y_{grad,t}$ are a measure of the business cycle at application and at graduation for each cohort. Our preferred regressor is the change in the unemployment rate. To show the robustness of our results we also estimate all specifications with unemployment levels, GDP growth and NBER recession indicators as measures of the business cycle. For conciseness, we focus our interpretation on the effect of unemployment change on our dependent variables and only highlight if differences arise from using one of the other measures. As control variables, we include dummies for the full set of interactions of university and graduation decade. These dummies pick up the (changing) quality differences of PhD education among universities over time and they control for the higher standards of publication in recent decades (e.g. Ellison 2002a, Ellison 2002b). Additionally, we report regressions controlling for a time trend in Appendix B.4.5 and controlling for academic subfields in Appendix B.4.6.

We estimate Equation (3.2) using linear regressions. To identify the average treatment effect of the business cycle measure on the respective outcome variable, we assume that the productivity and the career decisions of a cohort of (potential) PhD economists do not contemporaneously affect the business cycle in a given year. This assumption excludes potential reverse causality. Furthermore, no third factor is allowed to directly influence the business cycle and the career decisions as well as productivity. This means that our parameters are identified by the arguably exogeneous variation of the business cycle.[20] To be able to interpret β and γ exclusively as the causal parameters of the selection effect discussed in the theory section, we need an additional exclusion restriction to be satisfied: we assume that unemployment change affects a cohort's career decisions and publications only in terms of changing their choice of the sector to apply to (the selection effect).

The last assumption might not be strictly true for the business cycle at graduation, because the state of economy affects an economist's first job placement and the first placement in turn influences productivity (Oyer 2006).[21] Therefore our estimate of γ is a combination of the selection effect and the placement effect. In contrast β, the estimated influence of the business cycle at application, measures cleanly the selection effect as we control in all regressions for the degree-granting university, i.e. for the placement to different PhD programs.

Table 2.5 summarizes the main regression results of the three specifications, each in one column. Every column contains four independent regressions using alternative business cycle measures for the two explanatory variables. The estimated coefficients of the different regressions are reported one below the other. The following subsections explain the results for the three outcome variables in turn.

2.4.1 Effect on the Publications of all PhDs

The first column of Table 2.5 shows the effect of the business cycle on the publication output of an average PhD graduate in the sample. Unemployment change, both at time of application and at graduation, has a significantly positive effect on research productivity at the five and one percent level, respectively. These two results are also economically substantial: a cohort on the 90% quantile of unemployment change at time of application is expected to achieve 3.7 publication points more than a cohort

[20] Consequently, we do not need any control variables to consistently estimate the coefficients. We nevertheless include them to increase the precision of our estimates.

[21] We explain in Section 2.4.3 that given Oyer's result we might actually underestimate the causal effect of selection in our regressions.

Table 2.5: The Main Regression Results

	Productivity	Academic	Productivity
Unempl Change (Application)	1.50**	-0.89	3.23***
	(0.67)	(0.58)	(0.96)
Unempl Change (Graduation)	2.31***	1.36**	2.72**
	(0.65)	(0.61)	(1.20)
Unemployment (Application)	1.54**	-0.75	2.94**
	(0.65)	(0.79)	(1.11)
Unemployment (Graduation)	1.78**	-0.24	3.04**
	(0.74)	(0.59)	(1.26)
GDP Growth (Application)	-0.65**	0.47*	-1.45***
	(0.29)	(0.24)	(0.43)
GDP Growth (Graduation)	-0.70**	-0.41	-0.74
	(0.33)	(0.27)	(0.56)
Recession (Application)	2.08	-3.25**	5.28*
	(2.11)	(1.55)	(2.95)
Recession (Graduation)	4.49**	2.16	4.96
	(2.15)	(1.28)	(3.58)
Subsample	All	All	Academic
University-Decade Dummies	Yes	Yes	Yes
Observations	1068	1068	1047

NOTE.—Standard errors clustered on the graduation year in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

on the 10% quantile. This is approximately 12% of the mean. Similarly, if we do the same calculation for the graduation cohort, the difference is 5.5 points, which is 17.6% of the mean.[22]

Using the alternative measures of the business cycle as regressors deliver qualitatively similar results as unemployment change. A higher unemployment rate is associated with higher productivity at application and at exit. Positive GDP growth leads to a lower publication productivity and NBER recessions go hand in hand with more publication success. All coefficients are statistically different from zero at the five percent level. The only exception is the estimated coefficient for NBER recessions at application which is not significant at conventional levels.

Therefore, the effect of the business cycle at graduation is in line with empirical prediction 3: PhDs who graduate during a recession publish more on average. As

[22]Referring to Table 2.4 above, the difference between the 10% and the 90% quantiles of unemployment change at time of application is 2.4. Multiplying this by the parameter estimate of 1.54 gives a difference in average productivity between "boom" and "recession" cohorts of 3.7 publication points. Referring to Table 2.3, this is about the number of publication points one gets assigned for an article in "Economica' ' during the 1990s. From Table 2.4, we also find that the "average" PhD graduate achieves 31.49 publication points. Similarly, multiplying the difference between the 90% and 10% quantile of unemployment change with the paramter estimate of 2.31 at graduation yields 5.549 publication points. This is about 17.6% of the mean of 31.49.

noted above, the measured overall effect might be a combination of three effects: an "extensive margin" effect whereby more PhDs stay in academia, an "intensive margin effect" whereby a better selection of PhDs stay in academia and a "placement effect".

The theory does not make a prediction which overall effect the business cycle at time of application should have on the publication output of an average PhD graduate. On the one hand, according to Proposition 2.2.1, graduates who experienced a recession at time of application constitute a better selection of individuals. On the other hand, according to Proposition 2.2.2, fewer of these individuals are expected to stay in academia and publish after the PhD. Empirically, it seems that the former effect dominates the latter, as a worse business cycle (measured by a large positive change in the unemployment rate, a higher unemployment rate or lower GDP growth) at time of application is associated with a higher publication output of an average PhD.

2.4.2 Effect on Career Decisions

The second column of Table 2.5 reports how the business cycle is related to economists' career decisions after the PhD. Individuals are more likely to stay in academia when the economy is ailing according to our preferred business cycle measure of unemployment change at graduation. The estimated coefficient is significant at the five percent level. The mean estimates point in the same direction for two of the three alternative measures, but they are not significantly different from zero on conventional levels.

These findings give qualified support for empirical Prediction 3 from the theory section: PhD graduates are more likely to stay in academia if there is a recession at graduation. At least part of the increased average output of a recession cohort could therefore come from this "extensive margin" effect. Taking the mean estimates for unemployment change literally, a member of the cohort on the 90% quantile of unemployment change at graduation (+1.5%) has a 3.24 percentage points higher probability to become an academic compared to a PhD student graduating on the 10% quantile (-0.9%). The average propensity to become an academic is 60%.

The theory also predicts that economists who experience a recession at application to the PhD are less likely to stay in academia afterwards because some of them will enter only *because of* the recession (Prediction 1). The evidence in Table 2.5 suggests the existence of this effect. The estimated coefficient for unemployment change is of the predicted sign but not statistically different from zero. Also the parameter estimates of all other measures are of the predicted sign. For GDP growth and recession indicators they are significantly different from zero at the ten and the five percent level, respectively.

More generally, we are not sure how to measure the decision between academia and business correctly. We think three different concepts of someone being an "academic" are conceivable: First, one could only consider faculty members of higher learning institutions as academics. This definition leaves out staff at international organizations, central banks and other research-focused (governmental) institutions. Second, one could argue that the relevant distinguishing characteristic of an academic is producing novel and original research. And finally, one could more generally consider anyone an academic who works on research-related topics and upholds a relationship with the academic community.

The evidence reported in Table 2.5 is based on the third notion of an academic by classifying anyone as such who is either a faculty member or a member of the American Economic Association (AEA) after the PhD. Table 2.6 additionally reports the measures of being an academic according to the first two notions.

Column two in this table shows the propensity to become an academic measured by whether graduates end up as members of faculty at an economics, business or finance department of a college or university in the United States according to the listings published by Hasselback (2001). The direction of the effect is the same as in column one and in the main results table except for unemployment levels. However, the resulting coefficients are mostly not statistically significant for either point in time. This might be the case because the employed faculty listings are US focused and not exhaustive.[23]

Column three defines an academic as an individual who, according to our data, publishes at least one article in a ranked scientific journal after his or her PhD. The estimated effect for the business cycle at application points in the predicted direction for three out of four measures. The estimated coefficients are significantly

[23]For example, faculty on leave are not included and we do not have faculty directories for other departments, such as law and agriculture. Furthermore, our faculty listings are strongly focused on US institutions. Thus, they miss many foreign graduates who become professors in their home countries and are members of the American Economic Association.

Table 2.6: Different Measures for Being Classified as an Academic

	Academic	Faculty	Publish	Academic
Unempl Change (Application)	-0.89	-0.43	-0.91*	-1.72***
	(0.58)	(0.47)	(0.48)	(0.58)
Unempl Change (Graduation)	1.36**	0.53	0.45	2.87***
	(0.61)	(0.41)	(0.40)	(0.94)
Unemployment (Application)	-0.75	0.08	0.09	-1.23
	(0.79)	(0.38)	(0.40)	(1.03)
Unemployment (Graduation)	-0.24	0.60	-0.00	-0.07
	(0.59)	(0.36)	(0.40)	(0.92)
GDP Growth (Application)	0.47*	0.25	0.42**	0.75**
	(0.24)	(0.19)	(0.19)	(0.29)
GDP Growth (Graduation)	-0.41	-0.04	0.03	-1.25***
	(0.27)	(0.19)	(0.23)	(0.36)
Recession (Application)	-3.25**	-1.43	-1.65	-5.73***
	(1.55)	(1.06)	(1.23)	(1.73)
Recession (Graduation)	2.16	1.82**	1.26	3.95**
	(1.28)	(0.76)	(0.87)	(1.67)
Subsample	All	All	All	Tier 1
University-Decade Dummies	Yes	Yes	Yes	Yes
Observations	1068	1068	1068	234

NOTE.—Standard errors clustered on the graduation year in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

different from zero on the five percent level for unemployment change and for GDP growth. The business cycle at graduation is weak and not significant for any of the independent variables.

Column four in Table 2.6 also shows regressions for the propensity to become an academic (according to our preferred academic measure) for a subsample of graduates from the six top-ranked universities, i.e. the tier one schools. The effect here is significant at least on the five percent level and in the predicted direction for three out of the four business cycle measures. We interpret this as evidence that it is actually the individuals at the very top of the skill distribution which are most able to successfully switch back and forth between academia and business and who thus possess what one could call general skills.

Overall, we conclude that the results lend support to the predictions made by our theory about the career decisions of PhD graduates.[24]

[24]There might be concern about the behavior of foreign students over the business cycle. We discuss this issue in Appendix B.4.4

2.4.3 Effect on the Publications of Academics

The last column of Table 2.5 shows the results of regressing the publication output of individuals classified as academics on our four business cycle measures. The results here are largely robust to the sample selection according to any of the three definitions of an academic that were discussed above.

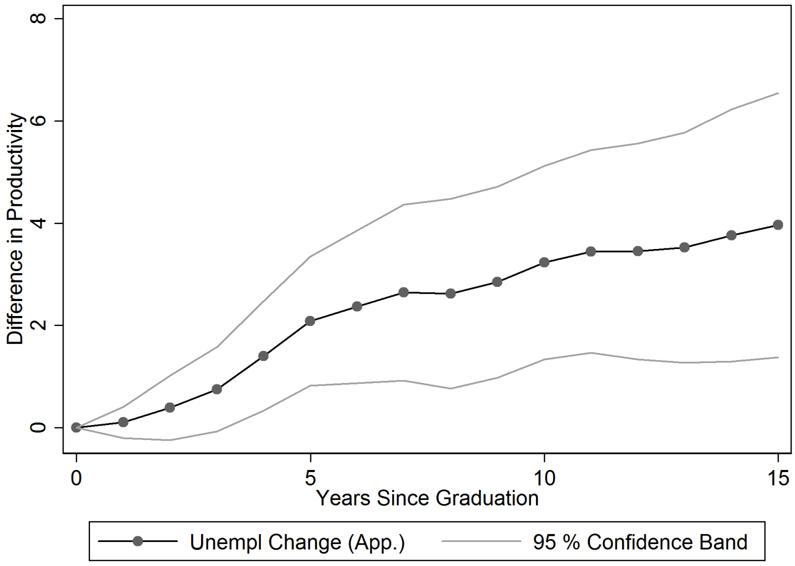
For all the different measures, the productivity of academics who experienced a recession at time of application is significantly higher than that of academics who applied during a boom. The coefficient is significant at the one percent level for unemployment change and of economically relevant magnitude: comparing the average member of the cohort on the 90% quantile of unemployment change at time of application to a cohort member on the 10% quantile, the former is on average 10.47 publication points better than the latter. This is about 16% of the mean.[25]

Column three in Table 2.5 reports the regression coefficients of our measures of the business cycle on PhD cohorts' cumulative publication success of an academic over the first ten years after graduation. In order to obtain a more dynamic picture of the business cycle's effect, Figure 2.6 plots the coefficients for unemployment change at application that we would have obtained if our cumulative publication measure had been defined for each out of 1 up to 15 years after the PhD instead of just year 10. According to this figure, the effect of the business cycle at application on productivity is truly long term: the publication gap between academic cohorts who experienced a recession versus a boom at application widens monotonously over time—although the slope seems to slightly flatten after year five or seven.

These findings are in line with Prediction 2 which states that the selection of PhD entrants is better during economically difficult times and that this better selection persists to the PhD graduates who stay in academia. In fact, Prediction 2 states that a generic recession at time of application cohort should first order stochastically dominate a generic boom at time of application cohort with respect to academic skill. Therefore, not only the mean but the whole distribution of academic skills should shift to the right if the economy worsens. Table 2.7 shows the effect of the business cycle on the distribution of publication output within each cohort using quantile regressions. The unit of observation is now an individual academic's publication

[25]The 10% quantile of unemployment change at time of application is -0.9 percentage points, the 90% quantile is 1.5 percentage points and the difference is therefore 2.4 percentage points. Multiplying this difference with the mean estimate of 3.27 yields 7.86. The mean productivity for an academic is 48.14 publication points.

Figure 2.6: Long-term effect of the unemployment change at graduation on the cummulative publication output of the average academic



NOTE.—The figure shows coefficients from regressing the (cumulative) publication productivity of an academic for different time-spans on the change in unemployment rate at application, controlling for the change in unemployment rate at graduation and university-graduation decade fixed-effects. Since we only observe 15 years of publication history for cohorts graduating before 1990, we use correspondingly shorter publications histories for academics graduating after that.

output.[26] Among those PhDs who are considered academics according to our "academic" measure, 45 percent do not publish at all. We therefore restrict Table 2.7 to the effect of the business cycle on the median of the publication distribution and above.

The estimates are in the predicted direction and significant for the upper quantiles of the publication distribution, but they become less significant for the lower quantiles. The reason for this is probably that the "academic" measure is not perfect at separating academics who do not publish from individuals who have left academia after the PhD. We know that there are more such individuals among the recession at application cohort, some of which are thus mistaken as low-skill academics. This downward-biases the difference between the publication distributions, most strongly so at the lower quantiles.[27]

[26]We only control for university tier–graduation decade fixed effects and their interactions here, because the quantile estimation becomes much less reliable with a large number of dummy controls. The standard errors are not clustered on the graduation year level as this is not straightforward to implement with quantile regressions.

[27]If we define an academic according to whether he publishes in a ranked journal instead of AEA membership or appearance in a faculty listing, and thus condition on non-zero publications, our quantile regressions yield positive and significant effects of unemployment change in line with the theory over the entire publication distribution.

Table 2.7: Quantile Regression for the Academic Subsamples

	50%	65%	80%	95%
Unempl Change (Application)	0.00	0.31	3.66*	9.48*
	(0.49)	(1.01)	(2.06)	(5.70)
Unempl Change (Graduation)	-0.00	1.24	3.73*	0.70
	(0.51)	(1.05)	(2.14)	(5.93)
Unemployment (Application)	-0.00	0.69	4.02*	11.56*
	(0.53)	(1.16)	(2.25)	(6.14)
Unemployment (Graduation)	-0.00	2.71**	5.09**	12.13**
	(0.49)	(1.06)	(2.06)	(5.61)
GDP Growth (Application)	0.00	-0.32	-1.55*	-4.80*
	(0.22)	(0.46)	(0.93)	(2.61)
GDP Growth (Graduation)	0.00	-0.06	-1.05	1.65
	(0.23)	(0.48)	(0.97)	(2.75)
Recession (Application)	0.00	1.03	6.90	17.12
	(1.39)	(2.92)	(5.90)	(16.03)
Recession (Graduation)	0.00	4.87*	7.55	-0.93
	(1.37)	(2.87)	(5.80)	(15.77)
Subsample	Academic	Academic	Academic	Academic
Tier-Decade Dummies	Yes	Yes	Yes	Yes
Observations	8248	8248	8248	8248

NOTE.—Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.5 also reports the effect of the business cycle at graduation on the research productivity of academics. According to the evidence in section 2.4.2 somewhat more PhDs enter an academic career if there is a recession at graduation. Without a specific assumption on the distribution of skills of PhD economists, our theory does not make a prediction whether the additional academics who enter at the "extensive margin" are of higher or lower academic skill than the average of those graduates who always decide to stay in academia after the PhD. However, as discussed in the theory section, if the number of spaces in academia is not completely flexible at graduation, there is also an "intensive margin" effect which stems from a higher competitiveness to enter academia. This improves the composition of talent in academia for cohorts who faced a recession at graduation.

The empirical results in Table 2.5 are consistent with this idea. The estimated coefficients are significant at the five percent level for unemployment changes and levels. They are not significant but in the right direction for GDP growth and the recession indicators. According to our estimates, an academic graduating on the 90% quantile of unemployment change is on average 6.67 publication points better than an academic graduating on the 10% quantile. This is about 13% of the mean of 48.14. Moreover, if there is an "intensive margin" effect at graduation, it should be weaker for the elite tier one universities whose students may virtually always be able to get an academic job if they want to. Indeed, table B.6 in the appendix shows that the allocation response over the business cycle is stronger for the tier one group and the productivity effect on tier one academics is insignificant.[28]

Yet the size of the estimated coefficient for the business cycle at graduation (in contrast to the coefficient for the business cycle at application) should be interpreted with caution. As noted above, we are measuring at graduation only the composite effect of the selection effect and a placement effect within academia. According to Oyer (2006), students who graduate in a recession receive a worse placement, which in turn results in fewer publications. This suggests that our estimated coefficient at graduation is biased towards zero and we are underestimating the quantitative importance of the selection effect of the business cycle at graduation.[29]

2.5 Conclusion

Recent studies have shown that aggregate labor market conditions can have strong and persistent effects on individuals' careers via the choice of initial jobs. Our article investigates the implications of this result on the composition of talent and thus on productivity in a sector. To guide our empirical study, we build a Roy-style model of occupation choice over the business cycle if the number of workers is fixed in one occupation and if it is flexible: In the first case, the quality of talent in a relatively more stable industry increases in recession while in the latter case its size rises and a quality change depends on the distribution of talent. In the market for economists we find that recessions indeed increase the publication output of the academic sector for the long term by altering the allocation of talent between academia and business. Depending on whether human capital is drawn to sectors with low or high social value during recessions, this reallocation effect might reinforce or cushion the massive harm done by a downturn. Moreover, our results indicate that it is easier for the

[28]The results at graduation could also be driven by graduation timing: if bad students systematically delay their graduation, it might result in a positive effect of recessions on publications. In that case, however, the estimated effect of the business cycle at application is underestimated and the true effect is even larger. Nevertheless, there is systematic evidence that by and large no graduation timing takes place: Oyer (2006) finds that there is no correlation of the business cycle and the size of the graduating cohort. Similarly we find no correlation between graduation numbers and our business cycle indicators. This result is reported and discussed in Appendix B.3.

[29]In Appendix B.4.7 we provide empirical evidence for the direction of this bias.

public sector to attract talented workers during recessions.

A contribution of our article is the theoretical model about talent selection for absolute and comparative advantage, which gives predictions about talent flows and the resulting sectoral productivity under both regimes. Comparative advantage has been dominant in the literature on the Roy model, but absolute advantage, when the number of jobs is fixed, seems equally relevant: for example, Groes, Kircher, and Manovskii (2010) show that the dynamics of occupational mobility are explained better by absolute than relative advantage. Borjas and Doran (2012) find that the number of faculty positions in mathematics are limited such that a supply shock can push less talented mathematicians into lower-ranked departments or out of the research community. In his survey of assignment models, Sattinger (1993) argues that absolute advantage selection often occurs when the resources that workers utilize for production are scarce and thus have an opportunity cost.

The empirical part of our study provides direct evidence for the impact of the business cycle on output and productivity due to the reallocation of talent. This is possible by considering the particular labor market for PhD economists, which provides a measure of productivity with two important advantages: first, publications are largely exogenous to the business cycle while wages and output are directly affected by recessions. Second, publications are easy to attribute to specific individuals and we can assess their quality quite well. In contrast, output in most other sectors is the result of a collaborative process and therefore it is hard to disentangle the individual (cohort's) contribution. The other specific feature of our setting is its two step selection process with competitive admission and the academic versus non-academic career choice six years later. At first glance this seems quite unique. However, early careers in other knowledge-intensive industries are not completely dissimilar: for example, starting positions in law or consulting firms feature an informal training phase of several years with a performance appraisal and promotion decision at the end.

If our findings generalize to the labor market as a whole—in this study we are only looking at a particular labor market at the top of the skill distribution—is left for future research. An encouraging step in this direction is the recent paper by Shu (2012), which finds that MIT graduates produce more patents if they graduate during a recession. Her paper also uses a Roy-style model of occupational choice. This points to a broader applicability of our ideas beyond the market of PhD economist.

However, sensible measures of productivity are much harder to come by in sectors that are not as transparent and individualized as science.

Chapter 3

Concentration Versus Re-Matching? Evidence on the Locational Effects of Commuting Costs

Concentration Versus Re-Matching? Evidence About the Locational Effects of Commuting Costs

Michael J. Boehm *

March 2013

Abstract

Using administrative employer-employee data from Germany, I exploit two reductions of tax breaks for commuting in 2003/4 and 2006/7 to estimate commuting costs' effect on the decision to switch job and move house. Standard theory predicts that higher commuting costs should lead to increased concentration in urban centers. However, I find that re-matching of existing jobs and houses to reduce commuting distances is much more prevalent in the data. With these estimates I calculate the effect of a complete abolition of the tax breaks on overall travel distance, fuel usage, greenhouse gas emissions, the tax base, and the de-population of the countryside.

Keywords: Work/Residence Location Choice; Commuting Costs; Environmetal Effects of Tax Policy; Employer-Employee Data

JEL Classification Numbers : R00; J61; J68; Q48; Q58

---

* London School of Economics and Centre for Economic Performance; Houghton Street, London WC2A 2AE, UK; Email: m.j.boehm@lse.ac.uk; Tel: +44 751959 4464. I am grateful to Joern-Steffen Pischke, Luis Garicano, Daniel Sturm, Henry Overman, Stephen Redding, Guy Michaels, Gabriel Ahlfeldt and seminar participants at LSE for comments and suggestions. I would also like to thank the Institute for Employment Research in Nuremberg (IAB) for their very friendly and helpful support. All errors and omissions are my own.

3.1 Introduction

There is an ongoing debate about what are the factors that shape our cities and determine the use of the land. While theorists in urban economics and economic geography have developed many interesting models and predictions, causal empirical evidence is relatively scarce. More specifically, many theories are concerned with the effect of transport costs on the concentration of individuals and firms within and between cities, but much of the existing empirics is either descriptive or anecdotal.[1] In this paper I provide causal evidence on the effect of an increase in a particular component of transport costs - commuting costs - on individuals' location decisions.

I estimate the effect of two reductions in per kilometer tax breaks for commuting in Germany on employees' decisions to move house and to switch jobs. Further, I analyze whether these moves or switches leave employees with a shorter commute and whether they make them locate their residences or jobs in concentrated locations, i.e. cities. Theoretical models in urban economics and economic geography are unanimous in their prediction that higher transport or commuting costs will lead to increased concentration of economic activity and population. However, these models are concerned with the long run general equilibrium, while from a public policy point of view we are also interested in the effect that occurs within a couple of years.

A convincing alternative hypothesis to concentration predicts that higher commuting costs will lead to a substantial amount of "re-matching", which is the process of occupying existing residential units and jobs differently but not changing their relative supply or utilization. Assume that some individuals live in location A and work in B while others work in A and live in B. If commuting costs rise, it becomes more attractive to live *and* work in *either* A *or* B and some (pairs of) individuals may now find it optimal to re-match in order to reduce their commuting distance. A precondition for this hypothesis is that individuals are heterogeneous in their residential preferences and in their productivity in different jobs. I investigate in how much the response to the tax break changes fits the concentration and in how much it fits the re-matching hypothesis.

I find that individuals strongly react to higher commuting costs by switching jobs and moving residence. In my preferred regressions, a 1000 euros reduction

[1] For a descriptive study, see Bento, Cropper, Mobarak, and Vinha (2005). It has also been argued that low fuel prices are the main reason for the spread-out shape and the lack of viable public transport of American cities compared to European ones (Krugman 2008).

in commuter tax breaks per year, which amounts to about a 300 euro increase in taxes paid for the average tax payer, leads to a 2.7 percent increased probability to change job location on average and it leads to a 1.7 percent increased probability to change residential location. The probability to change job or residence such that the commuting distance becomes shorter even rises by 8.9 percent.

In addition, there is evidence for both of the theoretical hypotheses in the data. Individuals become equally more likely to switch jobs from rural to urban locations as from urban to rural locations. However, they become more likely to move house from rural to urban locations but they do not become more likely to move house from urban to rural locations. Thus, on the one hand, there is no further concentration of jobs in cities due to the tax break changes. On the other hand, individuals' residential moves clearly show signs of an increased population concentration in cities, much as the standard theory predicts. Yet, the concentration effect for residential moves is substantially smaller than the re-matching effect.

I also do not find an increase of residential property prices in urban areas, which suggests that it is not inelastic housing supply or occupation rates that prevent stronger population concentration, but that the tax break changes' effect on demand for urban property is weak in the first place. In addition, a more detailed look at the dynamics of the adjustment shows that the bulk of the effect happens within two years after the change in commuting costs. This suggests that the short run effects considered in this paper may not be too different from the long run effects that the standard theory is concerned with.

Finally, I calculate the effect of the commuting tax breaks on outcomes that are directly relevant for policy. Assuming that they only have an effect on journeys via commuting and abstracting from general equilibrium effects (on fuel prices and congestion, for example), the tax breaks' hypothetical complete abolition in 2003/04 would have led to about a four percent reduction in the overall number of kilometers traveled in Germany within a year. Due to this, fuel usage would have declined by 5.2 percent and emissions by 3.4 percent while the effect on the movement from rural locations to cities would have been negligible. Despite the behavioral response of switching to job-residence combinations that feature a shorter commute, the increase in the tax base would have been substantial.[2] These results are also informative

[2] I refrain from comparing the environmental gains to the welfare losses for workers who lose part of their commuting subsidies as this would require strong assumptions about the distribution function of alternative work-residence combinations.

about the potential effects of reductions in similar commuting subsidies that exist in several other advanced countries such as the Nordic- and the Low Countries, Austria, Switzerland, Italy, France, and Japan.

My results contribute to several discussions in policy and academia. First, they show that an increase in commuting costs leads to increased concentration of population in cities, but not of jobs. This relates to the debate about the effect of transport costs and commuting costs in economic geography and in urban economics, respectively. The results also show that the bulk of the adjustment is re-matching, which requires substantial individual heterogeneity. This relates to the thriving literature on local labor markets, an important assumption of which is heterogeneity. Moreover, I estimate the effect of the commuting tax break change on overall tax revenues, fuel usage, greenhouse gas emissions, commuting distance, and the depopulation of the countryside. This relates to the policy debate about the benefits and costs of the abolition of this subsidy to commuting. Finally, I show that German employees are on average more mobile in terms of their jobs than in terms of their residences, which can be interpreted as that they value their private lives more than their careers.

For my analysis I use a representative two percent sample of all German employees provided by the Federal Employment Agency. This quarterly updated panel features locational information on residence and employment at the municipality level. I supplement the information with geocodes for all of the ca 12,500 municipalities in order to calculate commuting distances for each individual in the sample. I exploit the fact that tax breaks of commuting were reduced substantially at the turn of 2003/4 and again in 2006/7 to estimate the effect of an increase in commuting costs on individuals' likeliness to switch job location, move residence location, switch such that the new combination features a shorter commute, and the commuting distance itself. Moreover, I estimate in how far the switches constitute relocations toward one of the 80 cities of more than 100,000 inhabitants in the country.

The paper is related to several strands of literature. First, I contribute to the ongoing debate about the determinants of land use and urban shape. The standard monocentric urban model (for example, see Brueckner 1987, Fujita 1989) predicts that higher commuting costs lead to an increased population concentration near the urban center and to a steeper rent gradient. Classic economic geography shifts the focus from locational decisions within cities to between cities (for a survey, see Moretti 2011), while the new economic geography literature has started to explicitly model transport costs and firms' locational decisions (e.g. Helpman 1998, Redding and Sturm 2008). In general, all of these models predict that higher transport or commuting costs should lead to more concentration in terms of population and of economic activity (i.e. firms). I provide causal evidence for the existence of this concentration effect for individuals' residences, but don't find evidence for it in terms of jobs.

Second, the more important component of the adjustment to the tax break changes is what I call re-matching. This constitutes evidence for substantial individual heterogeneity in terms of productivity in different jobs and in terms of preferences for residential locations. Heterogeneity of workers is an important assumption of the thriving literature on local labor markets (for example, see the model in Moretti 2011), but it was considered an important feature of realistic theoretical models even much earlier (e.g. Michel, Perrot, and Thisse 1996).

Third, there is a policy debate about the effect of gasoline, carbon, and public transport prices on travel demand, CO2 emissions, household location, commuting choice and economic activity.[3] I show that travel distance, CO2 emissions, and locational decisions are strongly affected by commuting tax breaks. In general, the tax break change experiment seems more suited to analyzing the locational impact of a change in transport or commuting costs than many other experiments: the decline in the value of a job-residence combination is independent of the chosen transport mode or of intensive margin responses, such as driving more fuel-efficiently. This lets me focus on the extensive margin response of moving house or switching job. Moreover, the tax break changes are arguably fully exogenous to the effect analyzed - they were decided as part of an across-the-board cut in subsidies due to a dire fiscal situation - while fuel prices or most public transport investments are endogenous to travel demand.

Fourth, there is a domestic German debate about the reasons for and against the abolition of the commuter tax breaks. The claims that such a policy change would have no effect or only long run effects on peoples' locational decision (e.g. Distelkamp, Lutz, Petschow, and Zimmermann 2008, Schulze 2009, Graeb and

[3]For example, see Ahlfeldt and Feddersen (2010), Bento, Goulder, Jacobsen, and von Haefen (2009), Knittel and Sandler (2010), Busse, Knittel, and Zettelmeyer (2009), and Li, von Haefen, and Timmins (2008). A paper that is methodologically related to mine is by Molloy and Shan (2010), who study the effect of gasoline prices on household location via commuting costs. They find that construction activity reacts strongly to changes in locations' relative attractiveness due to increased transport costs but, just as this paper, find no significant effect on house prices.

Vorgrimler 2005) are closest related to my analysis. I show that individuals do react to the decreases in commuting tax breaks and that they do it swiftly - to a substantial degree even before the changes are implemented. Moreover, I estimate their effect on the overall reduction in gasoline consumption, emissions, travel distance, and the tax base. Nonetheless, I make the theoretical argument that the tax breaks, when they are set at the right level, support efficient matching in the housing and in the labor market. Environmental and fiscal goals should be pursued using different policy tools, such as gasoline taxes.[4]

The fifth group of literature this paper deals with is concerned with job mobility and the migration decision (most notably, consider Topel and Ward 1992, Bartel 1979). I show that there is a causal effect of pecuniary (i.e. wage) changes for the current job-residence combination on mobility. Moreover, by observing the strength of the reaction in terms of moving and in terms of switching jobs, I provide evidence that the average employee values her residence (and thus her private life) more than her job.[5]

The remainder of the paper proceeds as follows. I explain the data in the next section. Then I present the theoretical hypotheses about the effect of a reduction in commuting tax breaks on an individual and an aggregate level as well as the empirical strategy. Section 3.4 reports and interprets the main results, and section 3.5 estimates effects on variables that are of immediate concern for policy makers. Section 3.6 concludes.

3.2 Data

3.2.1 German Employment Records

The data that I use are a two percent representative sample of the quarterly updated administrative records of all German employees, the so-called "BA employment panel", collected by the Federal Employment Agency.[6] Employers have to

[4] Knittel and Sandler (2013) estimate empirically the welfare losses of taxing a variable (commuting in our case) that is imperfectly correlated with an externality and find that they are substantial.

[5] Again, pinning down the rents from existing work places and residences over alternatives that might feature a shorter commute would require putting a lot of structure on the distribution of these combinations in the population and I refrain from it in this paper.

[6] The weighting of observations as in survey data is not necessary as the sample is representative of the population of German employees and there is no panel attrition in the sense that workers only disappear when they cease to be employed in actuality.

provide quarterly notices about their employees in order for the public administration to determine entitlements to unemployment insurance and the accumulation of retirement benefits. The information provided should therefore be highly accurate and up to date.

The data range from 1999 to 2007 with about 500,000 to 600,000 individuals per year. They include information on each individual's age, gender, education, income, and municipality of work and residence. They also include job and employer characteristics, such as industry sector and size of the employer and the worker's position (e.g. in training, regular worker, or foreman). Unfortunately, there is no information on home ownership, marital status, children, and place of birth provided. I also don't know which individuals are occupying a second home near their workplace from where they travel to work during the week.

Table 3.1 reports important descriptive statistics on the individuals in the data. A person stays about seven out of nine years in the dataset on average (so the panel is quite balanced), she switches job location about 0.9 times and house location about 0.7 times during that period.[7] Underlying these averages is substantial heterogeneity in all the characteristics across individuals as indicated by the reported distribution quantiles of the variables. Tables 3.3 and 3.2 show that there are about 12,500 municipalities of residence and 10,500 municipalities of workplace in the dataset. Of course, the size distribution between municipalities is very dispersed with the biggest municipalities being many times larger than the smallest ones. Moreover the tables report the number of connections per work or residence municipality, i.e. the number of different municipalities from- or to which workers commute, respectively.

I supplement the data by adding detailed geographic coordinates (geocodes) for each municipality, which I retrieved using Google Maps.[8] I use this information to estimate the commuting distance by car (termed cardistance in the following) for every individual in the dataset.[9] The cardistance is the relevant distance between residence and workplace because the actual tax breaks are calculated according to

[7]Throughout the paper, I use the female form to refer to females and males alike unless I specifically make clear that I'm referring to a female only.

[8]The geocodes were downloaded on 5 May 2010 from `http://www.gpsvisualizer.com/geocoder/` (last accessed 2013-03-11).

[9]The cardistance estimate was calculated following Einig and Puetz (2007) who use the crow fly distance between the midpoints of two municipalities and multiply the result by 1.3. The latter is the average ratio between car- and crow fly distance. Idiosyncratic mistakes in this approximation method should not matter given the large number of municipality combinations in the dataset. However, one may imagine a bias for municipalities of large (area) size and the under- or overestimation of the commuting distance for some of their residents or workers.

the fastest car distance, although they are granted independent of the actual means of transport used. Tables 3.4, 3.5, and 3.6 document the distribution of cardistance in the years 2002, 2004, and 2007. We see that the average cardistance has mostly been increasing over time, which is also documented in Grau (2009). Therefore, the identification strategy below will use year fixed effects to control for overall differences in switching behavior across years.

3.2.2 Commuter Tax Breaks

Tax breaks for commuting apply in most European countries including Germany.[10] As early as 1920, the actual costs of traveling to the workplace were acknowledged in Germany as income-related expenses and thus could be deducted from the income tax bill. Initially, only the cost of public transport was accepted, but in 1955 the federal constitutional court allowed each kilometer traveled by car to be deducted with 0.50 Deutschmark. From the year 2001 onward, a flat rate irrespective of the means of transport of 0.36 euros for the first ten kilometers and 0.40 euros thereafter applied.

Over the years, the reduction (or even the full abolition) of these tax breaks became a constant matter of political debate. The critics, often from the political left or liberals, argued that the tax breaks are environmentally- and fiscally damaging. The supporters, often politically conservative or with a mandate from a rural constituency, countered that the tax breaks support rural- and family life, and that they are enhancing mobility in the labor market because they allow individuals to travel longer distances to their workplace.

Real change on commuter tax break had to come for another reason—the dire fiscal situation that the country faced throughout the early 2000s. In September 2003, two powerful state premiers from the big political parties, the conservative Roland Koch from Hesse and the social democrat Per Steinbrueck from North-Rine Westphalia, published a joint proposal to cut subsidies across the board in order to free resources for the government's budget. The commuter tax breaks were herein considered as a subsidy and in December an arbitration commission in parliament

[10] The following historical review for Germany is based on `http://de.wikipedia.org/wiki/Entfernungspauschale#cite_note-21`, `http://www.pendlerrechner.de/geschichte.shtml`, and the database of parliamentary events `http://dipbt.bundestag.de/dip21.web/bt` (last accessed 2013-03-11). For international comparisons see Borck and Wrede (2009) and the references therein.

decided to reduce them to 0.30 euros per kilometer from the first of January 2004 onward. The second, and larger, reduction of the subsidies came during 2006 when the two big parties had formed a formal coalition and the government decided to abolish the tax breaks starting in 2007 for commutes below 20km. However, on 9 December 2008, the federal constitutional court ruled that this new regulation was an unequal and inconsistent treatment of citizens before the law and that it was against the constitution. Hence, the pre-2007 situation was reinstated, but this is beyond the reach of my dataset, which ends in 2007.

In practice, there have thus been two reductions in the value of a given commute during my sample period. For individuals who live very close to their workplace this generally matters less than for individuals who commute long distances, since the tax break that an individual receives is the oneway cardistance between residence and workplace times

- 0.36 euros for the first ten kilometers and 0.40 euros thereafter from 1999 to 2003,
- 0.30 euros per kilometer during 2004 to 2006, and
- 0.30 euros per kilometer from kilometer 20 onwards for 2007.

In order to estimate the yearly tax-deductible amount for each worker, I multiply the resulting amount by 220 workdays per year (Schulze 2009) and divide by 1000 in order to report the tax breaks in thousands of euros.[11]

Figure 3.1 displays the tax breaks enjoyed by the median, the average, the 90, and the 95 percentile commuter in terms of cardistance over the years. We see that the tax breaks follow an upward trend in years where there is no policy change, which reflects the fact that cardistances generally have been increasing as mentioned above. However, in the years of policy changes, the tax breaks drop, and they drop more the longer is the cardistance. The strongest and most equally impacting drop happens in 2007 when tax breaks are only granted above 20km cardistance. For the

[11] There is a complication about the tax breaks for the years from 1999 to 2001 because these were determined as a flat rate only for car commuters while commuters on public transport had to prove the actual incurred cost. For simplicity, I ignore this issue in my calculations. It should not distort my results too much, since, according to Grau (2009), the majority of commuters are still using their cars to travel to the workplace and more than three quarters of commuters of distances above 25km are using their cars—the group that is most affected by the tax break (changes). I also ignore slightly lower commuter tax breaks before 2001, when the rate for commutes beyond 10km was 0.36 instead of 0.40 euros.

longer distance commuters, the overall drop is substantial: the individual at the 95 percentile experiences an overall reduction in tax breaks of almost 1,500 euros per year.

3.2.3 Strength of the Experiment

Before moving on to the specific empirical predictions, it is helpful to develop a better understanding of how important the tax break changes are for different individuals in the sample and thus the strength of the effects that one could reasonably expect. This discussion and the analysis in the following will be done in euros of tax break changes rather than the net income gain that they imply. The reason is that marginal tax rates depend on workers' marital status as well as other income components and deductibles, which are not reported in the data. These problems in determining the marginal tax rate - in addition to the fact that high-earners may react differently to the same (relative) change in the net euro value of a job-residence combination than low-earners - also prevents me from using the difference in the net value of tax break changes for different earnings groups as additional identifying variation in the estimation below.

For the determination of the strength of the experiment, I start with the net effect of a 1000 euros tax break change on an average individual.[12] If this person earns 25,000 euros a year and thus faces a marginal tax rate of 30 percent, the tax break change reduces her net yearly income by 300 euros or by about 2.5 percent. If she has a planning horizon of 15 years and discounts the future by five percent, this amounts to a net present value of the change of about minus 3,300 euros. Thus, the extents of some very common tax break reductions seem already non-negligible.

In further calculations I consider the extent of the tax break changes in relation to estimated overall costs of commuting, including the time use, and compared to the variation in annual fuel costs of commuting. I find that the commuter tax break changes make up about one ninth of overall commuting costs, while the yearly variation in gasoline costs is generally a fraction of the net value of the tax break reduction, in particular for the 2006/07 change.[13]

Another more important consideration is what fraction of my sample actually

[12]Note that 1000 euros is not excessively high, since all individuals who live more than 20km away from their workplace in 2006 face a tax breaks reduction of about 1,400 euros.

[13]For conciseness, I do not report the details of these calculations but they are available upon request.

benefits from the tax breaks. In general, any individual who files a tax return can claim the tax breaks. This should be a very high portion among the individuals covered by the social security system and thus in the data.[14]

However, there may be individuals who benefit from the tax breaks but are systematically not identified in my data as such and vice versa. The former group could be commuters who live in the same (large) municipality as they work but still travel a substantial distance to work. In this case, my estimator of the average effect of a tax break change would be upward biased because I attribute a given overall effect to a too small treated group. The latter group may be individuals who work part time, who have a second home near their workplace from which they commute, who do not earn enough to pay taxes at all, or who do not exceed a general annual allowance of income related expenses of currently 920 euros.[15] The existence of these groups downward-biases my estimate of the average treatment effect of the tax break change.

3.3 Theoretical Hypotheses and Empirical Strategy

In this section, I first derive individual-level predictions and then aggregate implications. The third part presents the empirical strategy.

3.3.1 Micro-Level Predictions

Consider an individual i who currently works at job (location) w and lives at residence (location) r but who is aware of all the other latent jobs and residences that are available to her. Assume her utility to be additively separable in money, more exactly, tax break euros. I can then write

$$u_i(r, w) = v_i(r, w) + TB(\overline{rw}), \tag{3.1}$$

[14] According to the Federal Statistics Office, out of 35.7million employees, 30.3million considered themselves commuters in 2004.

[15] `http://de.wikipedia.org/wiki/Werbungskostenpauschbetrag` (last accessed 2013-03-11). Kloas and Kuhfeld (2003) from the German Institute for Economic Research (DIW) provide a more complete list of such cases and estimate that they constitute about 0.5 percent of total employment. Also, there is a maximum annual claimable amount of commuter tax breaks (4,500 euros in December 2010).

where $v_i(r, w)$ is the non-tax break component of utility including the disutility of commuting and the gross tax break $TB(\overline{wr})$ is an increasing function of the commuting distance and the tax break rate per kilometer. Clearly, when the tax break rate per kilometer falls or when tax breaks are abolished for the first 20 kilometers, latent combinations of work and residence that feature shorter commutes than w and r will become relatively more attractive.

Note that the components of equation (3.1) should be interpreted as flows for a correct utility maximization via choosing the highest $u_i(r, w)$: the tax breaks $TB(\overline{rw})$ occur every year that the person lives and works in the work-residence combination r, w) and accordingly $v_i(r, w)$ is the flow utility per year in this combination.

Figure 3.2 illustrates how the relative attractiveness of combinations changes with the commuting distance for the specific policy changes of 2003/4 and 2006/7. In the case of 2003/4, the relative change in tax breaks is higher the larger is the difference between the two commuting distances. A proportional relationship also holds for the combinations within 20 kilometers for the 2006/7 change, but thereafter the relative attractiveness doesn't change any further. This is because, for example, a commute of 30 kilometers has lost the same money value as a commute of 50 kilometers.

The valuations of the currently chosen and the latent job-residence combinations fluctuate all the time for every individual,[16] but from figure 3.2 I would expect them to fluctuate more when changes in commuting tax breaks occur. The reason is that these changes add to the "normal" variation in relative valuations by making short cardistance combinations relatively more attractive compared to long cardistance combinations. I would thus expect that individuals are more likely to switch jobs and/or residence in years where commuting tax breaks fall. Moreover, they are more likely to switch in a way such that the new cardistance is shorter.

It is also interesting to understand whether individuals are more likely to react to tax break changes by changing job or by changing residence. On the one hand, it provides evidence on the causal effect of pecuniary changes for the current job or residence location on the likeliness to switch job or to move house. This is of interest in relation to the job mobility and migration literature (see, for example, Bartel 1979, Topel and Ward 1992) and location-based policies, which have a rationale if

[16]This is the reason for the turnover in jobs and residences that we observe in the data.

individuals are geographically immobile to a substantial degree (see the discussion in Moretti 2011). On the other hand, it reveals what the average person values more: her residence location, and thus a significant part of her private life, or her job? More precisely, the coefficient on the tax break change in the regressions on the likeliness to switch jobs or move residence will provide me with the fraction of individuals for whom the difference in the value of the current job or residence over the best alternative job or residence is less than a constant ($prob(u(w,r) - u(w',r) < c)$ and $prob(u(w,r) - u(w,r')) < c)$.[17] Note that the differences $u(w,r) - u(w',r)$ and $u(w,r) - u(w,r')$ are actually rents. Thus, I can examine whether the average individual has a higher valuation of her current house or her job over the best alternative, or, equivalently, whether she is more job- or residentially mobile.

I construct outcome variables to analyze the predictions on the individual level: four indicators which assume the value of one if, from the previous to the current year, the individual concerned changes her location of workplace (*Work Switch*),[18] her location of residence (*Residence Switch*), either of those (*Any Switch*), or she switches such that she ends up with a shorter commute (*Closer Switch*). Moreover, I construct the change in cardistance from the previous to the current year (*Cardistance Ch*). If individuals behave like hypothesized in equation (3.1) and figure 3.2, and if moving costs are not prohibitively high compared to the money value of the tax break change, I would expect that the tax break changes make it more likely that the indicator variables assume the value of one. Moreover, the average change in cardistance should turn out more negative.

Table 3.7 provides the means per worker for the indicator outcome variables in the year 2002, while tables 3.4, 3.5, and 3.6 describe the distribution of the change in cardistance in general and conditioning on a change occurring for the years 2002, 2004, and 2007. We see that, per year, almost 14.9% of individuals change their job location and 12.1% change residence location. Overall, they switch 18.4% of times

[17]In fact, as I argue below, it provides me with a lower bound on that fraction.

[18]If I were interested in job mobility without the geographical component, it would be preferable to examine the effect of the tax break changes on whether individuals change the establishments they work instead of looking at whether they change their municipality of workplace. I don't report these alternative regressions here because I am indeed focusing on the locational impacts. Nonetheless, the direction of the effect on establishment switches is the same as the ones for the work (location) switch, though the magnitude is somewhat lower.

while they switch closer only 5.6% of times.[19][20] Tables 3.4, 3.5, and 3.6 support this last point, showing that most of the switches are resulting in longer cardistances.

3.3.2 Macro-Level Predictions

The previous section has shown that, ceteris paribus, the reduction in commuter tax breaks should make every employee weakly prefer a shorter work-residence combination. According to standard theory in urban economics and in economic geography, this shift in aggregate demand for a short commute should lead to increased concentration of population and economic activity.

First, the Alonso-Mills model in the urban economics literature (e.g., see Brueckner 1987, Fujita 1989) focuses on the spatial equilibrium for one city with a central business district in which all firms are exogenously concentrated and around which consumers locate. If commuting costs rise, demand for more central locations of residence increases, which drives up their relative price and the rent gradient (the ratio between rents in the urban center and the periphery) becomes steeper. Further, if housing supply is not completely inelastic, either because of variable lot sizes or the possibility of construction of new houses, the population density close to the center will increase (concentration).

Second, classic economic geography a la Rosen-Roback (see Roback 1982, Moretti 2011) shifts the focus from locational decisions within cities to between cities. Again, there exists a spatial equilibrium in which the marginal consumer-worker is indifferent between locations. Transport costs, which are the equivalent to commuting costs in urban economics, are only introduced explicitly in the new economic geography literature. For example, Helpman (1998) proposes a general equilibrium model in which different cities are characterized by an exogenously given amenity (i.e. housing stock), and firms and individuals optimally locate under agglomeration

[19] These numbers seem quite high and I can only speculate why this is the case. One potential reason may be mis-recording postcodes for work or residence municipalities in some years which is corrected in other years and thus increases the measured job and residence turnover. Whatever the reason, if the upward-bias in measurement of the switches variables is unrelated to the explanatory variables in the regressions below, this should not be a problem.

[20] There also seem to be more changes in job- and work locations in 2005 than in other years. I have searched for explanations for this myself and I enquired about it at the Institute for Employment Research. I found that in 2005 the distinction between East- and West Berlin was abolished and thus municipality assignments have changes. Also, there might have been some updating of employee information from part of employers because there were administrative changes in the pension insurance system. At this stage, there is no reason for me to believe that these changes should be systematically and substantially correlated with the tax break changes of individuals.

economies and transport costs. If transport costs increase, the new equilibrium is characterized by a stronger concentration of individuals and firms in the locations which have more housing stock. Redding and Sturm (2008) generalize this model by endogenizing the housing stock, but the concentration implication remains.

Yet, there exists a convincing alternative hypothesis to the concentration prediction on how the adjustment to a rise in commuting costs may look like. Assume individuals are heterogeneous in terms of their residential preferences and in their productivity in different jobs. It may then be the case that some of them live in municipality A and work in B, while others work in A and live in B. If commuting costs rise, it becomes more attractive to live *and* work in *either* A *or* B and some (pairs of) individuals who were sufficiently close to indifference initially may now find it optimal to re-match in order to reduce their commuting distance. I term this the "re-matching" hypothesis.

A key feature of the re-matching hypothesis is that existing residential units and jobs are occupied in a different way but their relative supply or utilization is unchanged. It also requires sufficient heterogeneity in individuals' locational preferences - maybe because they strongly value the place where they grew up - and in job match quality. This further implies that there exist rents for the current job-residence match and not every worker-consumer is exactly indifferent between locations or the distance that she lives from the urban center as in the homogeneous version of the spatial equilibrium.[21]

I expect the existence and the relative strength of the concentration and the re-matching effect to depend on several key factors. First, if housing supply or occupation rates of the existing housing stock (lot sizes in the terminology of urban economics) are elastic, many individuals will be able to move to urban centers where jobs are disproportionately located.[22] This leads to a strong effect on population concentration and we may also see more concentration of jobs in urban centers.[23] Second, if housing supply or lot sizes are inelastic, the higher demand for more

[21]In his survey paper, Moretti (2011) argues that heterogeneity and rents are an economically important feature of reality and he incorporates them into his generalized spatial equilibrium model.

[22]Occupation rates may also adjust if there are search frictions in the housing market and thus there exists a natural rate of vacancies which is affected by the change in demand.

[23]Employers might however be located in the center already so that no further adjustment on this margin is possible. Moreover, an employer's commuting distance minimization problem is much more complicated than the one of an employee, because the employer has to take into account the commuting distances of all of her employees.

central locations will be absorbed into prices, and property prices in urban areas will rise while concentration will hardly be affected. In terms of the re-matching hypothesis, I expect to see a lot of it if jobs are located in diverse municipalities and if individuals are sufficiently heterogeneous such that there is a lot of scope for re-matching while not being too heterogeneous such that rents are not unsurpassably high. Finally, the concentration hypothesis is an implication from long run general equilibrium models, while the time frame of adjustment that can be considered in this paper is only two years. Re-matching may be much easier during a short period because, by definition, it does not need the housing stock or lot sizes to adjust.

I construct variables that take the value of one if an individual changes job or residence location from a non-urban municipality to one of the 80 cities with more than 100,000 inhabitants, i.e. *Work Switch Urban* and *Residence Switch Urban*, and the value of minus one if they do the opposite switch, i.e. *Work Switch Rural* and *Residence Switch Rural*.[24] If the hypothesized concentration effect is at work in my dataset, I would expect positive coefficients on the tax break change regressor for the *Work Switch Urban* and *Residence Switch Urban* outcome variables and less positive, or negative, coefficients on *Work Switch Rural* and *Residence Switch Rural*. Table 3.7 reports that only a small fraction of overall job switches or house moves are switches from a rural to an urban municipality or vice versa.

3.3.3 Empirical Strategy

In order to examine the causal impact of changes in tax breaks on individuals' locational decisions, and thereby assess the empirical value of the two hypotheses, I run the following general regression:

$$\text{location change} = \beta * \text{tax break change} + \gamma * \text{controls} + \epsilon \tag{3.2}$$

A unit of observation in this regression is an individual in a given year. Location change on the left hand side of the equation refers to the different outcome variables defined above, while tax break change refers to the change in tax breaks for an individual's work-residence combination.

One important feature that I have to control for in the regression is the cardis-

[24] The list of cities with more than 100,000 inhabitants is from Wikipedia `http://en.wikipedia.org/wiki/List_of_cities_in_Germany_with_more_than_100,000_inhabitants` (last accessed 2013-03-11).

tance of last year's work-residence combination. This is because individuals who have a high value of commuting distance are generally more likely to change location and the commuting distance is mechanically related to the change in tax breaks, since the latter is calculated using the former. In my main regressions I prefer to be conservative and to include municipality combination fixed effects to account for this.[25] I also include year fixed effects to account for general differences in the likeliness to change location between different years. My identification thus relies on a systematically different likeliness to change location for individuals in far-distance municipality combinations relative to individuals in short-distance municipality combinations around the tax break change years.

In some of the regressions I even include an interaction of individual fixed effects together with the municipality fixed effects so that the identification relies solely on individuals who do move or switch from the respective municipality. This is because those individuals who don't change location during the sample period are absorbed by the individual-municipality-combination fixed effect. Hence, I can separately examine the tax break changes' effect on the direction and timing of existing moves.

In order for regression (3.2) to identify the average treatment effect of the tax break change on an individidual's location change,[26] the following assumption needs to hold: There is no other factor than the tax break change that affects the relative likeliness of location changes for far-distance municipalities compared to short-distance municipalities over different years. One such factor may be gasoline prices, which I control for explicitly in the regressions. Other control variables that are less central for claiming causality of the regression coefficient, but are nonetheless included in the regression, are dummies for age quartiles, income quartiles, plant size quartiles and the individual's position in the job.

One concern for the causal interpretation of the estimates themselves may be that it is not clear how far in advance individuals expect the tax break changes to happen. The regression results in the next section show that some of them anticipate tax break reductions by moving in the preceding year - even before the contemplated changes were finally decided. However, the optimal response in terms of minimizing

[25]This is only possible because my dataset is truly large so that I have enough realized municipality combinations for statistical inference.

[26]In fact, this is potentially a short run general equilibrium effect because many individuals are hit by the tax break decrease which may systematically affect fuel prices and wages even in the short run. For simplicity and because the overall size of changes are not large enough to impose it, I abstract from general equilibrium effects in this paper.

the negative impact is to a first degree independent of the information release about the policy change: ideally, one would want to move at the 31st of December before the new rules come into effect. Therefore, the timing of information release should not constitute a major problem for the causal interpretation of the estimated effects.

Before getting to the presentation of the empirical results, it is important to explain the construction and the timing of the outcome and the explanatory variables: In terms of the outcome variables, I consider every individual's work-residence combination in the fourth quarter of each year and compare it to the combination in the fourth quarter of last year to calculate the switches and the change in cardistance. In terms of the regressors, unless a switch took place, I do not know which specific work-residence combination an individual considered an attractive alternative to the prevailing work-residence combination last year. Therefore, I cannot calculate the relative money worth change between the two combinations as a regressor. Instead, I rely on calculating the change in the tax break worth of last year's combination that occurred at the last turn of the year ("TB Ch") and that occurs at the next turn of the year ("TB Ch (Next Yr)") just after the fourth quarter that the observation refers to. The coefficient on the "TB Ch" regessor captures the effect of the tax break change last year on location change between last year and this year while the coefficient on "TB Ch (Next Yr)" captures the effect of the tax break change this year, i.e. on an anticipatory move. Figure 3.3 illustrates using a timeline when the respective regressors may assume values different than zero (they are negative then, as tax breaks only decline during the sample period).

3.4 Regression Results

In this section I first present the results from the individual-level regressions and then discuss their implications in light of the concentration and the re-matching hypothesis.

3.4.1 Individual-Level Results

I start by estimating regression equation (3.2) for the likeliness to switch jobs, switch residence, switch either, and to switch closer using a linear probability model. The results of my preferred specification with municipality combination fixed effects are displayed in table 3.8. The change in gasoline cost for the cardistance this year and next year as well as dummies for years, age quartiles, income quartiles, plant size quartiles, and the individual's position in the job are included as control variables throughout. Standard errors are clustered on the municipality combination level.

We see from the table that the tax break change for last year's cardistance between this year and last year has a statistically significant and positive effect on the likeliness to switch or move this year as well as last year. Individuals are also more likely to switch closer. This is all in accordance with the individual-level predictions from section 3.3.1: individuals are supposed to be more likely to move and switch jobs when tax breaks change and they are more likely to do so in order for the resulting cardistance to be shorter. They also engage in anticipatory moves and switches in the sense that some of them switch already in the run up to a tax break change at the turn of the year.

Moreover, the effect on switching closer in column four is stronger than the overall effect on switching in column three. This is as one would expect because increases in commuting costs should encourage switches that lead to a shorter cardistance while at the same time discouraging switches that lead to a longer cardistance. Thus, when tax breaks for commuting fall, the number of shortening switches should increase more than the overall number of switches.[27]

Next, I examine the magnitude of the effect. A tax break change of 1000 euros - the net equivalent of this for an average individual is about 300 euros - increases the likeliness of switching jobs, moving house, switching either, and switching closer by about 0.4, 0.2, 0.5, and 0.5 percentage points, respectively. Given that the averages of these variables are 14.9, 12.1, 18.4, and 5.6 percent, a 1000 euros tax break change increases the likeliness to switch jobs, move house, switch either, and switch closer by 2.7, 1.7, 2.7, and 8.9 percent, respectively. Thus, the effect of the tax break change is significant statistically and in magnitude, and it is strongest for the likeliness to switch closer as we should expect.

Tables 3.9 and 3.10 rerun the main regressions replacing the municipality combination fixed effects by individuals within municipality combination fixed effects and last year's cardistance as the main controls, respectively. The results are quite

[27]It also suggest that the following general equilibrium effect, though plausible, is not of great empirical importance: some long-cardistance combinations become more attractive (possibly because of lower rents or house prices) for people who are not affected very much by the tax break change (possibly because of a low marginal tax rate).

similar but somewhat stronger. As mentioned above, in the case of individual municipality combination fixed effects, the variation exploited only comes from individuals who do move or switch from the respective municipality. Thus, the tax breaks not only induce new switches or prevent distance-increasing moves that otherwise would have happened, but they also influence the direction or the timing of existing switches. We also see that the coefficient on either work or residence switches is now strictly smaller than the coefficient on the closer switches. This implies that the tax break changes prevent or postpone some longer-distance switches on top of encouraging additional shorter-distance switches.

The effect of changes in fuel prices on switching decisions turns out very inconclusive in the tables. This may not be very surprising. First, individuals continuously change their expectations about future fuel prices and price changes affect them immediately, contrary to tax break changes. Thus, when new information is revealed, they may find it optimal to move right away. Such an adjustment process can hardly be fully captured by the yearly average gasoline price changes included in my regressions. Second, individuals have different margins of adjustment to a gasoline price change - such as driving less, changing transport mode, or engaging in car sharing - that are not available in the case of a tax break change and that may be preferred to moving or switching job. In fact, the recent literature in transport economics and in environmental economics finds strong effects of gasoline price changes on driving behavior, new car purchases, and vehicle scrappage decisions (see Li, von Haefen, and Timmins 2008, Bento, Goulder, Jacobsen, and von Haefen 2009, Knittel and Sandler 2010).

From table 3.8 we see that the number of individuals who switch jobs in response to a tax break change is higher than the number who move residence. As mentioned above, the likeliness that the average person's flow valuation of her job over the next best alternative is below 300 euros (tax breaks of 1000 euros) is about 0.4 percent while the corresponding likeliness for her residence is about 0.2 percent.[28] Thus, there are much less individuals who derive a relatively low utility rent from their current house than there are individuals who derive a relatively low utility rent from their current job. I interpret this as evidence that the average person values

[28] These numbers are strictly lower bounds because in fact only the relative attractiveness of combinations with a shorter cardistance increases in response to the tax break reduction. Moreover, the relative attractiveness of combinations that feature a non-zero cardistance does not rise by the full 300 euros because tax breaks for these combinations themselves fall.

her residence (location) more than her job (location).

In principle, one could use this information to try to determine the distribution of- or the average rents from existing work places and residences in the population. However, this would require strong functional form assumptions and strong assumptions about which alternative work-residence combinations individuals are aware of or, equivalently, the combinations' arrival rates in a search model. I thus refrain from such an exercise in this paper.

I conclude from this discussion that the tax break changes for some individuals are substantial enough to change their preferred job-residence combination and to raise the utility differential between the old and the new combination above potential fixed costs of moving house or switching job. Moreover, the direction of the effect turns out as expected. Finally, the average individual seems much more likely to be willing to switch jobs rather than to move house in response to a fixed change in the value of her current combination, which can be interpreted as a higher valuation of her private life than her job.

3.4.2 Concentration Versus Re-matching

In order to understand how much of the overall adjustment to the tax break changes stems from concentration versus re-matching, I need to examine to what extent the effect leads to rural-to-urban and urban-to-rural moves.

Table 3.11 reports the results of the regressions that address this question.[29] We see that individuals are significantly more likely to switch their jobs and their residences from a rural to an urban location. They are equally likely to switch jobs from an urban to a rural location whereas the tax break reduction leads to a much weaker increase in urban-to-rural relocations of residence. Furthermore, the magnitude of each of the effects is a fraction of the overall adjustment reported in table 3.8.

The last observation indicates that the re-matching effect in terms of job as well as residence relocation is much stronger than potential concentration effects. Indeed, the concentration effect for jobs seems to be non-existent because there are as many induced urban-to-rural moves as there are rural-to-urban moves. As the overall distribution of jobs between rural and urban locations remains unchanged,

[29] The results are again similar but somewhat stronger when using individual-municipality combination fixed effects or cardistance as controls instead of municipality combination fixed effects.

these switches should therefore be considered as rematches. In terms of residential relocations, there seems to be a concentration effect, since more individuals are induced to move into cities than are induced to switch in the opposite direction.

What does this imply about the empirical validity of the urban and the economic geography models in the short run? First of all, jobs do not get more concentrated, potentially because they are already to a great degree located in urban centers or because minimizing the overall commuting distance of their employees is simply too complicated and costly for employers.[30] Therefore, the assumption of the vast majority of models that firms are exogenously located in the central business district seems to be harmless.

A stronger concentration effect on residential switches may be restrained either because the demand for relocating toward urban areas is not strongly affected by the tax break changes or because the supply of housing stock and the occupation rates are not very flexible in cities. According to the standard theory, the latter reason would imply a substantial increase in the relative price of existing urban housing stock. I examine this.

Figure 3.4 plots the relative property price index for big and medium-sized cities compared to the overall property price index for old and new flats and houses.[31] There seems to be a general upward trend in the relative price of flats in cities and a u-shape for houses in cities, but it is hard to see any effect on prices around the time of the tax break changes, 2003/04 and 2006/07. A set of formal regression analyses with different specifications in order to account for the time trend also fail to discover a relationship.[32] I abstain from reporting these regressions in the paper for conciseness.[33]

[30] It seems that the latter explanation might be the more relevant one since we do observe individuals switch their jobs from urban to rural locations.

[31] The data were downloaded from the Bank for International Settlements website `http://www.bis.org/statistics/pp.htm` on 17 November 2010. Property prices seem to be a better measure of the effect on the housing market than rental rates because of two reasons. First, rent increases within a short time frame (like one or two years) are restricted by law for privately used properties in order to protect tenants. Second, property prices should factor in the whole net present value of the effect, including the short as well as the long run.

[32] The non-findings of an effect of the tax break changes in property prices are also robust to using property price indices for big cities instead of big and medium-sized cities and for rural areas instead of an overall price index.

[33] Given that I find hardly any effect of the tax break changes on relative property prices, the tiny effect on moving residence from an urban to a rural location should stem from re-matching: either because the employer is located in the rural area that the person is moving to, or the rural area is better connected to the urban area where the employer is located than the previous urban area.

Although there may be other reasons why urban versus rural property prices do not visibly react to the tax break changes,[34] the overall message from the results above seems to be that the relative demand for urban locations is not very strongly affected. This implies that re-matching is much more important as a channel of adjustment to increased commuting costs than concentration. It further implies that the assumptions underlying the re-matching hypothesis seem to be economically meaningful: individuals are substantially heterogeneous in terms of their locational preferences as well as how productive they are in different jobs. Moreover, there exist rents for the currently occupied job-residence combinations.

So far this analysis has not focused on the dynamics of the adjustment because it only considers the overall switches in the years before and after the tax break changes take effect. Yet, it is interesting to understand better how fast people react and in how much their movements precede or follow the changes. Further, the dynamics of the adjustment might inform us on the long run effect of changes in commuting costs or transport costs more generally. Unfortunately, in this data, I cannot analyze longer time periods than two years, because there is a new policy change coming up in 2006/07 for the 2003/04 change and the available data end in 2007. Moreover, even if it were possible to observe a longer time series of the data, it is unclear if this by itself were much more informative. The reason is because the effects of the tax break changes might be contaminated by other substantial long run shifts that affect land use and urban shape.

Therefore, I separate the available time periods into smaller units instead of of looking at longer horizons. Table 3.12 reports the regression results for the effect on the switches for quarters around the tax break changes.[35]

We see that individuals react already in the first quarter of the year of the change by moving house or switching jobs indicated by the coefficient on "TB Ch Q1 (Next Yr)". The effect then grows until the 5th or 6th quarter ("TB Ch Q1 (Last Yr)" and "TB Ch Q2 (Last Yr)") after the tax break change before it drops back to zero.[36]

[34]Foremost, there might be institutional reasons which prevent prices to reflect supply and demand in the German property market in the short run. For example, rental rate adjustment is very constrained due to laws that protect (private) tenants from high raises. If property prices reflect the net present value of rental income, this should dampen the adjustment to a tax break change. In general, it is a widely held point of view that the German property market is not very free.

[35]The results are again similar but somewhat stronger when using individual-municipality combination fixed effects or cardistance as controls instead of municipality combination fixed effects.

[36]Note that the coefficients for the tax break change last year, i.e. quarters five to eight, are based solely on the 2003/04 change and might therefore not reflect the average adjustment to both

This seems to indicate that most of the adjustment happens in the short term already. Alternatively, it may simply take very long for housing supply to change, but the change might be continuous and small in the subsequent periods, so that we have a strong and visible effect of mostly re-matching and some concentration in the short term and a continuous and small per quarter effect on concentration via new housing supply in the medium and the long term. The lack of an effect on property prices casts doubt on this second explanation, however.[37]

3.5 Implications for Public Policy

Up to this point, the discussion has ignored the effect of the tax break reduction on the average distance commuted. The reason for this neglect was that estimation of equation (3.2) with (the change in) cardistance on the left-hand side is in fact biased in panels with a short time dimension and fixed effects (Nickell 1981), and the focus of the analysis was on the causal identification of the treatment effect of the tax break change.[38]

However, in this section I want to focus on the implications that the tax break change has for policy-relevant variables such as fuel consumption, CO2 emmissions, and tax payments and revenues. Therefore, I need its effect on the commuting distance in the first place. Table 3.13 reports this information. The first column reports the preferred regression with municipality fixed effects as controls. On average, a one thousand euro decrease in tax breaks leads to an overall decrease on the commuting distance of about 0.79 kilometers (summing the coefficients on "TB Ch" and "TB Ch (Next Yr)"). This is about a decrease of one thirtieth of the average cardistance in the sample according to tables 4-6.

Compared to column (1) in table 3.13, the second column distinguishes between the tax break changes in 2003/4 and those in 2006/7 and the third and fourth column examine the change in cardistance conditioning on the event that a switch takes place. We see that the 2006/7 tax break changes seem to have had more than double the impact on the commuting distance than the 2003/4 tax break changes.

events.

[37]Yet, Moretti (2011), in his survey paper, quotes studies that find that the adjustment to local demand shocks take around 10 years.

[38]Nickell's result is that in short panels with lagged dependent variables and individual fixed effects the lagged dependent variables are correlated with the component of the observation's error term that is constant over time.

Moreover, conditioning on the event that a switch takes place, the impact on the distance is (unsurprisingly) very high: a 1,000 euro tax break change makes the switch lead to a more than four kilometer lower commuting distance.

I want to consider the effect on the commuting distance and fuel usage in three scenarios of tax break changes: the actual reductions of 2003/4 and 2006/7 as well as a hypothetical complete abolition of tax breaks in 2003/04. Table 3.14 lists the preferred coefficient of 0.79 kilometers lower cardistance for a 1,000 tax break change together with further information that is used and the respective sources. Using the average tax break changes of 0.5 thousand euros in 2003/4 and 0.6 thousand euros in 2006/7 as well as the average overall tax break in 2003 of 2.3 thousand euros, I arrive at an overall effect of the three scenarios of a decline in the average annual commute of 0.40, 0.47, and 1.82 kilometers, respectively. This is displayed in the first row of table 3.15.[39] Using the average number of workdays per year, the fact that the above distance is just oneway, and the total number of employees, the effect of the three scenarios of tax break changes on the overall cardistance commuted becomes 6,714, 8,124, and 30,886 million kilometers, respectively. These are 0.98, 1.18, and 4.53 percent of the 690 billion kilometers traveled in the country overall per annum.

In order to compute the estimated fuel savings for the whole economy in terms of liters and money value, I assume that every commuter goes to work by (gasoline engine) car by herself. The estimates in the following should thus be interpreted as an upper bound, since going by car is known to be the most fuel-intensive and CO2 emitting transport mode.[40] Using the data on the average fuel consumption and the fuel price in the respective years from table 3.14, the amount of fuel saved becomes 537, 626, and 2,471 million liters, and 577, 793, and 2,654 million euros, respectively (see table 3.15). Overall, the country-wide fuel usage in the transport sector is 48 billion liters per year, hence a full abolition of commuter tax breaks in 2003/04 would have reduced fuel usage in the transport sector by up to 5.2%.

These savings in fuel consumption also have an effect on the emission of greenhouse gases, notably CO2. Burning one liter of gasoline generates about 2.32 kilogram of carbon dioxide, hence the tax break changes reduce emissions by an esti-

[39] Note that there is no incentive for an "intensive margin" of adjustment (apart from an income effect), since the tax breaks are independent of transport modes, car sharing, or the actual distance traveled per journey.

[40] Yet, note that this assumption is in fact not very extreme since about two thirds of all commuters use the car (Grau 2009).

mated 1.25, 1.45, and 5.73 tonnes. Using data on European Union emission rights trading between firms, the market would price this at "only" 12.96, 9.36, and 59.62 Mio euros of environmental savings.[41] In terms of overall emissions in passenger traffic, the tax breaks lead to an emission reduction of 0.74, 0.94, and 3.39 percent, respectively.

While the impact of the tax break change on kilometers traveled, fuel burnt, and greenhouse gases emitted seems unambiguously positive, its expansionary effect on the tax base may be good news for the exchequer but not for the taxpayer. Using the respective formulae to calculate the tax breaks before and after the changes in 2003/04 and 2006/07, the first row in table 3.16 provides the average per person reduction for the old cardistance, i.e. without taking into account individuals' reaction to the change. As individuals switch their workplace and residence closer together, the claimable tax breaks decrease even further or, to put it from the exchequer's perspective, the tax base rises even further. Row two of the table displays this effect.[42] Overall, the tax base per year increases by 21, 26, and 98 billion euros, respectively, helped by individuals' behavioral responses of moving residence and job location in order to reduce commuting distances. This is a substantial amount and assuming that the average marginal income tax rate is around 30 percent, it provides the government with additional tax revenues of 6.3, 7.8, and 27 billion euros, respectively.[43]

Conservative commentators and politicians have argued for a long time that the commuter tax breaks serve the purpose of supporting and preserving life on the countryside. Indeed, the results in table 3.11 show that the tax break changes make individuals more likely to move from rural areas to urban areas more than they make them engage in the opposite move. Yet, the concentration effect is only 0.1 percent per 1000 euros tax break change. Thus, even tax breaks' hypothetical full abolition in 2003/04 would have increased the likeliness to move from the countryside to a city of more than 100,000 inhabitants by just 0.23 percentage points temporarily.

[41] There is widespread criticism claiming that the practice of allocating a large number of emission rights to firms for free leads to a too low price for the emission rights. Therefore, the above numbers might severely underestimate the true social benefits from the carbon emissions reduction.

[42] In order not to have to deal with the exact distribution of cardistances for the 2006/07 change, I assume that all the individuals are in fact able to claim positive tax breaks for every kilometer, i.e. I ignore the 20km with zero tax breaks. Therefore, the estimated effect again should be considered an upper bound of the true effect.

[43] In addition to the changing commuting distances, wages might respond to in general equilibrium which would affect tax revenues. The direction and the extent of such an effect is hard to assess without putting a lot of specific structure on the problem, however, and from which I refrain.

This effect can hardly be termed as "landflight".

Despite all the positive effects on the environment and travel expenses that the reduction in tax breaks seemed to have, I cannot make a normative statement whether it was "beneficial". In fact there is a sound economic justification for the commuter tax breaks:

Suppose that one can split up the overall utility $u(r,w)$ from a each job-residence match into all other benefits $b(r,w)$ and commuting costs $c(r,w)$. It is then efficient for every individual to choose

$$u(r^*,w^*) = b(r^*,w^*) - c(r^*,w^*) = max\{u(r_j,w_k)\}$$

for all j,k. If proportional income taxation is used, in order not to distort the choice of efficient job-residence matches, the tax rate should be applied to the commuting costs as well, i.e. $(1-t)[b(r,w)-c(r,w)]$. Individuals should thus be allowed to deduct the exact commuting costs from their gross taxable income. More generally, in order to preserve efficient matches, proportional income taxation should only be applied to the "rent" from these matches.

If it is very costly and subject to fraud to have each individual prove their exact commuting costs to the tax authorities, a tax break that reflects the average costs per kilometer traveled may be a second best solution to this problem, i.e. $TB(\overline{rw}) = avg[c(r,w)]$.[44] This is the rationale for the introduction of the commuting tax breaks in the first place. The full abolition of commuting tax breaks in 2003/04 might therefore seem attractive from an environmental, travel expenses, and tax revenues point of view, but it might have distorted efficient matching in the housing and the labor market.[45][46]

[44] Although Knittel and Sandler (2013)'s results suggest that using such a proxy is fraught with error.

[45] Note that if the individual's costs $c(r,w)$ do not reflect society's cost from commuting - which is likely the case - there is a rationale for bringing $c(r,w)$ to its efficient level trough taxation.

[46] There exists a crude test whether the original tax breaks were higher than the actual commuting costs per kilometer: if $TB(\overline{rw}) > avg[c(r,w)]$, the chosen distances would have been inefficiently long. Hence, a reduction in tax breaks would increase $b(r^*,w^*) - c(r^*,w^*)$. If we think of $b(r^*,w^*)$ as mainly the wage and note that $-c(r^*,w^*)$ always increases because of the decreasing commuting distance, an increasing wage as a response to an increase in tax breaks would constitute evidence that the tax breaks were too high initially. In unreported regressions I find that there is no clear effect on wages. Thus there is no strong evidence for too high commuter tax breaks in the first place.

3.6 Conclusion

This paper has shown that individuals switch job or move residence in order to reduce their commuting distance when the costs of commuting rise. It has also provided strong evidence that higher commuting costs strengthen forces of population concentration, which is a core result from standard theory in urban economics and in economic geography. However, concentration is just a small part of the overall adjustment in terms of individuals' residence location and there is no evidence for job concentration. This is interesting because another margin of adjustment that has not received much attention in the urban and geography literature seems to account for the majority of the reaction - namely that individuals change the occupation of existing jobs and houses to reduce commuting distances. I term this margin of adjustment "re-matching".

The analysis has ignored some potentially important factors that were beyond the scope of the paper. Most importantly, no broadly encompassing general equilibrium notion of the effects was developed apart from a preliminary analysis of the effect on relative housing supply and property prices. For example, one could have argued that the tax break changes and the resulting lower cardistances might also have an effect on fuel prices and even wage rates, which in turn affects individuals' location decisions. Moreover, in the theoretical part I focused entirely on the substitution effect of increases in relative prices whereas the policy change may also have heterogeneous income effects for every individual in the data.[47]

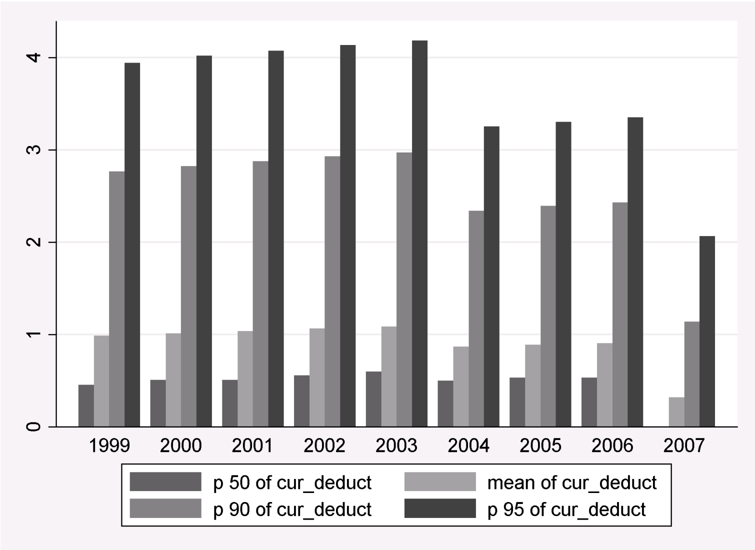
Naturally, the question arises whether results from the specific experiment exploited in this paper can be generalized to other contexts. Germany is a decentralized and densely populated country.[48] The fact that many employees can choose between jobs in different employment centers may favor the re-matching effect over the concentration effect - compared to centralized countries like France or the United Kingdom. It may also favor the re-matching effect compared to a large country like the United States, where switching employment between different urban areas while residing at the same place seems less feasible. For cultural and institutional reasons, Germans are generally less mobile than Americans in the labor market as well as

[47]For example, the Fujita (1989) book assumes positive income effects on commuting distance, i.e. that wealthier households prefer to locate farther away from the urban center.

[48]When comparing it with other developed countries it also seems to have an efficient public and private transport infrastructure but an underdeveloped housing market.

in the housing market. Therefore, and because of the greater distances involved, one might expect the overall adjustment to a given change in commuting costs on distances between home and work to be even larger in the United States than in Germany.

Figure 3.1: The Commuter Tax Break Distribution Over the Years

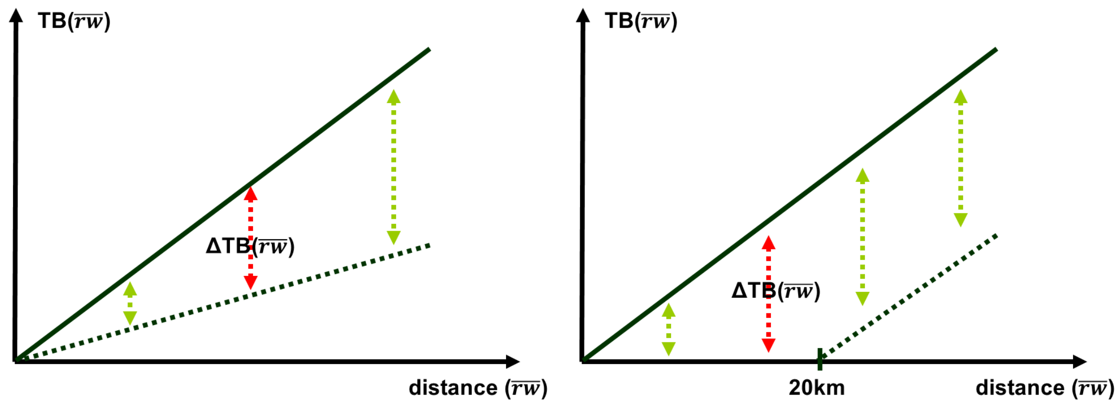


Note.—The figure depicts the median, mean, 90th, and 95th percentile of the tax break distribution in thousand euros over the sample years.

Figure 3.2: Tax Break Reductions and the Resulting Changes in the Relative Attractiveness of Commuting Distances

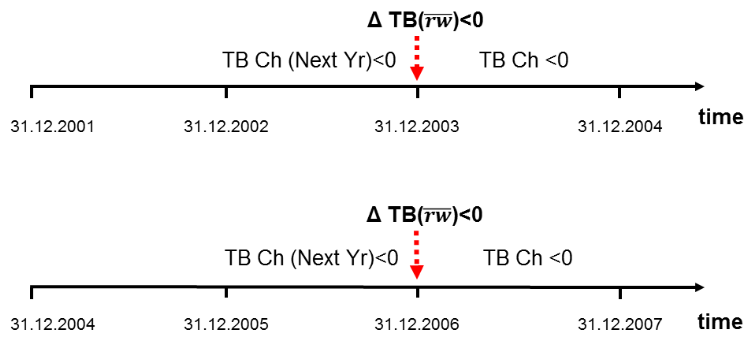
(a) Policy change in 2003/4

(b) Policy change in 2006/7



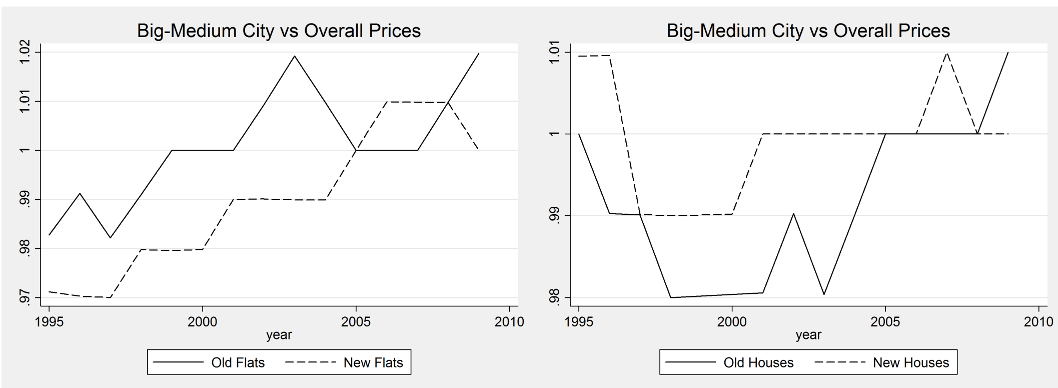
Note.—The figures illustrate the decline in the tax breaks by cardistance $\overline{rw}$ for the 2003/4 (left) and the 2006/7 (right) change.

Figure 3.3: The Timing of the Explanatory Variables



Note.—The figure depicts when the explanatory variables TB Ch and TB Ch (Next Yr) may assume values different from zero. TB Ch is negative for individuals observed in the fourth quarter of 2004 and 2007 who had work-residence combinations in the fourth quarter of 2003 and 2006, respectively, that featured a non-zero commuting distance. TB Ch (Next Yr) is negative for individuals observed in the fourth quarter of 2003 and 2006 who had work-residence combinations in the fourth quarter of 2002 and 2005, respectively, that featured a non-zero commuting distance.

Figure 3.4: Relative Property Prices over Time



Note.—The figures depicts the time series of relative price indices for new and old flats (left) and houses (right) in big and medium cities versus the overall index.

Table 3.1: Summary Statistics per Individual

	count	mean	p10	p25	p50	p75	p90
Years in Sample	942746	7.0	2.0	4.0	8.0	10.0	10.0
Work Switch	942746	0.9	0.0	0.0	0.0	1.0	3.0
Residence Switch	942746	0.7	0.0	0.0	0.0	1.0	2.0
Age (Years)	930122	38.5	20.0	27.0	38.0	48.9	57.5
Female	942746	0.4	0.0	0.0	0.0	1.0	1.0
Monthly Wage (euro)	914878	1703.8	297.0	660.5	1488.4	2475.1	3505.7
Cardistance (km)	911373	28.2	0.0	0.0	9.0	23.7	64.0
Urban Workplace	930122	0.5	0.0	0.0	0.3	1.0	1.0
Urban Residence	925221	0.4	0.0	0.0	0.0	1.0	1.0
Observations	942746						

Note.—The table reports means and quantiles for the number of years individuals are in the sample and the number of work- and residence switches they made during that time (the first three variables). For the remainder of the variables it reports means and quantiles in the fourth quarter of each person-year in the sample (1999-2007).

Table 3.2: Summary of Municipalities by Residence Numbers

	count	mean	p10	p25	p50	p75	p90
Number of Residents	12585	53.6	2.0	4.0	11.0	33.0	92.0
Number of Connections	12585	48.1	2.0	4.0	11.0	33.0	92.0
Observations	12585						

Note.—The table summarizes the distribution of the number of residents in the dataset (i.e. about two percent of the actual number of residents) for municipalities that report at least one resident in the year 2002. It also provides the distribution of the number of different employment municipalities in the data that these persons commute to.

Table 3.3: Summary of Municipalities by Employment Numbers

	count	mean	p10	p25	p50	p75	p90
Number of Employees	10451	64.5	1.0	2.0	8.0	29.0	92.0
Number of Connections	10451	58.0	1.0	2.0	8.0	28.0	92.0
Observations	10451						

Note.—The table summarizes the distribution of the number of employees in the dataset (i.e. about two percent of the actual number of employees) for municipalities that report at least one employee in the year 2002. It also provides the distribution of the number of different residence municipalities in the data that these persons commute from.

Table 3.4: Summary of Cardistance and Tax Breaks in 2002

	count	mean	p25	p50	p75	p90	p95
Cardistance (Last Year)	545486	25.8	0.0	6.4	19.6	45.4	105.0
Tax Break (Last Year)	545486	2.2	0.0	0.5	1.6	3.9	9.1
Tax Break (Change)	545486	0.0	0.0	0.0	0.0	0.0	0.0
Cardistance (Change)	516699	-0.2	0.0	0.0	0.0	0.0	7.0
Cardistance (Change) if Switch	61698	-1.5	-16.8	1.5	18.6	73.1	197.2
Observations	545486						

Note.—(Last Year) refers to the fourth quarter of the previous year, i.e. approximately the beginning of the year considered. (Change) refers to the change during the considered year, i.e. between the fourth quarter of the previous year and the fourth quarter in the current year. Cardistances are in kilometers and tax breaks in thousand euros.

Table 3.5: Summary of Cardistance and Tax Breaks in 2004

	count	mean	p25	p50	p75	p90	p95
Cardistance (Last Year)	544465	26.2	0.0	7.0	20.6	46.5	105.0
Tax Break (Last Year)	544465	2.3	0.0	0.6	1.7	4.0	9.1
Tax Break (Change)	544465	0.5	0.0	0.1	0.4	0.9	2.2
Cardistance (Change)	516456	0.3	0.0	0.0	0.0	0.0	5.7
Cardistance (Change) if Switch	56299	2.9	-15.9	2.8	20.3	80.2	217.5
Observations	544465						

Note.—(Last Year) refers to the fourth quarter of the previous year, i.e. approximately the beginning of the year considered. (Change) refers to the change during the considered year, i.e. between the fourth quarter of the previous year and the fourth quarter in the current year. Cardistances are in kilometers and tax breaks in thousand euros.

Table 3.6: Summary of Cardistance and Tax Breaks in 2007

	count	mean	p25	p50	p75	p90	p95
Cardistance (Last Year)	552369	27.9	0.0	7.6	21.8	49.9	116.6
Tax Break (Last Year)	552369	1.8	0.0	0.5	1.4	3.3	7.7
Tax Break (Change)	552369	0.6	0.0	0.5	1.3	1.3	1.3
Cardistance (Change)	533305	0.4	0.0	0.0	0.0	0.0	7.4
Cardistance (Change) if Switch	63302	3.0	-18.1	1.6	20.8	86.0	235.3
Observations	552369						

Note.—(Last Year) refers to the fourth quarter of the previous year, i.e. approximately the beginning of the year considered. (Change) refers to the change during the considered year, i.e. between the fourth quarter of the previous year and the fourth quarter in the current year. Cardistances are in kilometers and tax breaks in thousand euros.

Table 3.7: Summary of Switches in 2002.

	count	mean
Work Switch	622461	0.149
Residence Switch	622461	0.121
Any Switch	622461	0.184
Closer Switch	516699	0.056
Work Switch Urban	544112	0.014
Work Switch Rural	544112	0.014
Residence Switch Urban	540988	0.008
Residence Switch Rural	540988	0.008
Observations	622461	

Table 3.8: Main Regressions using Municipality Combination Fixed Effects as Controls

	(1) Work	(2) Res	(3) Work or Res	(4) Closer
TB Ch (Next Yr)	0.004***	0.002**	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.002)
TB Ch	0.003***	0.002**	0.004***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Petrol Cost Ch (Next Yr)	0.019*	-0.028***	0.000	0.002
	(0.011)	(0.005)	(0.013)	(0.015)
Petrol Cost Ch	0.027**	-0.001	0.032**	0.033**
	(0.011)	(0.006)	(0.013)	(0.014)
Main Control	Munic FE	Munic FE	Munic FE	Munic FE
Year Dummies	Yes	Yes	Yes	Yes
Observations	4504185	4504185	4504185	4498019

Note.—The table reports regression results of the *Work Switch*, the *Residence Switch*, the *Any Switch*, and the *Closer Switch* indicators in columns 1-4, respectively, on tax break changes that apply to last year's work-residence combination at the last turn of the year ("TB Ch") and the coming turn of the year ("TB Ch (Next Yr)"). Controls are average petrol cost changes between these years and not reported dummies for age quartiles, income quartiles, plant size quartiles, and the individual's position in the job. Moreover, year fixed effects and municipality combination fixed effects are included. Standard errors in parentheses: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 3.9: Main Regr. using Individual-Municipality Combination FE as Controls

	(1) Work	(2) Res	(3) Work or Res	(4) Closer
TB Ch (Next Yr)	0.003***	0.002**	0.004***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)
TB Ch	0.007***	0.003**	0.009***	0.013***
	(0.001)	(0.001)	(0.001)	(0.001)
Petrol Cost Ch (Next Yr)	-0.047***	-0.028***	-0.062***	-0.046***
	(0.006)	(0.005)	(0.006)	(0.014)
Petrol Cost Ch	-0.013*	-0.009	-0.011	0.004
	(0.007)	(0.007)	(0.007)	(0.013)
Main Control	Mu*Ind FE	Mu*Ind FE	Mu*Ind FE	Mu*Ind FE
Year Dummies	Yes	Yes	Yes	Yes
Observations	4504185	4504185	4504185	4498019

Note.—The table reports regression results of the *Work Switch*, the *Residence Switch*, the *Any Switch*, and the *Closer Switch* indicators in columns 1-4, respectively, on tax break changes that apply to last year's work-residence combination at the last turn of the year ("TB Ch") and the coming turn of the year ("TB Ch (Next Yr)"). Controls are average petrol cost changes between these years and not reported dummies for age quartiles, income quartiles, plant size quartiles, and the individual's position in the job. Moreover, year fixed effects and individual-municipality combination fixed effects are included. Standard errors in parentheses: * p<0.1, ** p<0.05, *** p<0.01.

Table 3.10: Main Regressions using the Cardistance as Control

	(1) Work	(2) Res	(3) Work or Res	(4) Closer
TB Ch (Next Yr)	0.005***	0.001**	0.006***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
TB Ch	0.003***	0.002***	0.005***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Petrol Cost Ch (Next Yr)	0.016***	-0.051***	-0.021***	-0.031***
	(0.004)	(0.003)	(0.004)	(0.003)
Petrol Cost Ch	0.014***	-0.012***	0.010**	0.006**
	(0.004)	(0.004)	(0.004)	(0.002)
Main Control	Cardist	Cardist	Cardist	Cardist
Year Dummies	Yes	Yes	Yes	Yes
Observations	4504185	4504185	4504185	4498019

Note.—The table reports regression results of the *Work Switch*, the *Residence Switch*, the *Any Switch*, and the *Closer Switch* indicators in columns 1-4, respectively, on tax break changes that apply to last year's work-residence combination at the last turn of the year ("TB Ch") and the coming turn of the year ("TB Ch (Next Yr)"). Controls are average petrol cost changes between these years and not reported dummies for age quartiles, income quartiles, plant size quartiles, and the individual's position in the job. Moreover, year fixed effects and the cardistance are included. Standard errors in parentheses: * p<0.1, ** p<0.05, *** p<0.01.

Table 3.11: Urban-Rural Switch Regressions using Municipality-Combination Fixed Effects as Controls

	(1) Work Urb	(2) Work Rur	(3) Res Urb	(4) Res Rur
TB Ch (Next Yr)	0.001***	0.001***	0.001***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
TB Ch	0.001***	0.001**	0.001***	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Petrol Cost Ch (Next Yr)	0.002	0.007**	-0.005***	-0.005***
	(0.002)	(0.003)	(0.001)	(0.001)
Petrol Cost Ch	0.005***	0.009***	0.001	-0.001
	(0.002)	(0.003)	(0.001)	(0.001)
Main Control	Munic FE	Munic FE	Munic FE	Munic FE
Year Dummies	Yes	Yes	Yes	Yes
Observations	4504175	4504175	4502743	4502743

Note.—The table reports regression results of the *Work Switch Urban*, the *Work Switch Rural*, the *Residence Switch Urban*, and the *Residence Switch Rural* indicators in columns 1-4, respectively, on tax break changes that apply to last year's work-residence combination at the last turn of the year ("TB Ch") and the coming turn of the year ("TB Ch (Next Yr)"). Controls are average petrol cost changes between these years and not reported dummies for age quartiles, income quartiles, plant size quartiles, and the individual's position in the job. Moreover, year fixed effects and municipality combination fixed effects are included. Standard errors in parentheses: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 3.12: Quarterly Regressions using Municipality-Combination Fixed Effects as Controls

	(1) Work	(2) Res	(3) Work or Res	(4) Closer
TB Ch Q1 (Next Yr)	0.001***	0.001***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
TB Ch Q2 (Next Yr)	-0.001***	0.001***	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
TB Ch Q3 (Next Yr)	0.002***	0.000*	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
TB Ch Q4 (Next Yr)	0.001***	0.001***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
TB Ch Q1	0.001***	0.001***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
TB Ch Q2	-0.001***	0.001***	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
TB Ch Q3	0.003***	0.001***	0.004***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
TB Ch Q4	0.000**	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
TB Ch Q1 (Last Yr)	0.000	0.002*	0.003***	0.004***
	(0.001)	(0.001)	(0.001)	(0.000)
TB Ch Q2 (Last Yr)	0.003**	0.007***	0.010***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
TB Ch Q3 (Last Yr)	-0.001	-0.001	-0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.000)
TB Ch Q4 (Last Yr)	-0.002**	-0.001	-0.002	-0.001*
	(0.001)	(0.001)	(0.001)	(0.000)
Petrol Cost Ch (Next Yr)	0.002	-0.009***	-0.006	-0.003
	(0.003)	(0.002)	(0.005)	(0.005)
Petrol Cost Ch	0.005*	-0.004***	0.001	0.003
	(0.003)	(0.001)	(0.004)	(0.004)
Main Control	Munic FE	Munic FE	Munic FE	Munic FE
Year Dummies	Yes	Yes	Yes	Yes
Observations	18327603	18327603	18327603	18319705

Note.—The table reports regression results of the *Work Switch*, the *Residence Switch*, the *Any Switch*, and the *Closer Switch* indicators at a quarterly frequency in columns 1-4, respectively, on tax break changes that apply to last quarter's work-residence combination at the last turn of the year ("TB Ch"), the previous to last turn of the year, and the coming turn of the year ("TB Ch (Next Yr)"). Controls are average petrol cost changes between the years and not reported dummies for age quartiles, income quartiles, plant size quartiles, and the individual's position in the job. Moreover, year fixed effects and municipality combination fixed effects are included. Standard errors in parentheses: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 3.13: Regr. on Cardistance using Municipality-Combination FE as Controls

	(1)	(2)	(3)	(4)
TB Ch (Next Yr)	-0.294	-0.236	-4.647***	-4.159***
	(0.206)	(0.224)	(1.052)	(0.970)
TB Ch	-0.494**	-0.373*	-4.472***	-4.323***
	(0.203)	(0.222)	(0.897)	(0.884)
Petrol Cost Ch (Next Yr)	8.861***	8.905***	-20.340**	-19.512**
	(1.560)	(1.554)	(9.304)	(9.055)
Petrol Cost Ch	-0.287	-0.111	-18.950**	-18.088**
	(1.431)	(1.417)	(8.307)	(7.944)
TB Ch 0607 (Next Yr)		-0.558**		-7.140***
		(0.242)		(2.590)
TB Ch 0607		-1.074***		-3.228
		(0.245)		(2.974)
Main Control	Munic FE	Munic FE	Munic FE	Munic FE
Sample	Full	Full	Only Switches	Only Switches
Year Dummies	Yes	Yes	Yes	Yes
Observations	4498019	4498019	535783	535783

Note.—The table reports regression results of the cardistance on tax break changes that apply to last year's work-residence combination at the last turn of the year ("TB Ch") and the coming turn of the year ("TB Ch (Next Yr)"). Compared to the first column, the second column distinguishes between the tax break changes in 2003/4 and those in 2006/7 while the third and fourth column examine the change in cardistance conditioning on the event that a switch takes place. Controls are as in the previous tables and standard errors in parentheses: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 3.14: Information and Sources for the Policy Effect Calculations

Variable	2003/04	2006/07	Source
Employees (in mio)	38.63	39.00	destatis.de
Workdays per year	220	220	Schulze (2009)
Avg fuel usage (liters per km)	0.080	0.077	autopresse.de
Fuel price (euro per liter, yearly average)	1.07	1.27	mwv.de
CO2 emissions (in kg per liter of petrol)	2.32	2.32	de.wikipedia.org
CO2 price (in euro per tonne)	10.40	6.45	eex.com
Overall CO2 emissions (in mio tonnes p.a.)	889	867	umweltbundesamt.de
Fraction of CO2 emission in traffic	0.19	0.18	umweltbundesamt.de
Overall fuel usage in traffic (bio liters p.a.)	48	47	umweltbundesamt.de
Overall person road travel (bio km p.a.)	682	687	umweltbundesamt.de
Avg cardist	26.20	27.90	iab data
Tax break rate before	0.40	0.30	iab data
Tax break rate after (<20km)	0.30	0.00	iab data
Tax break rate after (>20km)	0.30	0.30	iab data
Avg tax break (tsd euro p.a.)	2.30	1.80	iab data
Avg tax break change (tsd euro p.a.)	-0.50	-0.60	iab data
Estimated effect of tax break change on cardist (in km)	0.79	0.79	iab data

Table 3.15: Estimated Effect on Cardistance, Fuel Usage, and CO_2 Emissions

	2003/04	2006/07	full abolition
p.p. cardistance reduction (in km)	0.40	0.47	1.82
Reduction overall distance (mio km)	6,714	8,124	30,886
Reduction overall distance (in % of person road travel)	0.98	1.18	4.53
Fuel savings (in mio liters)	537	626	2,471
Fuel cost savings (in mio euro)	577	793	2,654
Fuel savings (in % of fuel usage)	1.12	1.33	5.15
CO2 emissions reduction (in mio tonnes)	1.25	1.45	5.73
CO2 emissions reduction (in mio euro)	12.96	9.36	59.62
CO2 emissions reduction (in % of traffic emissions)	0.74	0.94	3.39

Table 3.16: Estimated Effect on the Tax Base

	2003/04	2006/07	full abolition
Avg tax break reduction for original cardist (tsd euro p.a.)	0.50	0.60	2.30
Tax break reduction due to cardist change (tsd euro p.a.)	0.052	0.062	0.240
Overall increase in tax base (bio euro p.a.)	21	26	98

Chapter 4

Bibliography

Abowd, J. M., and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57(2), 411–45.

Acemoglu, D., and D. Autor (2010): "Skills, Tasks and Technologies: Implications for Employment and Earnings," Working Paper 16082, National Bureau of Economic Research.

Ahlfeldt, G. M., and A. Feddersen (2010): "From periphery to core: economic adjustments to high speed rail," Working Papers 2010/38, Institut d'Economia de Barcelona (IEB).

Altonji, J. G., P. Bharadwaj, and F. Lange (2008): "Changes in the Characteristics of American Youth: Implications for Adult Outcomes," Working Paper 13883, National Bureau of Economic Research.

Altonji, J. G., and L. M. Segal (1996): "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14(3), 353–66.

Aughinbaugh, A., and R. M. Gardecki (2007): "Attrition in the National Longitudinal Survey of Youth 1997," Mimeo.

Autor, D. H., and D. Dorn (2012): "The Growth of Low Skill Service Jobs and the Polarization of the U.S. Labor Market," MIT Working Papers 15150, MIT.

Autor, D. H., D. Dorn, and G. H. Hanson (2012): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," Working Paper 18054, National Bureau of Economic Research.

Autor, D. H., and M. J. Handel (2012): "Putting Tasks to the Test: Human Capital, Job Tasks and Wages," Journal of Labor Economics, forthcoming 15116.

Autor, D. H., L. F. Katz, and M. S. Kearney (2006): "The Polarization of the U.S. Labor Market," *American Economic Review*, 96(2), 189–194.

Autor, D. H., L. F. Katz, and A. B. Krueger (1998): "Computing Inequality: Have Computers Changed The Labor Market?," *The Quarterly Journal of Economics*, 113(4), 1169–1213.

Autor, D. H., F. Levy, and R. J. Murnane (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 118(4), 1279–1333.

Autor, D. H., A. Manning, and C. L. Smith (2010): "The Contribution of the Minimum Wage to U.S. Wage Inequality over Three Decades: A Reassessment," Working Paper 16533, National Bureau of Economic Research.

Bartel, A. P. (1979): "The Migration Decision: What Role Does Job Mobility Play?," *American Economic Review*, 69(5), 775–86.

Becker, S. O., K. Ekholm, and M.-A. Muendler (2009): "Offshoring and the Onshore Composition of Tasks and Skills," CEPR Discussion Papers 7391, C.E.P.R. Discussion Papers.

Bedard, K., and D. A. Herman (2008): "Who goes to graduate/professional school? The importance of economic fluctuations, undergraduate field, and ability," *Economics of Education Review*, 27(2), 197–210.

Bento, A. M., M. L. Cropper, A. M. Mobarak, and K. Vinha (2005): "The Effects of Urban Spatial Structure on Travel Demand in the United States," *Review of Economics and Statistics*, 87(3), 466–478.

Bento, A. M., L. H. Goulder, M. R. Jacobsen, and R. H. von Haefen (2009): "Distributional and Efficiency Impacts of Increased US Gasoline Taxes," *American Economic Review*, 99(3), 667–99.

Black, S. E., and A. Sufi (2002): "Who Goes to College? Differential Enrollment by Race and Family Background," Working Paper 9310, National Bureau of Economic Research.

Blau, F. D., and L. M. Kahn (2005): "Do Cognitive Test Scores Explain Higher U.S Wage Inequality?," *The Review of Economics and Statistics*, 87(1), pp. 184–193.

Blinder, A. S. (2009): "How Many US Jobs Might be Offshorable?," *World Economics*, 10(2), 41–78.

Borck, R., and M. Wrede (2009): "Subsidies for intracity and intercity commuting," *Journal of Urban Economics*, 66(1), 25 – 32.

Borjas, G. J. (2006): "Immigration in High-Skill Labor Markets: The Impact of Foreign Students on the Earnings of Doctorates," Working Paper 12085, National Bureau of Economic Research.

Borjas, G. J., and K. B. Doran (2012): "The Collapse of the Soviet Union and the Productivity of American Mathematicians," Working Paper 17800, National Bureau of Economic Research.

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

Brueckner, J. K. (1987): "The structure of urban equilibria: A unified treatment of the muth-mills model," *Handbook of Regional and Urban Economics*, 2, 821–845.

Busse, M. R., C. R. Knittel, and F. Zettelmeyer (2009): "Pain at the Pump: The Differential Effect of Gasoline Prices on New and Used Automobile Markets," Working Paper 15590, National Bureau of Economic Research.

Cameron, A., and P. Trivedi (2005): *Microeconometrics: methods and applications*. Cambridge University Press.

Card, D., and T. Lemieux (2001): "Can Falling Supply Explain The Rising Return To College For Younger Men? A Cohort-Based Analysis," *The Quarterly Journal of Economics*, 116(2), 705–746.

Cortes, G. M. (2012): "Where Have the Middle-Wage Workers Gone? A Study of Polarization Using Panel Data," *Working Paper*.

Crinò, R. (2010): "Service Offshoring and White-Collar Employment," *The Review of Economic Studies*, 77(2), 595–632.

Devereux, P. (2002a): "The Importance of Obtaining a High-Paying Job," Mimeo.

——— (2002b): "Occupational Upgrading and the Business Cycle," *Labour*, 16(3), 423–452.

Distelkamp, M., C. Lutz, U. Petschow, and T. Zimmermann (2008): "Wirkungen fiskalischer Steuerungsinstrumente auf Siedlungsstrukturen und Personenverkehr vor dem Hintergrund der Nachhaltigkeitsziele der Bundesregierung," *Final Report for the Federal Institute for Research on Building, Urban Affairs and Spatial Development.*

Ebenstein, A., A. Harrison, M. McMillan, and S. Phillips (2011): "Estimating the impact of trade and offshoring on American workers using the current population surveys," Policy Research Working Paper Series 5750, The World Bank.

Einig, K., and T. Puetz (2007): "Regionale Dynamik der Pendlergesellschaft: Entwicklung von Verflechtungsmustern und Pendeldistanzen," *Informationen zur Raumentwicklung*, 2/3.

Ellison, G. (2002a): "Evolving Standards for Academic Publishing: A q-r Theory," *Journal of Political Economy*, 110(5), 994–1034.

——— (2002b): "The Slowdown of the Economics Publishing Process," *Journal of Political Economy*, 110(5), 947–993.

Firpo, S., N. M. Fortin, and T. Lemieux (2011): "Occupational Tasks and Changes in the Wage Structure," IZA Discussion Papers 5542, Institute for the Study of Labor (IZA).

Fitzgerald, J., P. Gottschalk, and R. Moffitt (1998): "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics," *Journal of Human Resources*, 33(2), 251–299.

Freeland, C. (2012): "A U.S. Campaign Fixated on Another Era," Editorial from August 30, International Herald Tribune, Global Edition of the New York Times.

Freeman, R. B., and J. van Reenen (2009): "What if Congress Doubled R&D Spending on the Physical Sciences?," Mimeo.

Fujita, M. (1989): *Urban Economic Theory: Land Use and City Size*. Cambridge University Press.

Gabaix, X., and A. Landier (2008): "Why has CEO Pay Increased So Much?," *The Quarterly Journal of Economics*, 123(1), 49–100.

Garicano, L., and E. Rossi-Hansberg (2004): "Inequality and the Organization of Knowledge," *The American Economic Review*, 94(2), pp. 197–202.

Genda, Y., A. Kondo, and S. Ohta (2010): "Long-Term Effects of a Recession at Labor Market Entry in Japan and the United States," *Journal of Human Resources*, 45(1).

Goldin, C. D., and L. F. Katz (2008): *The race between education and technology / Claudia Goldin, Lawrence F. Katz*. Belknap Press of Harvard University Press, Cambridge, Mass. :.

Goolsbee, A. (1998): "Does Government R&D Policy Mainly Benefit Scientists and Engineers?," *The American Economic Review*, 88(2), 298–302.

Goos, M., and A. Manning (2007): "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *The Review of Economics and Statistics*, 89(1), 118–133.

Gould, E. D. (2002): "Rising Wage Inequality, Comparative Advantage, and the Growing Importance of General Skills in the United States," *Journal of Labor Economics*, 20(1), 105–147.

Graeb, C., and D. Vorgrimler (2005): "Zur steuerlichen Bedeutung der Entfernungspauschale," *Wirtschaftsdienst*, 85.

Grau, A. (2009): "Commuters: Majority Still Going by Car," *STATmagazin, Federal Statistics Office Germany*, 30 October 2009.

Groes, F., P. Kircher, and I. Manovskii (2010): "The U-Shapes of Occupational Mobility," Working Paper.

GRONAU, R. (1974): "Wage Comparisons-A Selectivity Bias," *Journal of Political Economy*, 82(6), 1119–43.

GUSTMAN, A. L., AND T. L. STEINMEIER (1981): "The Impact of Wages and Unemployment on Youth Enrollment and Labor Supply," *The Review of Economics and Statistics*, 63(4), pp. 553–560.

HANSEN, W. L. (1991): "The Education and Training of Economics Doctorates: Major Findings of the Executive Secretary of the American Economic Association's Commission on Graduate Education in Economics," *Journal of Economic Literature*, 29(3), 1054–1087.

HASSELBACK, J. (2001): "US Faculty Directories for Economics, Finance and Management," last accessed 2010-10-25, CET 3:58 pm.

HECKMAN, J. (1974): "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42(4), pp. 679–694.

HECKMAN, J. J., AND B. E. HONORE (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58(5), pp. 1121–1149.

HECKMAN, J. J., AND G. SEDLACEK (1985): "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy*, 93(6), pp. 1077–1125.

HELPMAN, E. (1998): "The Size of Regions," *in Topics in Public Economics (New York: Cambridge University Press)*, pp. 33–54.

HSIEH, C.-T., E. HURST, C. I. JONES, AND P. J. KLENOW (2012): "The Allocation of Talent and U.S. Economic Growth," Working paper, Working Paper.

JAIMOVICH, N., AND H. E. SIU (2012): "The Trend is the Cycle: Job Polarization and Jobless Recoveries," Working Paper 18334, National Bureau of Economic Research.

KAHN, L. B. (2010): "The long-term labor market consequences of graduating from college in a bad economy," *Labour Economics*, 17(2), 303–316.

KALAITZIDAKIS, P., T. P. MAMUNEAS, AND T. STENGOS (2003): "Rankings of Academic Journals and Institutions in Economics," *Journal of the European Economic Association*, 1(6), 1346–1366.

Kim, E. H., A. Morse, and L. Zingales (2006): "What Has Mattered to Economics since 1970," *The Journal of Economic Perspectives*, 20(4), 189–202.

Kloas, J., and H. Kuhfeld (2003): "Entfernungspauschale: Bezieher hoher Einkommen begünstigt (Aktuelle Ergebnisse zum Verkehrsverhalten privater Haushalte)," *Wochenbericht des DIW Berlin*, 42/03.

Knittel, C. R., and R. Sandler (2010): "Carbon Prices and Automobile Greenhouse Gas Emissions: The Extensive and Intensive Margins," Working Paper 16482, National Bureau of Economic Research.

——— (2013): "The Welfare Impact of Indirect Pigouvian Taxation: Evidence from Transportation," Working Paper 18849, National Bureau of Economic Research.

Krugman, P. (2008): "Stranded in Suburbia," `http://www.nytimes.com/2008/05/19/opinion/19krugman.html`.

Laband, D. N., and M. J. Piette (1994): "The Relative Impacts of Economics Journals: 1970-1990," *Journal of Economic Literature*, 32(2), 640–666.

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

Li, S., R. von Haefen, and C. Timmins (2008): "How Do Gasoline Prices Affect Fleet Fuel Economy?," Working Paper 14450, National Bureau of Economic Research.

Liu, R., and D. Trefler (2011): "A Sorted Tale of Globalization: White Collar Jobs and the Rise of Service Offshoring," Working Paper 17559, National Bureau of Economic Research.

Machin, S., and J. Van Reenen (2008): "wage inequality, changes in," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan, Basingstoke.

MaCurdy, T., T. Mroz, and R. M. Gritz (1998): "An Evaluation of the National Longitudinal Survey on Youth," *Journal of Human Resources*, 33(2), 345–436.

McLaughlin, K. J., and M. Bils (2001): "Interindustry Mobility and the Cyclical Upgrading of Labor," *Journal of Labor Economics*, 19(1), pp. 94–135.

Michaels, G., A. Natraj, and J. Van Reenen (2010): "Has ICT Polarized Skill Demand? Evidence from Eleven Countries over 25 years," Working Paper 16138, National Bureau of Economic Research.

Michel, P., A. Perrot, and J.-F. Thisse (1996): "Interregional equilibrium with heterogeneous labor," *Journal of Population Economics*, 9, 95–113.

Molloy, R., and H. Shan (2010): "The effect of gasoline prices on household location," Working Papers 2010/28, Institut d'Economia de Barcelona (IEB).

Moretti, E. (2011): "Local labor markets," *Handbook of Labor Economics*, 4, 1237–1313.

Mulligan, C. B., and Y. Rubinstein (2008): "Selection, Investment, and Women's Relative Wages Over Time," *The Quarterly Journal of Economics*, 123(3), 1061–1110.

Murnane, R. J., J. B. Willett, and F. Levy (1995): "The Growing Importance of Cognitive Skills in Wage Determination," *The Review of Economics and Statistics*, 77(2), pp. 251–266.

Murphy, K. M., and R. H. Topel (1985): "Estimation and Inference in Two-Step Econometric Models," *Journal of Business & Economic Statistics*, 3(4), pp. 370–379.

Nickell, S. J. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49(6), 1417–26.

Okun, A. M. (1973): "Upward Mobility in a High-Pressure Economy," *Brookings Papers on Economic Activity*, 1973(1), pp. 207–261.

Oreopoulos, P., T. von Wachter, and A. Heisz (2012): "The Short- and Long-Term Career Effects of Graduating in a Recession," *American Economic Journal: Applied Economics*, 4(1), 1–29.

Ottaviano, G. I., G. Peri, and G. C. Wright (2010): "Immigration, Offshoring and American Jobs," Working Paper 16439, National Bureau of Economic Research.

Oyer, P. (2006): "Initial Labor Market Conditions and Long-Term Outcomes for Economists," *Journal of Economic Perspectives*, 20(3), 143-160.

——— (2008): "The Making of an Investment Banker: Stock Market Shocks, Career Choice, and Lifetime Income," *The Journal of Finance*, 63(6), 2601–2628.

Philippon, T., and A. Reshef (2009): "Wages and Human Capital in the U.S. Financial Industry: 1909-2006," Working Paper 14644, National Bureau of Economic Research.

Pischke, J.-S. (1995): "Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study," *Journal of Business & Economic Statistics*, 13(3), 305–14.

Raaum, O., and K. Røed (2006): "Do Business Cycle Conditions at the Time of Labor Market Entry Affect Future Employment Prospects?," *The Review of Economics and Statistics*, 88(2), 193–210.

Redding, S. J., and D. M. Sturm (2008): "The Costs of Remoteness: Evidence from German Division and Reunification," *American Economic Review*, 98(5), 1766–97.

Roback, J. (1982): "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 90(6), 1257–78.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.

Sattinger, M. (1993): "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature*, 31(2), pp. 831–880.

Schulze, S. (2009): "Einige Beobachtungen zum Pendlerverhalten in Deutschland," *HWWI Policy Paper*, 1-19.

Shu, P. (2012): "The Long-term Impact of Business Cycles on Innovation: Evidence from the Massachusetts Institute of Technology," MIT Job Market Paper.

Solon, G., R. Barsky, and J. A. Parker (1994): "Measuring the Cyclicality of Real Wages: How Important is Composition Bias," *The Quarterly Journal of Economics*, 109(1), 1–25.

Spitz-Oener, A. (2006): “Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure,” *Journal of Labor Economics*, 24(2), 235–270.

Stern, S. (2004): “Do Scientists Pay to Be Scientists?,” *Management Science*, 50(6), 835–853.

Stock, W. A., and J. J. Siegfried (2006): “Time-to-Degree for the Economics Ph.D. Class of 2001-2002,” *The American Economic Review*, 96(2), 467–474.

Stock, W. A., J. J. Siegfried, and T. A. Finegan (2011): “Completion Rates and Time-to-Degree in Economics PhD Programs (with comments by David Colander, N. Gregory Mankiw, Melissa P. McInerney, James M. Poterba),” *American Economic Review*, 101(3), 176–88.

Sullivan, D., and T. von Wachter (2009): “Job Displacement and Mortality: An Analysis Using Administrative Data,” *The Quarterly Journal of Economics*, 124(3), 1265–1306.

Tervio, M. (2008): “The Difference That CEOs Make: An Assignment Model Approach,” *American Economic Review*, 98(3), 642–68.

Topel, R. H., and M. P. Ward (1992): “Job Mobility and the Careers of Young Men,” *The Quarterly Journal of Economics*, 107(2), 439–79.

Vroman, W., and M. Wachter (1977): “Worker Upgrading and the Business Cycle,” *Brookings Papers on Economic Activity*, 1977(1), pp. 229–252.

Yamaguchi, S. (2012): “Changes in Returns to Task-Specific Skills and Gender Wage Gap,” *SSRN Working Paper*.

Appendices

Appendix A

Has Job Polarization Squeezed the Middle Class? Evidence from the Allocation of Talents

A.1 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 2010). 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 loose too many observations from the NLSY97 as, at the time of writing, data is 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 1.1 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 for the NLSY (e.g. Altonji, Bharadwaj, and Lange 2008, 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-attriters 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.[1] First, I follow ABL in excluding from the NLSY79

[1]I am extremely grateful to Prashant Bharadwaj for providing me with their data and do-files.

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 1.1 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.[2] 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 (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

[2]I thank Steve McClaskie and Jay Zagorsky for providing me with the official attrition-adjusted sample weighting program for the NLSY.

Reserve Bank.[3] 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 1.1 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.

A.2 Generalization of Results to the Three Occupation Case

In the following I derive predictions (1.7) and (1.8) for the three-occupation case. For ease of exposition, wages in occupations (1.3) are reproduced here:

$$w_{Kit} = \pi_{Kt} + \beta_{K0} + \beta_{K1} x_{1it} + ... + \beta_{KJ} x_{Jit} + u_{Kit}.$$

Note from equation (1.2) and the wages in occupations that:

$$w_{it} = \begin{cases} w_{Hit} = \pi_{Ht} + \beta_{H0} + \beta_{H1} x_{1it} + ... + \beta_{HJ} x_{Jit} + u_{Hit} & \text{if } H_{it} = 1 \\ w_{Mit} = \pi_{Mt} + \beta_{M0} + \beta_{M1} x_{1it} + ... + \beta_{MJ} x_{Jit} + u_{Mit} & \text{if } M_{it} = 1 \\ w_{Lit} = \pi_{Lt} + \beta_{L0} + \beta_{L1} x_{1it} + ... + \beta_{LJ} x_{Jit} + u_{Lit} & \text{if } L_{it} = 1 \end{cases}$$

When occupational wage rates change, by the envelope theorem, the marginal change

[3] Source: "Personal Consumption Expenditures: Chain-type Price Index (PCECTPI)", accessed 2012-8-14, `http://research.stlouisfed.org/fred2/series/PCECTPI`

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 propensity for an individual of talent vector x_{it} to enter occupation K under prices π_t. Exploiting that the three probabilities sum to one gives prediction (1.13):

$$dE(w_{it}|x_{it}, \pi_t) = d\pi_{Mt} + p_H(x_{it}, \tilde{\pi}_{HMt}, \tilde{\pi}_{LMt})d\tilde{\pi}_{HMt} + p_L(x_{it}, \tilde{\pi}_{HMt}, \tilde{\pi}_{LMt})d\tilde{\pi}_{LMt},$$

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

$$\begin{aligned} p_H(x_{it}, \tilde{\pi}_{HMt}, \tilde{\pi}_{LMt}) = Pr[&u_{Hi} - u_{Mi} > -(\pi_{Ht} - \pi_{Mt} + (\beta_H - \beta_M)'x_{it}), \\ &u_{Hi} - u_{Li} > -(\pi_{Ht} - \pi_{Lt} + (\beta_H - \beta_L)'x_{it})], \end{aligned}$$

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

For convenience, omit the dependence on x_{it} from now on. Holding constant $\tilde{\pi}_{HMt}$ and $\tilde{\pi}_{LMt}$ at $t = 0$ and integrating equation (1.13) with respect to π_{Mt} we get

$$E(w_i|\pi_{M1}, \tilde{\pi}_{HM0}, \tilde{\pi}_{LM0}) - E(w_i|\pi_{M0}, \tilde{\pi}_{HM0}, \tilde{\pi}_{LM0}) = \triangle\pi_M.$$

Similarly,

$$\begin{aligned} E(w_i|\pi_{M1}, \tilde{\pi}_{HM1}, \tilde{\pi}_{LM0}) - E(w_i|\pi_{M1}, \tilde{\pi}_{HM0}, \tilde{\pi}_{LM0}) &= \int_{\tilde{\pi}_{HM0}}^{\tilde{\pi}_{HM1}} p_H(\tilde{\pi}_{HMt}, \tilde{\pi}_{LM0})d\tilde{\pi}_{HMt} \\ E(w_i|\pi_{M1}, \tilde{\pi}_{HM1}, \tilde{\pi}_{LM1}) - E(w_i|\pi_{M1}, \tilde{\pi}_{HM1}, \tilde{\pi}_{LM0}) &= \int_{\tilde{\pi}_{LM0}}^{\tilde{\pi}_{LM1}} p_L(\tilde{\pi}_{HM1}, \tilde{\pi}_{LMt})d\tilde{\pi}_{LMt}. \end{aligned}$$

Summing these three expressions gives equation (1.14):

$$E(w_i|\pi_1) - E(w_i|\pi_0) = \triangle\pi_M + \int_{\tilde{\pi}_{HM0}}^{\tilde{\pi}_{HM1}} p_H(\tilde{\pi}_{HMt}, \tilde{\pi}_{LM0})d\tilde{\pi}_{HMt} + \int_{\tilde{\pi}_{LM0}}^{\tilde{\pi}_{LM1}} p_L(\tilde{\pi}_{HM1}, \tilde{\pi}_{LMt})d\tilde{\pi}_{LMt}$$

Finally, since we do not know the choice probabilities on the adjustment path, these have to be approximated analogously to equation (1.9) and figure 1.5

$$p_H(\tilde{\pi}_{HMt}, \tilde{\pi}_{LM0}) \approx p_H(\tilde{\pi}_{HM0}, \tilde{\pi}_{LM0}) + \frac{p_H(\tilde{\pi}_{HM1}, \tilde{\pi}_{LM1}) - p_H(\tilde{\pi}_{HM0}, \tilde{\pi}_{LM0})}{\tilde{\pi}_{HM1} - \tilde{\pi}_{HM0}}(\tilde{\pi}_{HMt} - \tilde{\pi}_{HM0})$$
$$p_L(\tilde{\pi}_{HM1}, \tilde{\pi}_{LMt}) \approx p_L(\tilde{\pi}_{HM0}, \tilde{\pi}_{LM0}) + \frac{p_L(\tilde{\pi}_{HM1}, \tilde{\pi}_{LM1}) - p_L(\tilde{\pi}_{HM0}, \tilde{\pi}_{LM0})}{\tilde{\pi}_{LM1} - \tilde{\pi}_{LM0}}(\tilde{\pi}_{LMt} - \tilde{\pi}_{LM0}),$$

which gives equation (1.16). Note 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.

A.3 Details of the Minimum Distance Estimation and Test

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.

First, I run seemingly unrelated wage and allocation regressions on the individual level in both points in time to obtain estimates $\hat{\delta}_{Ht}$, $\hat{\delta}_{Lt}$, and $\hat{\gamma}_t$ as well as an estimate of their joint covariance matrix. Second, I combine the wage rates $\triangle\tilde{\pi}_H \equiv \triangle(\pi_H - \pi_M)$, $\triangle\tilde{\pi}_L \equiv \triangle(\pi_L - \pi_M)$, and $\triangle\tilde{\pi} = [\triangle\tilde{\pi}_H, \triangle\tilde{\pi}_L]$. I also define the $J \times 1$ vectors $\hat{\triangle\gamma}$, $\hat{\bar{\delta}}_K \equiv \frac{\hat{\delta}_{K0}+\hat{\delta}_{K1}}{2}$ for $K\epsilon\{H, L\}$, and $m(\triangle\tilde{\pi}) = \hat{\triangle\gamma} - \hat{\bar{\delta}}_H\triangle\tilde{\pi}_H - \hat{\bar{\delta}}_L\triangle\tilde{\pi}_L$.

The minimum distance estimator minimizes the quadratic form

$$Q(\triangle\tilde{\pi}) = m(\triangle\tilde{\pi})'Wm(\triangle\tilde{\pi}),$$

with W being a $J \times J$ weighting matrix. Under some regularity conditions we can apply a central limit theorem to the OLS estimates $\hat{\delta}_{Ht}$, $\hat{\delta}_{Lt}$, and $\hat{\gamma}_t$ as well as to $m(\triangle\tilde{\pi})$:

$$\sqrt{N}m(\triangle\tilde{\pi}) \overset{a}{\sim} \mathcal{N}\left(Em(\triangle\tilde{\pi}), NVar(m(\triangle\tilde{\pi}))\right)$$

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

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

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

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

which provides the specification test.

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

Appendix B

The Allocation of Talent over the Business Cycle and its Long-Term Effect on Sectoral Productivity

B.1 Formal Results and Proofs

Without loss of generality, we define the density function of academic and business skills on the unit square, i.e. $f(\alpha, \beta) \geq 0$ for $\alpha, \beta \in [0, 1]$ and zero otherwise. Furthermore, rather than treating N as the *absolute* number of PhD places like in the main text, it is convenient here to redefine it to be the number of places in the PhD programs as a fraction of the whole population. As in the main text, we compare a generic boom to a generic recession cohort, i.e. $y^{Boom} > y^{Rec}$. Furthermore, a person applies for a PhD if he has skills such that $\alpha > \beta + y$.

In order to facilitate the proofs in the following, we do three more things: First, we define different sets of applicants to keep our notation concise in the rest of this section. Second, we define conditional probabilities to be able to compare different sets with each other. Third, we show that the least able (in terms of academic skills) individual admitted into academia in a recession is academically more able than the least able individual admitted in a boom. This result is used repeatedly in the proofs of the propositions.

1. The following distinct sets of applicants are used in the proofs and illustrated in Figure B.1:

   - C(onstant) applicants, who enter academia no matter what happens in the business cycle.

$$C = \{(\alpha, \beta) | \alpha \geq \alpha^{Rec} \wedge \alpha > \beta + y^{Boom}\}$$

- B(usiness inclined), who only select themselves into academia if the business climate necessitates it.

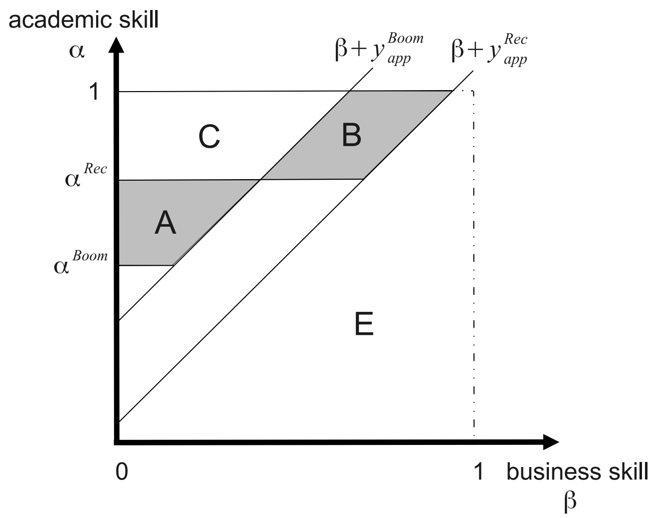
$$B = \{(\alpha, \beta) | \alpha \geq \alpha^{Rec} \wedge \beta + y^{Rec} < \alpha \leq \beta + y^{Boom}\}$$

- A(cademically inclined), who want to go into academia but only have the chance to if the group B members don't apply.

$$A = \{(\alpha, \beta) | \alpha^{Boom} \leq \alpha < \alpha^{Rec} \wedge \alpha > \beta + y^{Boom}\}$$

- E(xternals), who never go into academia.

Figure B.1: Example with a U(0,1) distribution of both skills



Note that $A \cup C$ is the boom cohort and $B \cup C$ the recession cohort. Furthermore, from our assumption that there are always more people applying for a PhD-program than there are spaces, it follows that y has an upper bound.

2. We introduce the following notation for the probability of being a member of the set X (or fulfilling the condition X) conditionally on being a member of the set Y:

$$P_Y(X) = \frac{P(X \cap Y)}{P(Y)}.$$

This conditional probability is always within [0,1] and can be interpreted as the fraction of members of Y who are also members of X. If the subscript Y is dropped, we refer to the fraction X compared to all potential applicants. As mentioned above, N is the fraction of individuals actually entering the academic sector, i.e. in a recession $N = P(C \cup B)$ and in a boom $N = P(C \cup A)$.

3. We show that the cut-off value α^s is weakly higher in recession than in boom. A higher cut-off value implies that the least able (in terms of academic skills) individual admitted into academia in a recession is academically more able than the least able individual admitted in a boom.

Lemma B.1.1 $\alpha^{Boom} \leq \alpha^{Rec}$.

Proof of lemma B.1.1: Let $g_y(\alpha) := \int_0^{\alpha-y} f(\alpha,\beta)\, d\beta$ be the percentage of students with academic skill α who will apply to the PhD-program. Obviously $y^{Boom} > y^{Rec} \Rightarrow g_{y^{Boom}} \leq g_{y^{Rec}}$ as $f \geq 0$ for all (α, β). Therefore $\alpha^{Rec} \geq \alpha^{Boom}$ as the equality $\int_{\alpha^{Rec}}^1 g_{y^{Rec}}\, d\alpha = N = \int_{\alpha^{Boom}}^1 g_{y^{Boom}}\, d\alpha$ has to hold. ∎

Proof of proposition 2.2.1: : First, note that by the definition of A and B, $P_A(x \geq \alpha) = 0$ if $\alpha > \alpha^{Rec}$ and $P_B(x \geq \alpha) = 1$ if $\alpha \leq \alpha^{Rec}$. Second, as $P(A) = P(B) = N - P(C)$ it follows that $P_{A\cup C}(x \geq \alpha) \leq P_{B\cup C}(x \geq \alpha)$, which is the definition of first order stochastic dominance. As the argumentation holds analogously for the business skills, this implies a joint stochastic dominance of academic and business skills of the recession cohort compared to the boom cohort. ∎

Proof of proposition 2.2.2: In case of $y_{grad} < y^{Boom}$ some or no people in set B leave the recession cohort and nothing changes in the boom cohort. If $y_{grad} \geq y^{Boom}$, all people in B leave. All remaining members of the recession cohort (who are member of set C and may or may not leave) are a subset of the boom cohort and therefore behave alike. Note that, as P(B) = P(A) and all members of B, but potentially only some members of A, leave for $y_{grad} \geq y^{Boom}$, there are always more leavers in the recession than in the boom cohort. ∎

Proof of proposition 2.2.3: Let B' be a subset of B. We show that $C \cup B'$ first order stochastically dominates $C \cup A$ in the partial distribution of academic skill, which is the proposition for $y^{grad} < y^{Boom}$. It follows for all α that

$$P_{C\cup B'}(x \geq \alpha) = P_{C\cup B'}(C)P_C(x \geq \alpha) + P_{C\cup B'}(B')P_{B'}(x \geq \alpha),$$

and analogously $P_{C\cup A}(x \geq \alpha) = P_{C\cup A}(C)P_C(x \geq \alpha) + P_{C\cup A}(A)P_A(x \geq \alpha)$. This means that the percentage of members in C and B' who have an academic skill larger than some arbitrary α is the weighted sum of the percentage of members in C and of the percentage of members in B' who have at least such a high academic skill. The respective weights are the percentage of members of C in $C \cup B'$ and the percentage of B' in $C \cup B'$. (Remember that $P_{C\cup B'}(C)$ is the percentage of members of C in the union of C and B'.)

Now one can show as in Proposition 2.2 :

- $P_{C\cup B'}(x \geq \alpha) \geq P_{C\cup B'}(C)P_C(x \geq \alpha) \geq P_{C\cup A}(C)P_C(x \geq \alpha) = P_{C\cup A}(x \geq \alpha)$ for $\alpha \geq \alpha^{Rec}$.

  The first inequality holds by the decomposition of $P_{C\cup B'}(x \geq \alpha)$ above, the second inequality holds because $P(A) = P(B)$ and the equality holds because $P_A(x \geq \alpha) = 0$ for $\alpha \geq \alpha^{Rec}$ by definition of the set A.

- $P_{C\cup B'}(x \geq \alpha) = 1 \geq P_{C\cup A}(C)\underbrace{P_C(x \geq \alpha)}_{=1} + P_{C\cup A}(A)P_A(x \geq \alpha) = P_{C\cup A}(x \geq \alpha)$ for $\alpha < \alpha^{Rec}$. The first equality holds by the definition of C and B', the first inequality by the definition of probability measures (it cannot exceed one) and the second equality holds by the definition above.

These two statements taken together prove the first order stochastic dominance in the partial distribution of the academic skill for the recession cohort compared to the boom cohort.

Note, that the same argument can be made if $y^{grad} \geq y^{Boom}$ with A' and C' being subsets of A and C, respectively, and $B' = \emptyset$. This completes the proof. ∎

For the proof of the last proposition we require one further piece of notation: Let y^{Boom}_{grad} denote the business cycle variable if there is a boom at graduation and y^{Rec}_{grad} if there is a recession at graduation. Note that $y^{Boom}_{grad} > y^{Rec}_{grad}$ and therefore $w^B_{Boom} = \beta + y^{Boom}_{grad} > w^B_{Rec} = \beta + y^{Rec}_{grad}$.

Proof of proposition 2.2.4 with unlimited academic jobs: The PhD students with $\{\alpha, \beta) | \beta + y^{Rec}_{grad} < \alpha \leq \beta + y^{Boom}_{grad}\}$ leave academia when there is a boom instead of a recession at graduation. As this set can be non-empty, weakly more students leave in a boom than in a recession. ∎

Proof of proposition 2.2.4 with a fixed number of jobs: Note that the decision to enter academia with a fixed number of jobs, N_{grad}, is very similar to the decision to enter the PhD program with a fixed number of PhD spaces. By applying the proof of lemma B.1.1 analogously, it is easy to show that the minimum academic skill to enter academia is higher for the recession-at-graduation cohort than for the boom cohort:

Let $h_{y_{grad}}(\alpha) := \int_0^{\alpha - y_{grad}} f(\alpha, \beta)\, d\beta$ be the percentage of students with academic skill α who will apply to the academic job market. Obviously $y_{grad}^{Boom} > y_{grad}^{Rec} \Rightarrow h_{y_{grad}^{Boom}} \leq h_{y_{grad}^{Rec}}$ as $h \geq 0$ for all (α, β). Therefore $\alpha_{grad}^{Rec} \geq \alpha^{Boom}$ because the equality $\int_{\alpha^{Rec}}^1 h_{y_{grad}^{Rec}}\, d\alpha = N_{grad} = \int_{\alpha^{Boom}}^1 g_{y_{grad}^{Boom}}\, d\alpha$ has to hold.

By applying the proof of proposition 2.2.1 analogously, the academic skill of the recession-at-graduation cohort first order stochastically dominate the academic skills of the boom-at-graduation cohort. ∎

B.2 Cyclicality of Academia versus Business

In our theory section we assume that compensation in the academic sector is less cyclical than in the business sector. In this section we provide evidence that this is a reasonable assumption. We focus on the cyclicality in the attractiveness of academia versus business only at graduation from the PhD. At application, graduate school seems to be clearly less cyclical than business—as was illustrated by the flood of applications to masters and PhD programs during the crisis of 2008/09 (see also Bedard and Herman 2008, Gustman and Steinmeier 1981, Black and Sufi 2002).

Ideally, we would like to compare the variability of the total expected lifetime compensation (consisting of pecuniary and non-pecuniary benefits) for the two sectors over the business cycle. Unfortunately, this is not possible for two reasons: First, (variabilities in) non-pecuniary benefits are hard to observe and difficult to compare across jobs. Second, even the monetary component of compensation is difficult to obtain or to approximate. Wage data for the business sector is not consistently available on a yearly basis over longer time periods for economics PhDs.[1]

[1] We do not have access to any employer-employee matched dataset as in Oreopoulos, von Wachter, and Heisz (2012).

Furthermore, even if wages were available, they are a result of the selection process we are trying to explain (e.g. Solon, Barsky, and Parker 1994). Consequently, it would be sensible to focus on wage offers in both sectors as used by Scott Stern in a similar setting (Stern 2004). Unfortunately we are unable to find such data.

In the following we approximate the relative attractiveness of the academic sector by comparing the number of academic versus non-academic job offers for economists over the business cycle.[2] The underlying assumption is that an additional vacancy (weakly) increases a sector's relative attractiveness. The number of new jobs is published annually in the Job Openings for Economists (JOE) director's report in the American Economic Review's Papers and Proceedings issue in May. The academic and non-academic openings are broken up by new and total jobs and listings (employers). Since we want to approximate the decision situation of a graduate in year t during his job market year, we focus on the sum of new job offers from August in year $t-1$ to July in year t.[3,4]

Figure B.2 plots the yearly sums of job offers over the years from 1977 to 2010. Academic jobs are displayed in the upper-left panel and non-academic jobs in the upper right panel. In the lower panel the overall number of job offers is plotted together with the number of academic per non-academic jobs. Academic and non-academic jobs move together in lockstep, which shows that the academic sector is in fact quite cyclical. However, the relative number of academic jobs to non-academic jobs appears to be countercyclical: even when the number of academic jobs rise, the number of non-academic jobs rises relatively more. The reverse is true in recessions. Therefore, graduates have relatively more business jobs (compared to academic jobs) to choose from in booms than in recessions.

To formally test if business jobs are indeed more pro-cyclical than academic jobs, we estimate the following system of equations

$$\log(\#\text{ Academic jobs})_t = \beta_{Academic} \cdot \mathrm{y}_t + \delta \cdot \text{controls} + \epsilon_t \tag{B.1}$$

$$\log(\#\text{ Non-Academic jobs})_t = \beta_{Non-Academic} \cdot \mathrm{y}_t + \delta \cdot \text{controls} + \epsilon_t \tag{B.2}$$

[2] Oyer (2006) uses the academic job offers as a measure of demand for economists in academia.

[3] The seasonality of job offers within a given year follows the job market for each cohort, especially for academic jobs. Job offers reach their trough in June after which they start rising. They literally jump up in October and stay high during the fall after which they decline. We therefore define each yearly sum of job offers according to job market years instead of calender years.

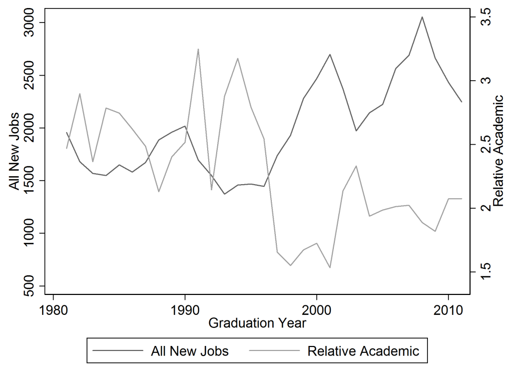
[4] We do not use total jobs as we do not know if these jobs are double counted in several months.

Figure B.2: Academic and non-academic job offers over time

(a) Academic job offers

(b) Non-academic job offers

(c) Overall job and relative Academic Job Offers



where the dependent variables are the log of the number of new academic and non-academic jobs, respectively, and y_t is a measure of the business cycle. Then we test if the business cycle has a larger influence on the number of non-academic jobs than on the number of academic jobs, i.e. if $\beta_{Non-Academic}$ is larger than $\beta_{Academic}$ in absolute values.

The regressor y_t is one of four business cycle measures: recession indicators, unemployment levels and changes, and the log of GDP. The business cycle variables are measured in October of the year before graduation when the mode of job offers for each cohort comes in. The controls include dummies for the switch from seven to ten monthly reports of job offers in 1999 and the JOE going online in 1995 interacted with a linear time trend. We estimate the outlined specification in levels with a time trend and in first differences. We do this to control for the potential trend or the non-stationarity of dependent and independent variables.

Table B.1 and Table B.2 report the results of these regressions in levels and in first differences. Unemployment and GDP are significantly related to academic

Table B.1: Differing Cyclicality of Academic and Non-Academic Jobs—Levels

	log(# Academic Jobs)	log(# Non-Academic Jobs)	z-Value
Unemployment	-0.05***	-0.09***	2.23**
	(0.01)	(0.02)	
GDP	1.96***	4.06***	-2.32***
	(0.62)	(1.08)	
Recession	0.02	-0.08	1.57*
	(0.05)	(0.09)	

NOTE.—Standard errors in parentheses. The z-Value is the test statistic of a one-sided test. for $|\beta_{Non-Academic}| > |\beta_{Academic}|$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.2: Differing Cyclicality of Academic and Non-Academic Jobs—First Differences

	FD log(# Academic Jobs)	FD log(# Non-Academic Jobs)	z-Value
Unempl Change	-0.05***	-0.06**	0.58
	(0.01)	(0.03)	
GDP Growth	2.57***	3.09**	-0.39
	(0.62)	(1.34)	
FD Recession	-0.00	-0.09	1.66**
	(0.03)	(0.06)	

NOTE.—Standard errors in parentheses. The z-Value is the test statistic for a one-sided test. for $|\beta_{Non-Academic}| > |\beta_{Academic}|$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

and non-academic job offers in the way that we would have expected from Figure B.2. Moreover, in levels, the relationship is significantly stronger for non-academic than for academic jobs. For example, a one percentage point increase in unemployment is approximately associated with a nine percent decrease in the number of non-academic jobs and "only" a five percent decrease in academic jobs. Recession indicators do not work that well. Although they are significantly different from each other in the right direction, the estimated coefficients are not significantly different from zero on their own. These results are qualitatively robust to using total job openings instead of focusing on new ones, variations in the control variables (e.g. including quadratic time trends), and a sensible alternative timing of the business cycle variables.

Overall, we would state that we find reasonable support for the assumption that the academic sector is less cyclical than the non-academic sector in the job openings for economists. We think this is some prima facie evidence for our assumption that in downturns the academic sector becomes relatively more attractive as an employer compared to the business sector. Moreover, we think that the above exercise is conservative because of the following reason: the (variation in the) number of job offers is unlikely to approximate well the (variation in) non-pecuniary benefits, which are substantial and probably stable in research related jobs (see Stern 2004). Thus, total compensation in the academic sector might be less cyclical than indicated by the number of job openings.

B.3 Cohort Sizes and Timing of Graduation

This section addresses potential concerns about factors that might confound our results and analyzes possible impacts on our estimates. In the following we address concerns about the size of the entry and exit cohort and the timing of graduation. Lastly, we address a potential correlation of the business cycle at application and graduation.

In order to do this, we calculate the number of graduates from our dataset (in the following listed as "# of Graduates (AEA)") and match it with the business cycle at application and at graduation. For conciseness, we focus on unemployment change as our preferred measure for the business cycle. Then, we supplement this data with data from the National Science Foundation's "Survey of Earned Doctorates" and the "Survey of Graduate Students and Postdoctorates in Science and Engineering".[5] From there we obtain the number of full-time, first-time graduate students ("# of Entrants (NSF)") and awarded doctorates ("# of Graduates (NSF)") for our top 30 universities since 1977.[6,7]

We report the partial correlation coefficient of unemployment change at application and at graduation with application and graduation numbers in Table B.3. In order to obtain the correct standard errors we aggregate the data to yearly averages. To keep this section concise, we only report for unemployment change and not for all four business cycle measures. These correlation tables are available upon request from the authors.

One might have the concern that the number of students admitted to the PhD

[5] These surveys are publicly available through the WebCASPAR Interface: "WebCASPAR Integrated Science and Engineering Resource Data System — NSF Survey of Earned Doctorates/Doctorate Records File," National Science Foundation, last accessed 2012-03-16, `https://webcaspar.nsf.gov/`.

[6] The number of full-time, first-time graduate students is only an imperfect proxy for the number of students entering a PhD, because it also includes master students.

[7] In previous versions we erroneously used NSF data on full-time, first-time graduate students and doctorates for all universities in the NSF sample.

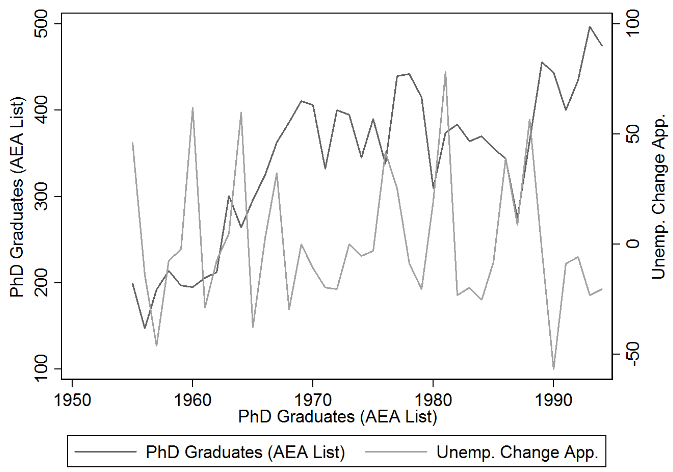
Table B.3: Correlation of Unemployment Change with the Number of Entrants and Graduates

	Unempl Change (Grad.)	Unempl Change (App.)	# Graduates (AEA)	# Graduates (NSF)	# Entrants (NSF)
Unempl Change (Grad.)	1.00				
Unempl Change (App.)	-0.13	1.00			
	(0.42)				
# Graduates (AEA)	0.02	-0.17	1.00		
	(0.91)	(0.30)			
# Graduates (NSF)	0.16	-0.12	0.35*	1.00	
	(0.41)	(0.52)	(0.06)		
# Entrants (NSF)	0.03	0.09	0.02	0.22	1.00
	(0.86)	(0.62)	(0.91)	(0.15)	
Observations	56				

p-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

systematically increases (decreases) in recessions.[8] Within the framework of our model, this would weaken (strengthen) the selection effect at application. However, according to the NSF data, the business cycle at application is not related to the number of full-time, first-time graduate students (compare row five, column two in Table B.3). This supports our assumption in the main text. Moreover, the number of graduates in the NSF data and our data (AEA) is unrelated to unemployment change at application to the PhD (compare row three and four, column two in Table B.3 and Figure B.3).

Figure B.3: Number of graduates and unemployment change at application



Another concern is that PhDs might time their graduation in order to circum-

[8]For example, an increase in PhD entrants during recessions may even happen if universities do not intend to increase their intake but more of the successful applicants take up their offers.

vent entering the private or the academic job market during a time of recession.[9] The effect of such graduation timing on our parameter estimates would depend on whether the high- or the low skilled delay their graduation date: If it is the academically strong students who delay their graduation during recession, this would lead to a downward-bias of our estimates at graduation on productivity and vice versa if it is the academically weak students who delay. In both cases, we only add measurement error to our results at application biasing our estimates towards zero, as long as the business cycle at application and graduation are not related. In general, since it is unlikely that either group has an incentive to graduate during recession, graduation timing should lead to procyclical cohort sizes. We do not find a relation between unemployment change and graduation numbers according to the NSF data and the AEA doctoral listings (compare row three and four, column one in Table B.3), which suggests that graduation timing is not much of an issue. This supports our results in the main text as well as the assumption of no graduation timing in Oyer (2006).

Finally, a last concern might be that, contrary to our assumption in the model, the business cycle is systematically correlated with itself in the six years between a cohort's application and graduation. Table B.4 reports this and the contemporaneous correlation exemplary for unemployment change and GDP growth. The correlation table with unemployment levels and recession indicators are available upon request from the authors. Unsurprisingly both measures are strongly contemporaneously related. However, there is no significant correlation, neither of unemployment change nor GDP change, between the time of application and graduation. If at all, there may be a very slightly reversing relationship over the six years. This could imply that we potentially underestimate the effect of the business cycle on academic performance because a recession cohort at graduation is more likely a boom cohort at application (and thus is inherently not as able) and vice versa for a boom cohort at graduation. For the same reason we might in this case overestimate the effect of the business cycle on the career decision (i.e. the academic variable) at application and at graduation.

[9] In Appendix B.2 we document that also academic job offers decline during recession.

Table B.4: Correlation of Unemployment Change and GDP Change at Application and at Graduation

	Unempl Change (App.)	Unempl Change (Grad.)	GDP Growth (App.)	GDP Growth (Grad.)
Unempl Change (App.)	1.00			
Unempl Change (Grad.)	-0.15	1.00		
	(0.27)			
GDP Growth (App.)	-0.79***	0.16	1.00	
	(0.00)	(0.25)		
GDP Growth (Grad.)	0.13	-0.86***	-0.11	1.00
	(0.34)	(0.00)	(0.41)	
Observations	57			

p-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.4 Robustness

B.4.1 Alternative Measures for Productivity

In this section we consider three alternative measures of academic productivity in Table B.5: the number of top five articles, the h-value and the raw number of articles.[10]

The h-index (Hirsch index or Hirsch number) is a measure based on citations and number of articles. The last measure is the raw number of articles written as recorded in JSTOR. In Table B.5 we report the results for these three alternative productivity measures for the full and the academic subsample. All mean estimates for every business cycle measure point in the same direction as the dynamic performance measure in the main text and as the selection theory predicts.

B.4.2 The Tier 1 Subsample

Next, we repeat our main regression for individuals who graduated from the elite Tier 1 schools. According to Table B.6, the magnitude of the effects appears to be larger in all considered dimensions. The estimated coefficients are in some specifications more, and in some specifications less, significant than in the main text. Taken together, the results for the Tier 1 graduates support our findings in the main text.

[10]We classify articles in “Econometrica”, “The American Economic Review”, “The Quarterly Journal of Economics”, “The Review of Economic Studies”, “The Journal of Political Economy” and “The Journal of Finance” as top journal articles.

Table B.5: Alternative Productivity Measures

	Top Five	h-index	# of Articles	Top Five	h-index	# of Articles
Unempl Change (App.)	1.51	1.14	1.21	3.48**	3.11*	5.65
	(1.02)	(1.13)	(3.24)	(1.39)	(1.59)	(4.81)
Unempl Change (Grad.)	3.93***	3.97***	5.01**	4.91***	4.81***	4.62
	(0.91)	(0.87)	(2.17)	(1.37)	(1.54)	(4.40)
Unemployment (App.)	0.62	0.97	3.36	1.94	2.54	7.75
	(1.08)	(1.14)	(2.78)	(1.74)	(1.90)	(4.96)
Unemployment (Grad.)	0.78	1.27	2.28	1.65	2.49	5.45
	(1.13)	(1.15)	(2.65)	(1.66)	(1.74)	(4.29)
GDP Growth (App.)	-0.59	-0.46	-0.46	-1.49**	-1.36*	-2.64
	(0.45)	(0.49)	(1.41)	(0.62)	(0.70)	(2.18)
GDP Growth (Grad.)	-1.26**	-1.24**	-1.21	-1.51**	-1.36*	-0.80
	(0.49)	(0.49)	(1.17)	(0.72)	(0.79)	(2.12)
Recession (App.)	-0.62	-0.70	-1.47	1.77	1.99	5.44
	(3.47)	(3.54)	(8.70)	(5.03)	(5.11)	(12.69)
Recession (Grad.)	6.20**	6.50**	6.97	6.87	7.08	4.22
	(2.84)	(2.82)	(6.11)	(4.45)	(4.82)	(11.18)
Subsample	All	All	All	Academic	Academic	Academic
Univ-Decade Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1068	1068	1068	1047	1047	1047

NOTE.—Standard errors clustered on graduation year in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.6: Main Regression Results (Tier 1)

	Productivity	Academic	Productivity
Unempl Change (Application)	5.39**	-1.72***	9.84***
	(2.14)	(0.58)	(2.96)
Unempl Change (Graduation)	4.34*	2.87***	3.94
	(2.39)	(0.94)	(3.45)
Unemployment (Application)	3.16	-1.23	5.87*
	(2.04)	(1.03)	(3.15)
Unemployment (Graduation)	2.55	-0.07	3.72
	(2.46)	(0.92)	(3.96)
GDP Growth (Application)	-1.99**	0.75**	-3.68***
	(0.90)	(0.29)	(1.24)
GDP Growth (Graduation)	-1.25	-1.25***	-0.83
	(1.10)	(0.36)	(1.57)
Recession (Application)	7.48	-5.73***	16.83*
	(6.15)	(1.73)	(8.49)
Recession (Graduation)	5.38	3.95**	4.08
	(6.91)	(1.67)	(9.93)
Subsample	Tier 1	Tier 1	Tier 1 Academic
University-Decade Dummies	Yes	Yes	Yes
Observations	234	234	232

NOTE.—Standard errors clustered on graduation year in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.4.3 Length of the PhD

In our main analysis we subtract six years, the median duration of a PhD, from the graduation date and then use our measure of the business cycle at this date as macroeconomic variation at entry. The median duration of a PhD stayed almost constant at around five to six years since the 1970s according to the data assembled in Table B.7.

Using the median duration of the PhD might be questionable, because there

Table B.7: Duration of a PhD

Year	1977	1986	1996	1997	2001
	5.7	6.3	5.3	5.25	5.5
	Median years of registered time to PhD	Median years of registered time to PhD	Time-to-degree	Median time-to-degree	Time-to-degree
Source	Hansen (1991)	Hansen (1991)	NSF*	Stock, Siegfried, and Finegan (2011)	NSF*

NOTE.—*NSF duration data includes masters degrees, therefore we subtract 1.5 years.

might be substantial variability in the duration of a PhD. Therefore we repeat our main analysis with a weighted average of the respective business cycle measure at application according to the distribution of completion times for the year 1997 described in Stock, Siegfried, and Finegan (2011). The results are reported in Table B.8. Note that the regressors have a much lower variation because we compute moving averages here. Thus, if we want to compare the results in Table B.8 to our main regressions in Table 2.5, we need to divide the point estimates for unemployment levels by about 1.2 and for the other regressors by about 2.6. Nonetheless, the mean estimates in Table B.8 are larger and more significant than in the main text. This suggests that the latter might be downward biased due to measurement error.

Table B.8: The Regression Results Using "Weighted Average" of PhD Entry

	Productivity	Academic	Productivity
Unempl Change (Application)	3.99**	-6.00***	10.01***
	(1.71)	(1.52)	(2.90)
Unempl Change (Graduation)	2.33***	1.27***	2.92**
	(0.65)	(0.45)	(1.20)
Unemployment (Application)	2.33**	-1.17	4.48***
	(0.96)	(1.44)	(1.65)
Unemployment (Graduation)	1.89***	-0.35	3.39***
	(0.67)	(0.67)	(1.18)
GDP Growth (Application)	-1.31	2.55***	-3.35**
	(0.86)	(0.58)	(1.29)
GDP Growth (Graduation)	-0.70**	-0.38	-0.80
	(0.34)	(0.25)	(0.60)
Recession (Application)	14.92**	-14.48**	33.86***
	(6.43)	(6.03)	(10.41)
Recession (Graduation)	5.25***	1.34	6.89**
	(1.86)	(1.24)	(3.01)
Subsample	All	All	Academic
University-Decade Dummies	Yes	Yes	Yes
Observations	1023	1023	1005

NOTE.—Standard errors clustered on graduation year in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.4.4 Foreign Students

One concern that was expressed to us is that foreign students may go back to their home country after the PhD. For example, Borjas (2006) shows that the share of foreign doctoral students has more than doubled since the 1970s. If hiring in the academic sector in the US is cyclical too, one might imagine that, in recessions, more foreign students go back to academic jobs in their respective home countries. We do not have information about whether students are natives or foreigners in our dataset. In terms of our model, if there are foreign academic programs whose hiring is less correlated with the US business cycle than US schools' hiring, this makes demand for economists more inelastic. If those graduates who take the option to go back more often in recessions appear in the faculty listings, the AEA listings, or if they publish in ranked journals, they are counted as academics. This fits our story. If they are not counted as academics, our estimates in Table 2.5 will understate the effect of the business cycle at graduation on the propensity to become an academic and, depending on whether it is the high-α or the low-α PhDs who react more to this, our estimates will under- or overstate the effect on the publications per graduate. Note that our model does not make predictions on the latter effect.

Another possible effect involving foreign students may be at PhD entry. For example, a recession in a big foreign sending country and a simultaneous boom in the US might lead to a higher proportion of foreign students starting a US PhD program. Since foreigners are more likely to go back to (academic positions in) their home countries after the PhD—and listed publications and AEA membership are less likely abroad—we might mistake them for having left academia. This may downward-bias our effect of the business cycle at PhD application on the likelihood to become an academic. In unreported robustness checks we therefore assemble data from the "Survey of Earned Doctorates" and examine how the fraction of foreign PhD entrants and graduates is correlated with the business cycle. We do not find a relationship between those variables. Furthermore if we control for the fraction of foreigners in a graduation cohort in our main regression, all our results remain the same.

B.4.5 Time Trend as Control

One concern might be that our graduation decade dummies inadequately control for the general trends in academia over time. In Table B.9 we therefore report the main regression with university dummies and a linear, quadratic and cubic time trend instead. The results of the main section on productivity are largely robust. Only the productivity of academics at graduation is not significant anymore, but theory made no prediction for the signs of this parameter in the first place. The results on the propensity to become an academic have the right sign (except at graduation when using unemployment levels) and at application they are significant at the 10% level.

Table B.9: Alternative controls: Time Trend

	Productivity	Academic	Productivity
Unempl Change (Application)	0.90*	-0.57	3.03***
	(0.51)	(0.35)	(0.99)
Unempl Change (Graduation)	0.65	0.39	1.44
	(0.46)	(0.39)	(1.20)
Unemployment (Application)	1.47***	0.05	2.83***
	(0.39)	(0.33)	(0.94)
Unemployment (Graduation)	0.57	-0.45	2.10*
	(0.42)	(0.34)	(1.21)
GDP Growth (Application)	-0.45*	0.25	-1.40***
	(0.22)	(0.16)	(0.46)
GDP Growth (Graduation)	-0.19	-0.07	-0.45
	(0.24)	(0.18)	(0.55)
Recession (Application)	2.11	-1.93*	6.56**
	(1.69)	(1.15)	(3.06)
Recession (Graduation)	1.74	-0.11	3.11
	(1.45)	(0.70)	(3.59)
Subsample	All	All	Academic
Time trend	Yes	Yes	Yes
Observations	3195	3195	1455

NOTE.—Standard errors clustered on graduation year in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.4.6 Controlling for Subfields

In our main regression we might bias our estimates by not controlling for the subfield of the considered PhD student. For example, if (hypothetically) during a recession more students chose macroeconomics and macroeconomic papers are published better on average, our result might be driven by a change in subfield choices. To address this concern we collect from the AEA listings the subfield for each considered PhD graduate. The considered subfields are Microeconomics, Macroeconomics, Econometrics/Statistics, Labor Economics, Industrial Organization, Public Economics, International Economics, Development Economics and History.

In an unreported regression we only find very weak evidence, that students change their subfield if they apply or graduate in a recession.[11] Nevertheless, in Table B.10 we repeat our main regressions with additional subfield fixed effects. Our main results are both, quantitatively and qualitatively robust to the inclusion of these additional fixed effects.

Table B.10: Alternative controls: Subfields

	Productivity	Academic	Productivity
Unempl Change (Application)	1.59**	-1.00*	3.52***
	(0.60)	(0.56)	(0.98)
Unempl Change (Graduation)	2.16***	1.31*	2.52**
	(0.62)	(0.68)	(1.14)
Unemployment (Application)	1.67**	-0.60	3.11**
	(0.66)	(0.72)	(1.27)
Unemployment (Graduation)	1.65**	-0.30	2.77**
	(0.62)	(0.59)	(1.17)
GDP Growth (Application)	-0.66**	0.56**	-1.57***
	(0.26)	(0.24)	(0.42)
GDP Growth (Graduation)	-0.62*	-0.39	-0.62
	(0.31)	(0.27)	(0.53)
Recession (Application)	2.19	-3.23**	5.51*
	(1.89)	(1.46)	(3.06)
Recession (Graduation)	4.43**	1.86	4.91
	(1.81)	(1.44)	(3.27)
Subsample	All	All	Academic
University-Decade Dummies	Yes	Yes	Yes
Subfield-Decade Dummies	Yes	Yes	Yes
Observations	5612	5608	4487

NOTE.—Standard errors clustered on graduation year in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.4.7 Placement Bias at Graduation

As noted in the main text, the estimated coefficients for the business cycle at graduation may measure the combined impact of the selection effect described in our theory section and the placement effect. In contrast, the estimated influence of the business cycle at application measures cleanly the selection effect because we control in all regressions for the degree-granting university, i.e. for the placement to different PhD programs. In this subsection we aim to sign the direction of the bias at graduation.

According to Oyer (2006), the first placements of graduates are on average worse

[11] These results are left out for conciseness and available from the authors on request.

in a recession. This leads in turn to fewer publications, as assistant professors publish significantly less at lower ranked institutions. It should therefore induce a bias towards zero in our estimates at graduation. In our study, we cannot estimate the exact size of the bias because we do not have comprehensive placement data for the last 50 years for the universe of US PhD students. Even if this data were available, controlling for quality of the first job is not straightforward since the first placement is influenced by the ability of the PhD graduate and therefore is an outcome of the selection process described in this article. Oyer (2006) tackles this problem by using the aggregated demand for PhD graduates as an instrument for first placements.

To get a sense of the direction of the placement bias, we substitute in the following our business cycle measure at graduation with the total number of new academic and non-academic jobs for economists from the "Job Openings for Economists" published by the American Economics Association. This data is only available for the shorter time-span from 1977 to 1994 and is also used in Appendix B.2. More academic job offers influence the placement according to Oyer (2006) but may also increase the number of graduates entering academia. In contrast, more non-academic job offers only influence the number of economists entering academia - if more jobs are available in the private sector more graduates want to go into business.

In the following we estimate our main specification, controlling for the number of non-academic job offers in the first specification, and controlling for both, academic and non-academic job offers, in the second specification. In the first specification the coefficient on non-academic job offers is a combination combination of selection and placement effect because non-academic job offers are correlated with the omitted academic job offers (see Appendix B.2). Our estimate should be biased towards zero. If we control for both types of job offers, then the estimated coefficient on non-academic jobs gives us the pure selection effect induced by an increase in the business sector's attractiveness at graduation, while the estimated coefficient for academic job offers remains a combination of the placement and the selection effect. Therefore academic job offers act as an imperfect proxy for placements and academic demand, and thus mitigate the bias on the coefficient for non-academic job offers.

We report the results of this exercise in Table B.11. More non-academic jobs lead to fewer graduates staying in academia supporting the idea that the available outside options drive the selection into academia. The reported coefficients are

Table B.11: Controlling for the Demand for Economists

	Academic	Academic	Academic	Productivity	Productivity	Productivity
Unempl Change (Application)	-1.48	-1.39*	-1.23*	2.82*	2.80*	2.95*
	(1.02)	(0.78)	(0.67)	(1.43)	(1.42)	(1.43)
log(Non-Academic Jobs)		-13.24**	-18.97*		1.88	-3.75
		(5.42)	(9.94)		(10.97)	(13.38)
log(Academic Jobs)			12.42			12.16
			(10.36)			(15.97)
Unemployment (Application)	-0.56	-0.02	-0.18	1.29	1.33	1.30
	(0.91)	(0.83)	(0.77)	(1.35)	(1.42)	(1.45)
log(Non-Academic Jobs)		-13.79**	-20.44**		-0.94	-2.13
		(5.18)	(9.26)		(11.70)	(11.65)
log(Academic Jobs)			15.65			2.81
			(10.71)			(15.81)
GDP Growth (Application)	0.74**	0.67**	0.61**	-1.46*	-1.46*	-1.50**
	(0.30)	(0.26)	(0.23)	(0.70)	(0.71)	(0.69)
log(Non-Academic Jobs)		-12.71**	-18.85*		0.80	-3.99
		(5.12)	(9.11)		(10.82)	(13.73)
log(Academic Jobs)			13.28			10.33
			(9.40)			(16.32)
Recession (Application)	-2.66	-2.62*	-2.26	0.64	0.62	0.77
	(2.17)	(1.33)	(1.32)	(3.89)	(3.87)	(3.85)
log(Non-Academic Jobs)		-13.79***	-20.08**		3.20	0.50
		(4.51)	(8.98)		(11.32)	(12.11)
log(Academic Jobs)			13.79			5.91
			(10.56)			(15.37)
Subsample	All	All	All	Academic	Academic	Academic
University-Decade Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	500	500	500	490	490	490

NOTE.—Standard errors clustered on graduation year in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

significantly different from zero at least on the 10% level. The mean estimate of the number of academic jobs is positive, but not significantly different from zero. In the productivity regressions, all coefficients are imprecisely estimated and therefore not significantly different from zero. If we interpret the mean estimates, we find a negative impact of non-academic jobs on productivity but only if we control for the number of academic jobs. For all business cycle measures the estimated coefficient for non-academic jobs is consistently more positive if we leave out the number of academic job offers. This points to the expected upward bias due to the placement effect of Oyer (2006).