THE UNIVERSITY OF CHICAGO

LEARNING SPEEDS AND LABOR MARKET ADJUSTMENTS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY
ROBERT JACKMAN

CHICAGO, ILLINOIS
JUNE 2020

Contents

List of Figures

List of Tables

Acknowledgments

I give a heartfelt thanks to my committee, Robert Shimer, Thibaut Lamadon, and Felix Tintelnot. It has been a pleasure working with you through these years, and I am grateful for your guidance. Collectively, many others at the University of Chicago have provided invaluable suggestions and support. A non-exhaustive list includes Rodrigo Adão, Fernando Alvarez, José Ignacio Cuesta, Jonathan Dingel, Andrew Gianou, Juanna Joensen, Robert Lucas Jr., Simon Mongey, Casey Mulligan, Karthik Nagarajan, David Novgorodsky, Mariel Schwartz, Alejandro Hoyos Suarez, and Karen Ye.

I owe a debt of gratitude to many at the Department of Economics and Business Economics at Aarhus University, with Elena Mattana and Rune Vejlin in particular. I also thank Jonas Bødker, Henning Bunzel, Andrei Gorshkov, Mads Hejlsen, Eva Rye Johansen, Anne Brink Nandrup, for support, helpful comments, and an inexhaustible willingness to talk about Denmark.

Finally, thank you to my family: Stefani, Graydon, Mom, and Dad. Graduate school can be tough on families, and I thank you for being at my side throughout. Your consistent support and encouragement will always stay with me.

Abstract

As the economy changes, some workers confront a labor market that no longer demands their set of skills. While these workers can adapt by learning new skills, not all workers learn at the same rate. This paper asks how heterogeneous learning rates influence labor market outcomes and occupational choice, particularly under economic volatility. First, using Danish administrative data, I document heterogeneous responses to job loss from establishment closures. For workers who earned high grades in secondary education, annual earnings rebound to those of their peers in the years following job loss. Workers with lower grades suffer a permanent earnings loss relative to their peers, showing no recovery whatsoever. Second, I study occupational choice in a dynamic Roy model where workers learn on the job at varying speeds. I estimate these learning speeds alongside the evolution of unobserved skills using a non-linear Kalman Filter, corrected to accurately estimate posterior covariances. High-type learners acquire new cognitive-based skills 65% more quickly than low-type learners, however that pattern is flipped for manual skills. High-type learners choose occupations they are initially unsuited to, especially early in their careers. This early mismatch delivers mid-career benefits in the form of higher skills and earnings. A higher learning rate also allows for quicker adjustment and recovery in the face of adverse shocks to occupational prices.

I. Introduction

The labor market is, at its heart, a market for human capital. Workers' experiences in the market – choosing and changing jobs, celebrating raises, handling layoffs – are simultaneously causes and consequences of human capital. This paper studies how different workers acquire human capital, and how those differences influence their careers. To do so, I focus on three facets of the labor market: recovery from adverse events, occupational choice, and wage growth, all of which are closely tied to human capital accumulation. I argue that workers acquire human capital at different rates, and these differing rates drive life-cycle labor market inequality. My hypothesis: what most shapes careers is not the possession of a stock of human capital, but the capacity to acquire it.

This paper pays special attention to recovery from adverse labor market shocks such as a layoffs, being fired, and wage/hours reductions. These can cause significant long run harm to workers in terms of both wages and wellbeing (Jacobson, LaLonde, and Sullivan (1993), Sullivan and von Wachter (2009), Eliason and Storrie (2006)). However, the long run costs are not uniform across workers, and this paper joins other recent work (e.g. Seim (2019)) asking why. While the outcome heterogeneity might be explained by a plethora of factors, I offer evidence that the ability to acquire new human capital quickly is a particularly important factor. Through an analysis of Danes who lost jobs in establishment closure, I show that quick learners—-defined using grades from secondary education and educational achievement—enjoy a long run recovery not only with respect to economy-wide averages, but also relative to their peers.[1] In contrast, all other types of workers suffer the persistent losses relative to their peers.

Throughout this paper, I use administrative data from Statistics Denmark. These feature a matched employer-employee dataset covering the universe of Danish workers from 1991

[1] In this respect, my argument echoes Neal (1998). I will, however, use a different mechanism that features differing implications.

through 2016, offering an unusually long panel for a population of this size. They include all Danish jobs over the 25 year period, as well as very good coverage of earnings and occupations by job. Additionally, I have secondary education grades for a subset of workers, a useful observation to identify learning types. On the employer side, I observe establishment closure as well as a variety of statistics describing each establishment's workforce.

The depth and breadth of these data allow for a fine-grained study of the heterogeneous recoveries from establishment closure. However, the evidence I marshal using closures cannot by itself speak to the broader relationship between labor markets and learning rates. Absent the structure provided by a model, I would not be able to control for the human capital acquisition which lies at the core of my hypothesis. In particular, reduced form analysis cannot distinguish between wage growth driven by a) worker ability, b) occupational characteristics (e.g. steep vs. shallow learning curves), and c) worker-occupation matches. Further, my estimates of the mismatch costs would be unacceptably biased by strong assumptions concerning unobserved worker skills. A detailed understanding of mismatch allows for a full picture of the costs and benefits associated with the worker's occupational alternatives. Also lost in a reduced-form context is the forward-looking nature of occupational choice. This is crucial to describe the qualitative patterns of occupational choice, which I will study across worker types.

To address these shortcomings, I construct a dynamic Roy model of occupational choice where workers accumulate human capital through learning-by-doing. The fundamental level of heterogeneity across workers is the rate at which they learn-by-doing. In choosing an occupation, a worker is also making a choice about her future comparative advantage. As she works, her multidimensional skill set gradually transforms into the skill set demanded by her chosen occupation. These skills follow the worker through her career, and can be applied to other occupations in a manner reminiscent of Gathmann and Schönberg (2010). In contrast to models featuring returns to occupational or firm tenure, my model's projection of occupations onto skill space allows for the accumulation of broadly applicable human capital while avoiding an untenable dimensionality problem. Notably, it delivers occupational choice that is potentially unrelated to observed comparative advantage. Rather, workers who learn quickly choose occupations that offer greater opportunities for skill accumulation. Workers who do not learn as quickly are more closely tied to their present-day comparative advantage. As my estimates will show, the quickest learners adjust to their job's cognitive skill set 65% faster than the rate of the slowest learners.

Estimating the model and its learning rates poses a particular challenge. Each worker has an unobserved, unique, and dynamic skill set which evolves according to occupational choice. Further, I do not observe individual-specific initial skills. Despite the dual challenge of unobserved latent variables *and* parameter values, I am able to estimate the full path of each worker's skills. I do so using a non-linear Unscented Kalman Filter with additive errors, popularized by Wan and van der Merwe (2002). In a methodological contribution, I propose a simple additional step to this filter, resulting in improved accuracy for the posterior covariance estimates. In the simple case of a linear model with Gaussian errors, my modification delivers a precise approximation of the standard Kalman Filter.

The estimated skill paths are commensurate with model predictions: fast learners start their careers by selecting occupations for which they have a high mismatch, and pay a price for this in early-career earnings. In the first year of their respective careers, fast learners on average choose occupations with a skill mismatch 10% higher than those chosen by slower learners. This is in spite of the fact that they start with a set of skills that is, on average, 5% higher than slower learners, and with cognitive skills that are 17% higher. Average first-year salaries are nearly equal between the quickest and slowest types, and significant separation of earnings does not arise until the fifth year of their respective careers.

For faster learners, the benefit of this tradeoff is increased skills in later periods. By choosing more difficult occupations in their early career, these workers acquire more skills than they would otherwise. After 20 years of experience, the mismatch pattern flips, and fast learners choose occupations with an average of 30% less mismatch than slower learners.

This difference, in combination with the movement up the occupational skill ladder, leads to average salaries that are 57% higher in mid-career.

I.A. Literature

This paper builds on and speaks to a variety of literatures. First is a burgeoning literature on multidimensional human capital. Similar to this paper, Lise and Postel-Vinay (2019) studies the acquisition of multidimensional skills, where workers learn skills to match the demands of their firm. They are especially interested in the learning rates unique to three skills, finding that manual skills adjust quickly but enjoy moderate wage returns, cognitive skills adjust slowly and enjoy high wage returns, and interpersonal skills are relatively time invariant. This paper studies the same three skills, constructed in a similar manner, but estimates heterogeneous adjustment rates across not just skill types but also worker types. Further, I directly estimate the evolution of skills from individual-level data, as opposed to identifying an initial skill set, followed by a non-stochastic evolution of skills during the worker's career.

Yamaguchi (2012) is also closely related to this paper, and is nearly unique in that it also uses a Kalman Filter to directly estimate the full path of worker skills. The study likewise allows for some heterogeneity of skill acquisition rates, however it suggests that the heterogeneity is independent of occupational choice. Instead, skills have a 'drift' that varies with worker type. My paper builds on Yamaguchi (2012) by allowing for flexibility in rates of on-the-job learning, and methodologically extends it by allowing for an arbitrary non-linear specification of both the learning and wage functions.

Other important and related contributions to the study of multidimensional skill space include Autor, Levy, and Murnane (2003), Lindenlaub (2017), Guvenen, Kuruscu, Tanaka, and Wiczer (2019), and Traiberman (2019). Traiberman (2019) shares with this paper not only a theoretical approach, but also an emphasis. It analyzes multidimensional skills within a dynamic discrete choice model, and focuses on margins of adjustment for workers who suffer adverse labor market shocks. While it allows for differing returns to experience across occupations, worker-level heterogeneity is confined to a fixed vector of comparative advantage and differing occupational switch costs. Alongside Dix-Carneiro (2014), Traiberman (2019) also highlights the heterogeneous labor market costs of import competition shocks within a dynamic Roy model. In a reduced form context, papers such as Autor, Dorn, and Hanson (2013), Autor, Dorn, Hanson, and Song (2014), and Hummels, Jørgensen, Munch, and Xiang (2014) indicate the extent to which losses from trade shocks differ across workers.

Finally, this paper addresses the literature on displaced worker & establishment closure literature introduced above. I estimate losses that are on average smaller than those reported in other papers and shed light on the variables that can explain differences in outcomes, joining papers such as Seim (2019) and Davis and von Wachter (2012) in doing so.

I.B. Discussion

Public policy often treats worker displacement as a retraining problem. Particularly when macroeconomic forces like recessions and import competition lead to job loss, a default response is to encourage workers to resume education in one form or another. Take the example of Trade Adjustment Assistance (TAA), the United States' primary government program directed towards workers who lose their jobs due to import competition or offshoring. According to its 2016 annual report, "The TAA Program seeks to provide adversely affected workers with opportunities to obtain the skills, credentials, resources, and support necessary to *(re)build skills for future jobs*" (emphasis in the original).

The take-up rate for the TAA's training support, however, is remarkably low, despite considerable recent attention to international trade as a cause of worker displacement. In the TAA's 2016 fiscal year, 126,844 new workers became eligible for TAA benefits, meaning the agency certified that their job losses were due to foreign competition or offshoring.[2] 24,433 took up training assistance that year, or less than 20% of the year's eligible workers. Another

[2]Certifications are done at the establishment-level, not the worker-level.

way of looking at this is that roughly 1 in 13,000 Americans received training through TAA in 2016. The same year, the unemployment and wage costs of trade rose to the forefront of the American Presidential election, and both major party candidates supported increased trade restrictions.

Emphasizing this response is not to minimize the scale of the job losses, nor the pain that comes with it.[3] Rather, it is to say that current policy approaches are not appealing to their target audience. This paper's model provides a framework for thinking about why these take-up rates are so consistently low. For some workers, retraining is unlikely to have a substantial effect. Those workers may not have faith in their own ability to retrain, especially as they age. Some of those who do enter retraining will have difficulty adapting their skill set, consequently suffering persistent wage losses or dropping out of the labor force entirely.

The paper continues as follows: Section II introduces the data used and discusses labor market institutions in Denmark. Section III analyzes heterogeneity in recovery from job loss with a differences-in-differences framework. Section IV introduces the skills-learning process and the dynamic Roy model. Section V estimates the learning process using a modified Unscented Kalman Filter within an Expectation-Maximization algorithm. Section VI concludes.

[3]For an excellent and highly readable perspective on personal experiences with job losses and plant closures in America, I suggest *Janesville* by Amy Goldstein.

II. Data and Institutional Background

The data used in this paper is provided by Statistics Denmark. The primary register is IDA, a matched employer-employee dataset covering the universe of the Danish working age population in the years 1991-2016.[4] Each individual is assigned a unique ID, allowing them to be tracked through their careers. At the individual level, I observe demographic characteristics, educational background, and yearly labor income. For a subsample of the population, I observe grades from secondary education. Finally, I observe the duration and type of governmental support for each worker.

Each individual is linked to the job or jobs that he or she held through each year. For many jobs, I observe an estimate of wages, total yearly earnings from that job, full/part time status, and an 8-digit occupational classifier. I also observe start and end dates for jobs through 2014, though these data are imperfect, particularly later in the sample. Data coverage is not uniform across jobs, and is superior for the primary job that a worker holds in November of any given year. Jobs are linked to an establishment and firm, providing further information such as industry, geographic location of the job, average wages, and number of workers at the establishment. Crucially, I also observe when an establishment permanently shutters, which will form the backbone of the event study in section III.

Labor income is central to this paper, and three potential measures of earnings merit consideration. The first is wage, a measure developed by Statistics Denmark. This statistic has drawbacks, mostly due to the fact that hours are not observed and Statistics Denmark uses a coarse rule to calculate wages. Lund and Vejlin (2016) puts considerable work into improving the measure, however their modifications cannot completely overcome the challenge of measuring wages. Moreover, wages are observed in only some jobs, with significantly less coverage than the other two potential measures. The second measure is monthly salary,

[4] See Bobbio and Bunzel (2018) for a comprehensive review of these data. IDA is the Danish acronym for *The Integrated Database for Labour Market Research*.

constructed by dividing yearly earnings at a job by the number of months spent on that job. While monthly wage is a potentially coarse measure due to an implicit assumption that all jobs share common hours, it has the advantage of excellent coverage across all jobs. The last potential measure is yearly labor income. This measure sidesteps the challenge of identifying which job is most reflective of a worker's true skill in a year, which is useful in some contexts. While yearly labor income has near-perfect coverage and high accuracy, it does not offer any correction for time not worked. For a paper concerned about early career wage growth, this is an issue. For example, initial wage growth will be overstated when comparing half a year's salary to a whole year's salary. On the other hand, when I want to measure high-level labor performance, then yearly earnings is an ideal metric.

Of those three measures, I primarily use monthly salary and yearly labor earnings. I rely on yearly labor earnings in section III, where I am interested in recoveries at the worker level, as opposed to the characteristics of any one job. In section V, I instead rely on the monthly salary measure in the structural estimation of learning rates. This nicely suits the needs of assessing renumeration due to human capital, due to the combined strong coverage of the measure and its correction for partial-year employment. Finally, all currency levels in this paper are stated in 2010 Danish Krone unless otherwise noted.

Turning to occupations, occupational codes are observed as DISCO-88 and later DISCO-08 codes. These are Danish adaptations of the European ISCO codes, which are first reliably observed in 1991. These are not recorded for every job, nor for every year within the same job, although coverage is very good. Thanks to Taber and Vejlin (2019), I am able link these occupational codes to the American Standard Occupational Classification (SOC) codes at the 4-digit level. In turn, this lets me link each job to a set of skills as measured by the now-famous O*NET database. I also improve the coverage of occupational codes by imputing missing occupations within jobs: if a job-year observation has no occupation recorded, but that same job has a known occupation in another year, I pull the code from the nearest year.

O*NET measures skills, tasks, abilities, and much more for SOC occupations. It is an initiative of the U.S. Department of Labor, primarily intended to help American workers in their job search. For each occupation, workers, occupational experts, or both assign scores to various measures that help codify the nature of a given occupation. For instance, "Persuasion" is an O*NET skill, and has both a Level (measured from 0-7) reflecting the difficulty of an occupation with regards to persuading others, and an importance (measured from 0-5), measuring how critical persuasion is to completing the job.

Similar to Traiberman (2019) and Lise and Postel-Vinay (2019), I use Principal Component Analysis (PCA) to reduce the considerable dimension of the O*NET database. I use the first three principal components to match the three skills in the model: cognitive, manual, and interpersonal. These three skills are matched to the first three principal components by rotating the linear space to match orthogonality conditions. I follow Lise and Postel-Vinay (2019) in my method, and I impose the following conditions:

1. The skill "Complex Problem Solving" is a strictly cognitive task.
2. The activity "Performing General Physical Activities" is a strictly manual task.
3. The skill "Social Perceptiveness" is a strictly interpersonal task.

In terms of degrees achieved and school attendance, coverage of education is excellent. Data on individual grades at the secondary level is also available, though they are limited to the subsample of the population who completed academic secondary education. For the period under consideration, the only grades observed are akin to a general GPA. I normalize these within cohorts, assigning each individual a within-cohort percentile. This will be a key source of sample selection and division in the remainder of the paper.

Finally, I observe establishment closures directly in the data. A closing establishment is marked in the year prior. I confirm that closures are not simply a shift in locations by verifying that no more than 30% of workers in each closing establishment moved to a common firm in the following year.

II.A. Institutional Background

"Flexicurity" is a popular way of describing the Danish labor market. The term reflects a system where companies can hire and fire workers without significant legal challenges or severance costs. However, the Danish government offers generous support for the unemployed, both in terms of benefits and in terms of assistance in returning to employment. This policy set generates substantial fluidity in the Danish labor market. Occupational and job switching are common, particularly in the early stages of a career. However, this should not be taken to mean that the Danish labor market is just like the U.S. labor market.

In particular, unions retain a strong presence in Denmark. Unions cover roughly 70% of the working population, and unionization typically occurs at the occupational level. While Denmark has no national minimum wage, collective bargaining agreements (CBAs) typically establish a wage floor, and these have very wide coverage. Having said this, the CBAs do not impose a rigid wage schedule on firms, as they do in some countries. Boeri and Calmfors (2001) constructs a central bargaining index from 0 (no central bargaining) to 1, and reports that Denmark has fallen significantly in their measure, from 0.64 in the 1970s to 0.34 in the 1990s. While wage-tenure contracts still exist, individual-level wage negotiation on a yearly basis is common in the private sector.

Secondary education in Denmark is not mandatory, and vocational training is a common alternative to schooling intended as preparation for higher education. During the 1980s—the time when many of our observations went to school—vocational training was the norm, and the various forms of academic secondary education were not as well attended. Once a student chooses a type of secondary education, the nature of the schooling is fairly uniform across the country, and tests for academic education are typically drafted and graded by a common central authority.

Finally, government support for higher education is generous. All schooling costs are covered and students receive additional direct financial support from the government while they remain in school. As a result, Danes often finish formal studies later than is observed in other countries. I account for this fact by measuring wages and salaries from "career start," defined by the first year when an individual is employed full-time for two consecutive years, uninterrupted by any formal schooling.

III. Recovery from Establishment Closures

III.A. Establishment Closures as a Random Event

To study responses to adverse labor market shocks, I focus on workers who lose their primary jobs due to establishment closure. In doing so, I follow a now-extensive literature dating back to Jacobson, LaLonde, and Sullivan (1993), who studied Pennsylvanian workers subject to layoffs in the 1980s. This literature is tied together by a common empirical approach, in the sense that layoffs—in this case, establishment closures—are treated as a random event at the worker-level. If true, then the layoffs lend themselves to the study of worker-level effects of job displacement in a differences-in-differences framework.

Establishment closures are particularly useful to estimate average losses and responses to job separations. Though they cannot address every plausible source of endogeneity, establishment closures address many of the selection concerns endemic to this problem. Theoretically, job separations initiated by the firm affect workers from the lower end of the productivity distribution. This remains true for layoffs, as firms prefer to fire lower-performing employees before higher-performing ones. On the other hand, job separations initiated by the worker could skew towards the upper end of the productivity distribution. Workers choosing to leave their job are more likely to enjoy a superior outside option, and they're more likely to be moving up the job ladder. Additionally, an extensive literature argues that workers initiate job separations in order to find better matches (e.g. Topel and Ward (1992), Jovanovic (1979), Pavan (2011)). Due to this dual selection problem, studying non-random job separations is unlikely to shed light on responses to adverse labor market events.

In contrast, establishment closures have no selection at the establishment level. This is not to say that this strategy avoids such issues altogether. Primary among remaining concerns include selection *into* establishments—it's possible that certain types of workers tend to find employment at establishments that are more likely to shutter. Additionally, I might be concerned about predictive selection: if workers can forecast establishment closure, then those with superior outside options might choose to leave prior to closure. I address these two potential issues in two separate ways. First, I address establishment-level selection by selecting a control group that reflects the treatment group on a wide variety of worker-level observables. Second, I address predictive selection by replicating the main differences-in-differences exercise using a sample selected years before the actual closures.

I construct the control by individual-level matching, as opposed to establishment-level matching. As a first pass, it might make sense to match at the level of the shock: this would directly address concerns about selection at the establishment-level. Instead, I match at the individual level for two reasons. First, my primary goal is to analyze individual heterogeneity in responses. Matching on individual-level heterogeneity leads to a natural definition of the relevant and distinct control groups in this context. Second, because I do not observe firm financials or anything else reasonably predictive of establishment closure, propensity score matching isn't likely to succeed in identifying establishments with roughly equal probabilities of closing.

III.B. Sample Selection

I construct the sample to select workers who lose a long run, stable job due to establishment closure. Closures from 1995-2010 are considered, meaning I can potentially observe work histories running for 4 years prior to and 6 years after each closure. For every worker affected by a plant closure, I impose the following criteria for sample inclusion:

1. The worker's employment at this establishment extends over three calendar years leading up to the closure.
2. The employment is full-time, and is the worker's primary job leading up to the closure.[5]
3. The worker is between 25 and 55 years old at the time of closure.
4. This is the first time the worker has lost a job due to plant closure.
5. The establishment has a minimum of 5 full-time employees in the year prior to closure.

[5]As measured by a worker's "November" job. Danish register data records the new year in November, so this month is when the data is most accurate.

6. The establishment is not connected to any other workplace in the following year, as measured by workers moving collectively to another establishment.
7. The worker is not receiving government support for education in the year prior to the closure.
8. The worker does not own the firm or establishment.

73,333 lost jobs meet these criteria, stemming from 9,128 establishment closures. There are far more closures and jobs lost due to closures over the 16-year period. By restricting the sample with these criteria, I can focus on my core question—how do individuals react to a significant adverse shock?

I select the control based on the coarsened exact matching (CEM) procedure, popularized in economics literature by Azoulay, Graff-Zivin, and Wang (2010) as well as Jäger and Heining (2019). This choice departs from the layoff literature's standard control group, which has typically been a subset of the population which meets the same sample eligibility requirements as the treatment group, aside from the closure (or layoff).[6] However, the standard choice of control is particularly inappropriate for my purposes. My goal is to analyze heterogeneity in recoveries, and for each treatment observation I ask what has been lost relative to the counterfactual world where the worker was not hit with this shock.

With that in mind, I use CEM to construct a control group that matches the treatment on a large set of observables in the year prior to closure. Each individual in the control group is matched to a single worker selected from the population of Danes who never experience job loss due to plant closure. All potential matches must be employed for at least 90% of the year in which they are matched. I match on the following set of observable characteristics:

1. Age
2. Sex
3. Highest education completed
4. 2-digit occupational code

[6] Or, when using register data, the population who meet this criteria. See, e.g. Seim (2019). Another potential method is to use workers who later join the treatment group as a contemporaneous control. This would be particularly useful if I had stronger concerns about selection into establishments that have a higher propensity to close. However, this concern is not as salient as the concern for heterogeneity analysis.

5. Tenure at current job (divided into low, medium, and high)
6. Whether or not they live in the Copenhagen area
7. Yearly earnings

Typically, this process returns multiple potential matches. I select the potential match with the closest yearly salary (in the year prior to closure) as a control group observation. In the few cases where this procedure returns no matches, I relax the age-matching requirement to allow for matches who are 2 years older or younger than the treatment worker. This leaves us with a match rate exceeding 99%. The few workers who are unmatched are excluded from the sample. Descriptive statistics for treatment and control groups are given in Table I.

This method of choosing a control group pays off when I turn to uncovering heterogeneity in recoveries. For any dimension of heterogeneity, I can select the appropriate treatment sample, and I have a control of peers at hand. This stands in contrast to a common control group, which could confuse unrelated phenomena for "recovery." For example, suppose the earnings of one worker type grew faster than the population average, while another worker type had slower than average growth. In the years post-closure, I would confuse those different growth rates for a recovery to the population baseline (and lack of recovery), though no recovery is realized: this was simply a secular trend. Instead, all coefficient estimates below should be interpreted as an effect *relative to peers*, due to the CEM procedure and its construction of a peer group.

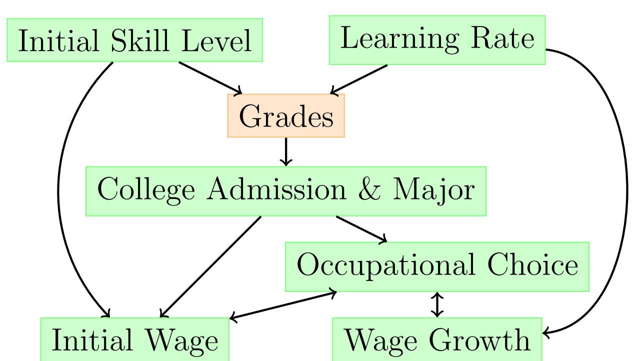


Figure I
Causes and Confounders

TABLE I

Descriptive Statistics: Treatment and Control Group Balance Tables in the Year Prior to Closure

Variables	High Educ. x High Grade		High Educ. x Low Grade		Low Educ. x High Grade		Low Educ. x Low Grade		Entire Sample	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
Matched										
Age	36.86	36.89	36.28	36.33	37.31	37.38	35.38	35.38	40.29	40.29
	(6.41)	(6.40)	(6.59)	(6.59)	(7.16)	(7.16)	(6.92)	(6.92)	(8.42)	(8.44)
Male	0.52	0.52	0.51	0.51	0.59	0.59	0.49	0.49	0.62	0.62
	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)	(0.49)	(0.50)	(0.50)	(0.49)	(0.49)
Copenhagen	0.27	0.27	0.15	0.15	0.21	0.21	0.17	0.17	0.12	0.12
	(0.44)	(0.44)	(0.35)	(0.35)	(0.41)	(0.41)	(0.37)	(0.37)	(0.32)	(0.32)
Earnings/2010$	85,361	82,739	70,905	69,764	73,917	71,489	63,462	62,088	61,930	61,087
	(62,808)	(50,568)	(44,131)	(41,882)	(46,253)	(39,964)	(41,347)	(33,533)	(36,921)	(32,526)
High-Skill Occ.	0.64	0.64	0.35	0.35	0.19	0.19	0.11	0.11	0.17	0.17
	(0.48)	(0.48)	(0.48)	(0.48)	(0.39)	(0.39)	(0.31)	(0.31)	(0.37)	(0.37)
Medium-Skill Occ.	0.32	0.32	0.56	0.56	0.65	0.65	0.67	0.67	0.41	0.41
	(0.47)	(0.47)	(0.50)	(0.50)	(0.48)	(0.48)	(0.47)	(0.47)	(0.49)	(0.49)
Low-Skill Occ.	0.01	0.01	0.04	0.04	0.07	0.07	0.11	0.11	0.29	0.29
	(0.11)	(0.11)	(0.20)	(0.20)	(0.25)	(0.25)	(0.31)	(0.31)	(0.45)	(0.45)
Tenure at Firm	5.04	4.92	4.96	4.98	5.19	5.16	5.17	5.12	5.59	5.65
	(2.87)	(2.99)	(2.99)	(3.14)	(2.92)	(3.28)	(3.02)	(3.25)	(3.42)	(3.61)
Unmatched										
Experience	10.54	10.62	11.34	11.71	12.70	13.58	12.39	12.55	14.77	14.83
	(6.05)	(6.20)	(6.11)	(6.48)	(7.15)	(7.27)	(6.67)	(6.83)	(7.00)	(7.02)
Big City	0.49	0.42	0.33	0.29	0.37	0.30	0.32	0.29	0.28	0.24
	(0.50)	(0.49)	(0.47)	(0.46)	(0.48)	(0.46)	(0.47)	(0.45)	(0.45)	(0.43)
Gov't Employee	0.26	0.30	0.31	0.35	0.27	0.22	0.22	0.24	0.18	0.23
	(0.44)	(0.46)	(0.46)	(0.48)	(0.44)	(0.42)	(0.41)	(0.43)	(0.38)	(0.42)
Manufacturing	0.10	0.12	0.12	0.13	0.09	0.08	0.11	0.11	0.18	0.19
	(0.30)	(0.33)	(0.32)	(0.34)	(0.29)	(0.28)	(0.31)	(0.31)	(0.39)	(0.39)
Observations	3,220	3,220	2,303	2,303	707	707	2,666	2,666	73,333	73,333

Standard deviations in parentheses. Grades are not observed for the entire sample, and therefore the number of observations in the first four groups do not sum to the sample size.

III.C. Heterogeneity Analysis

I hypothesize that fast-learning workers recover more quickly after an adverse shock than slow-learning ones. This is due to their ability to rapidly adjust their skill set and move to occupations in higher demand. There is, however, a complication to this hypothesis: faster-learning workers can develop a strong comparative advantage more quickly, providing a disincentive to leaving an occupation. In section IV of this paper, I will discuss this complication and the model's implications in detail. Here, for the sake of these empirical exercises, I predict that, in the wake of job displacement, higher-learning types will fare better relative to their peers than lower types.

To examine the variation in recoveries, I study differences-in-differences at the individual level. For each matched treatment-control pair, I define a differences-in-differences statistic as

$$Y_i = (Y^i_{Treat,Post} - Y^i_{Treat,Pre}) - (Y^i_{Control,Post} - Y^i_{Control,Pre}). \tag{1}$$

In this section, Y^i is the log of yearly earnings. I define this for a variety of pre- and post-periods in order to study short-run and long-run recoveries after an establishment closure. This implies that Y_i is smaller in absolute value for fast learners than for slow learners.

To focus on differences based on learning rates, I split workers into four subgroups. This split occurs along two dimensions: high vs. low education and high vs. low grades, with both serving as imperfect proxies for learning speeds. High education means a worker has completed some level of postsecondary education. High grades are defined as being in the top tercile of a secondary schooling cohort. Low grades imply the bottom tercile. The middle tercile is omitted from this categorization. Recall that I only observe grades for a subset of the sample. When using education and grades as proxies for learning speeds, we should be concerned about potential confounders, especially when so many features of the labor market correlate with educational outcomes. For example, we might be worried that people with higher grades and education sort into occupations with thicker labor markets.

A thicker market would make it easier to find a closely related job post-job loss, thus easing the path to recovery. It's also possible that highly educated Danes may hold a preference for the amenities available in Copenhagen or Denmark's mid-sized cities. Larger cities and towns might also feature thicker labor markets than small towns, and thus ease recovery.

To parse these potential explanatory variables, I regress a battery of fixed effects on the differences-in-differences statistic from (1). All regressors are defined in the year prior to establishment closure. The specification is

$$\begin{aligned} Y_i = &(\text{Educ x Grade})_i + (\text{Educ})_i + (\text{Age at Closure})_i + (\text{Occupation at Closure})_i + \\ &(\text{Geography at Closure})_i + (\text{Occ x Geography at Closure})_i + \upsilon_i \end{aligned} \tag{2}$$

Table II reports the results of this regression with a long-run difference.[7] Here, the pre-period is one year prior to the closure and the post-period is five years after. Notably, only the (High Education x High Grade) group is predictive of differences in post-closure recovery. I emphasize again that this is relative to their peers, not the general public. This group is recovering to their peers—in the differences-in-differences sense—with a gap in yearly earnings that is roughly 6.5 to 7.5 percentage points higher than baseline.

Am I focusing on a relevant margin of heterogeneity? The results are reassuring, and suggest that my chosen proxy for learning rates predicts recovery relative to peers. Moreover, this predictive capacity remains after controlling a battery of potential confounding variables. In the next subsection, I will therefore use this margin to break the sample apart in the more familiar differences-in-differences specification.

Do I observe the same differences in the short run? No. Table III documents the same regressions as above, but with the post-period changed to 1 year post-closure. Notably, no indicators are predictive of differences in outcomes. This result will stand up in the next

[7] These regressions are estimated on closures from 1995-2006. The expansion of data allowing for the extension of closures until 2010 is not yet prepared to be included in this exercise. For that reason, the sample sizes are lower in this regression than in later regressions.

TABLE II

Long Run Heterogeneity in Recoveries—from 1 Year Pre- to 5 Years Post-Closure

	Individual-pairing "DiD", log(Yearly Earnings)			
	(1)	(2)	(3)	(4)
$\mathbb{I}$\{HighEduc & HighGrade\}	0.0750***	0.0758**	0.0698**	0.0659**
	(0.0289)	(0.0319)	(0.0318)	(0.0314)
$\mathbb{I}$\{HighEduc & LowGrade\}	0.0115	0.0093	0.0049	0.0066
	(0.0298)	(0.0316)	(0.0317)	(0.0317)
$\mathbb{I}$\{LowEduc & HighGrade\}	0.0217	0.0201	0.0401	0.0306
	(0.0538)	(0.0557)	(0.0556)	(0.0554)
$\mathbb{I}$\{LowEduc & LowGrade\}	0.0128	0.0010	0.0116	0.0109
	(0.0272)	(0.0296)	(0.0298)	(0.0299)
Constant	-0.0646***	-0.0777**	0.00286	0.00888
	(0.00511)	(0.0394)	(0.0700)	(0.0765)
Education FE	N	Y	Y	Y
Age at Closure FE	N	Y	Y	Y
Occupation at Closure FE	N	N	Y	Y
Geography at Closure FE	N	N	Y	Y
Occupation x Geography FE	N	N	N	Y
Observations	34574	34322	34322	34322
R^2	0.000	0.003	0.007	0.012
Adjusted R^2	0.000	0.002	0.005	0.007

Standard errors clustered at the closure-firm level. Occupational FEs defined at the 2-digit level from the Danish ISCO classification system. Geographic fixed effects are limited to one for Copenhagen and a second covering all cities with populations greater than 100,000 in 2006 (Copenhagen, Aarhus, Odense, Aalborg). "High Grade" defined as the top tercile of observed secondary schooling grades. "Low grade" defined as the bottom tercile of observed secondary schooling grades. All grades are normalized within-cohort. Regressions include observations with no grade information observed. The "pre" period is one year prior to establishment closure and the "post" year is five years after the closure.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

subsection: the short-run costs of establishment closure are remarkably homogeneous.

Finally, I conduct a placebo exercise to check the matching process. Table IV reports coefficients from the same regression, but with both pre- and post- periods occurring before the closures. I report regressions from both 5 years pre-closure to 1 year pre-closure and 3 years pre- to 1 year pre. For the most part, the results are reassuringly statistically insignificant. However, as I will see for the (Low Education x High Grade) group, there's a problem with differential growth rates in the five years leading up to the closure. This suggests an issue with the matching for this specific group of workers.

TABLE III

Short Run Heterogeneity in Recoveries—from 1 Year Pre- to 1 Year Post-Closure

	Individual-pairing “DiD”, log(Yearly Earnings)	
	(1)	(2)
$\mathbb{I}$\{HighEduc & HighGrade\}	0.0412	0.0067
	(0.0251)	(0.0275)
$\mathbb{I}$\{HighEduc & LowGrade\}	0.0066	-0.0163
	(0.0276)	(0.0295)
$\mathbb{I}$\{LowEduc & HighGrade\}	-0.0255	-0.0776
	(0.0556)	(0.0575)
$\mathbb{I}$\{LowEduc & LowGrade\}	0.0655*	0.0247
	(0.0255)	(0.0274)
Constant	-0.1286***	0.0620
	(0.0045)	(0.0626)
Education FE	N	Y
Age at Closure FE	N	Y
Occupation at Closure FE	N	Y
Geography at Closure FE	N	Y
Occupation x Geography FE	N	Y
Observations	37826	37522
R^2	0.000	0.014
Adjusted R^2	0.000	0.009

Standard errors clustered at the closure-firm level. Occupational FEs defined at the 2-digit level from the Danish ISCO classification system. Geographic fixed effects are limited to one for Copenhagen and a second covering all cities with populations greater than 100,000 in 2006 (Copenhagen, Aarhus, Odense, Aalborg). “High Grade” defined as the top tercile of observed secondary schooling grades. “Low grade” defined as the bottom tercile of observed secondary schooling grades. All grades are normalized within-cohort. Regressions include observations with no grade information observed. The “pre” period is one year prior to establishment closure and the “post” year is one after the closure.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE IV

Heterogeneity in Pre-Trends? Log(Yearly Earnings) – 5 & 3 Years Pre-Closure to 1 Year Pre-Closure.

	5Y to 1Y Pre-Close		3Y to 1Y Pre-Close	
	(1)	(2)	(3)	(4)
$\mathbb{I}$\{HighEduc & HighGrade\}	0.0073	0.0113	-0.0111	-0.0281
	(0.0324)	(0.0350)	(0.0193)	(0.0211)
$\mathbb{I}$\{HighEduc & LowGrade\}	0.0068	-0.0163	-0.0141	-0.0248
	(0.0356)	(0.0375)	(0.0221)	(0.0229)
$\mathbb{I}$\{LowEduc & HighGrade\}	0.154**	0.168**	0.0103	0.0005
	(0.0568)	(0.0594)	(0.0389)	(0.0397)
$\mathbb{I}$\{LowEduc & LowGrade\}	0.0163	0.0129	-0.0135	-0.0297
	(0.0277)	(0.0294)	(0.0189)	(0.0201)
Constant	0.0010	0.0956	-0.0047	-0.0852
	(0.0043)	(0.0636)	(0.0051)	(0.0543)
Education FE	N	Y	N	Y
Age at Closure FE	N	Y	N	Y
Occupation at Closure FE	N	Y	N	Y
Geography at Closure FE	N	Y	N	Y
Occupation x Geography FE	N	Y	N	Y
Observations	36831	36544	42898	42466
R^2	0.000	0.010	0.000	0.008
Adjusted R^2	0.000	0.006	-0.000	0.005

Standard errors clustered at the closure-firm level. Occupational FEs defined at the 2-digit level from the Danish ISCO classification system. Geographic fixed effects are limited to one for Copenhagen and a second covering all cities with populations greater than 100,000 in 2006 (Copenhagen, Aarhus, Odense, Aalborg). "High Grade" defined as the top tercile of observed secondary schooling grades. "Low grade" defined as the bottom tercile of observed secondary schooling grades. All grades are normalized within-cohort. Regressions include observations with no grade information observed. The "pre" period is either five or three years prior to establishment closure and the "post" year is one year pre-closure.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

III.D. Recovery from Establishment Closure

Having shown the value of grades-education interaction in predicting recovery to peers, I now turn to the differences-in-differences regressions themselves. My main specification follows the standard in the establishment closure:

$$y_{it} = \alpha_i + \alpha_t + X'_{it}\Delta + \sum_{v=-5}^{6} \delta_v D^v_{it} + \sum_{v=-5}^{6} \delta^{Treat}_v D^v_{it} \text{ x } \mathbb{I}\{Treatment_i\} + \iota_{it} \qquad (3)$$

where

- y_{it} is some outcome. I focus on yearly earnings in the main text, with other outcomes considered in Appendix C.
- α_i, α_t are individual and year-fixed effects.
- X'_{it} is a quartic polynomial in age.
- D^v_{it} is an indicator that equals 1 if worker i is v years removed from the closure in period t. Workers in the control are assigned a "placebo closure" that is equal to the closure year of their match.

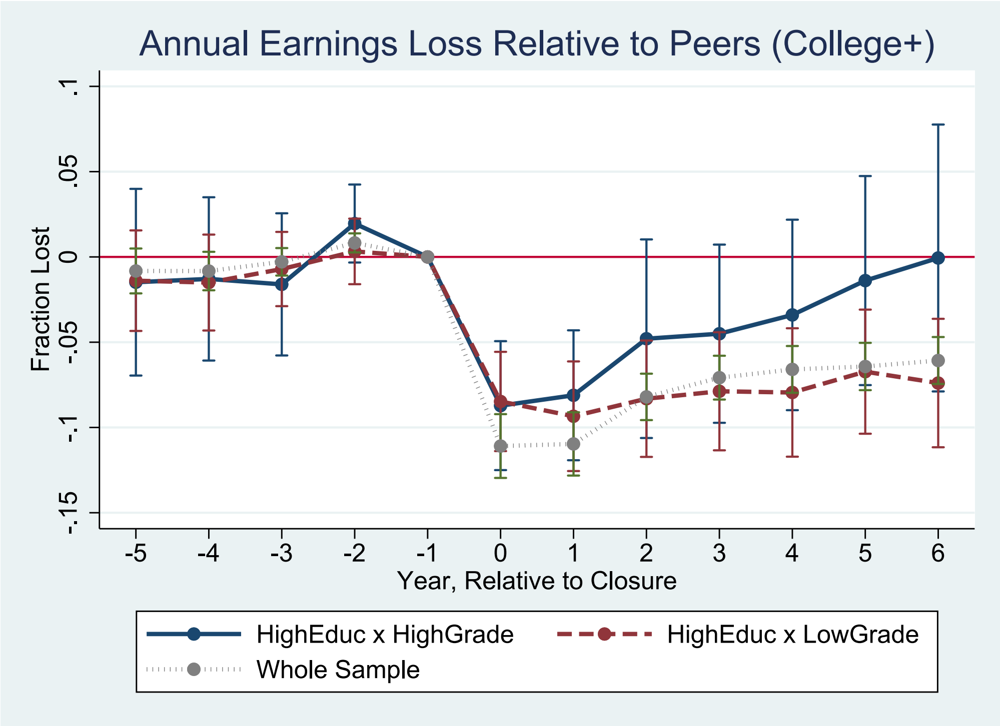
The variables of interest are δ^{Treat}_v. I interpret these as the effect of closure, v years removed from the event. Negative numbers indicate pre-closure years, and the year before the event is the omitted indicator. I estimate the main specification in levels of yearly earnings in order to be directly comparable to other papers in the job losses literature.

Critically, I estimate this regression separately for each of the four groups. This is again to establish that all reported results are relative to a peer group, not a commonly-defined control. The results are graphed in Figure II. The patterns we observe are consistent with the predicted pattern. Initial losses are remarkably homogeneous. The average earnings loss in the first year roughly lies between 30,000 and 40,000 DKK (∼5,750 to 7,750 USD in 2010). This represents an average loss of between 8-11% of the previous years' earnings. This estimate is substantially smaller than estimates in related literature.

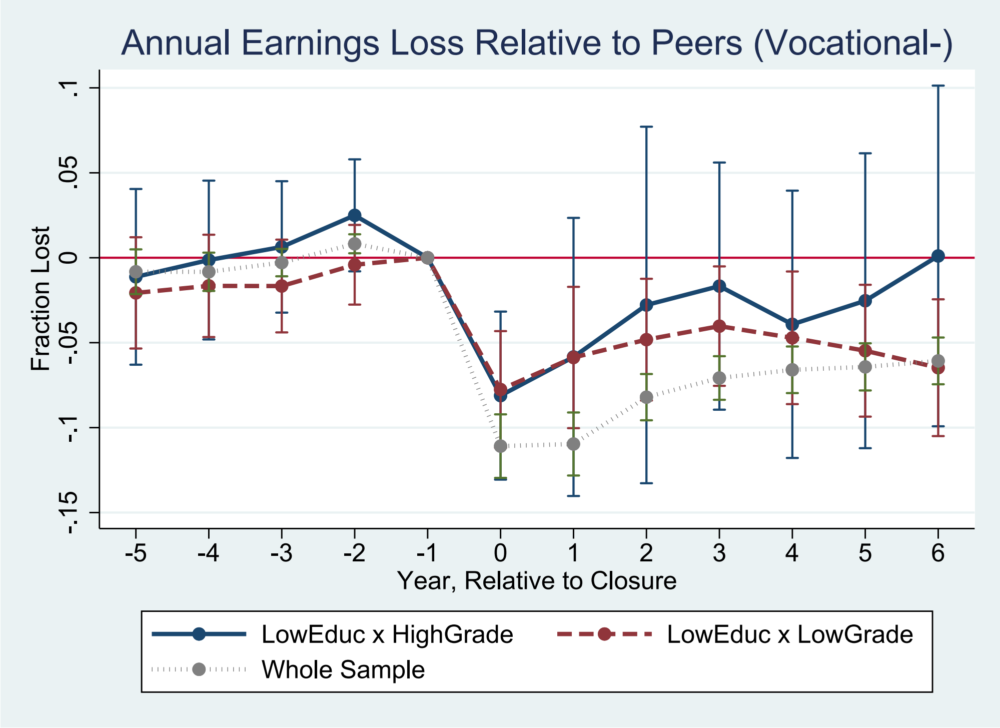
The distinct paths of recovery are remarkable. Both high-grade groups fully recover to the yearly earnings of their peers 6 years post-closure. In contrast, both low-grade groups suffer a persistently lower level of average yearly earnings. One key takeaway is that grades are a stronger predictor of recovery to peers than education: quicker learning leads to quicker recovery. The large standard errors are an unfortunate consequence of the data coverage on grades. While the whole sample includes 73,503 pairs, I only observe grades for 18% of that sample. That, combined with substantial within-group outcome heterogeneity, lead to these larger standard errors.

What's behind the divergence in recovery paths? There are two margins to investigate in these data: unemployment and occupational choice. In Figure III, we see that unemployment is higher (as always, relative to their peer group) for the low-grade groups. In contrast, Figure IV does not tell a clear story: across all skill types, the relative response is ambiguous. There does not appear to be evidence that the high-grade groups are "upskilling", at least relative to their peers who did not suffer an establishment closure. However, Figure V helps to resolve why this can be consistent with my core learning-rates hypothesis. In terms of cognitive and social skills, the high-grade types choose high-skilled occupations consistently, both before and after job loss. This pattern is predicted in our model, and established in Proposition I: high-learning types initially choose high-skill jobs to increase their option value. Therefore, we do not see substantial increases in the skills demanded by chosen occupations later in their careers. Colloquially, high-type learners jump in the deep-end at the outset of their career, so later career shifts are less common.

While these results offer support for a heterogeneous learning rates model of human capital, they cannot address by themselves the model's broader implications for the labor market. For example, if a worker has stagnant wages, is that because she isn't learning the skills required, or because she had all those skills at the outset? I next construct a model to draw further implications out of my hypothesis.

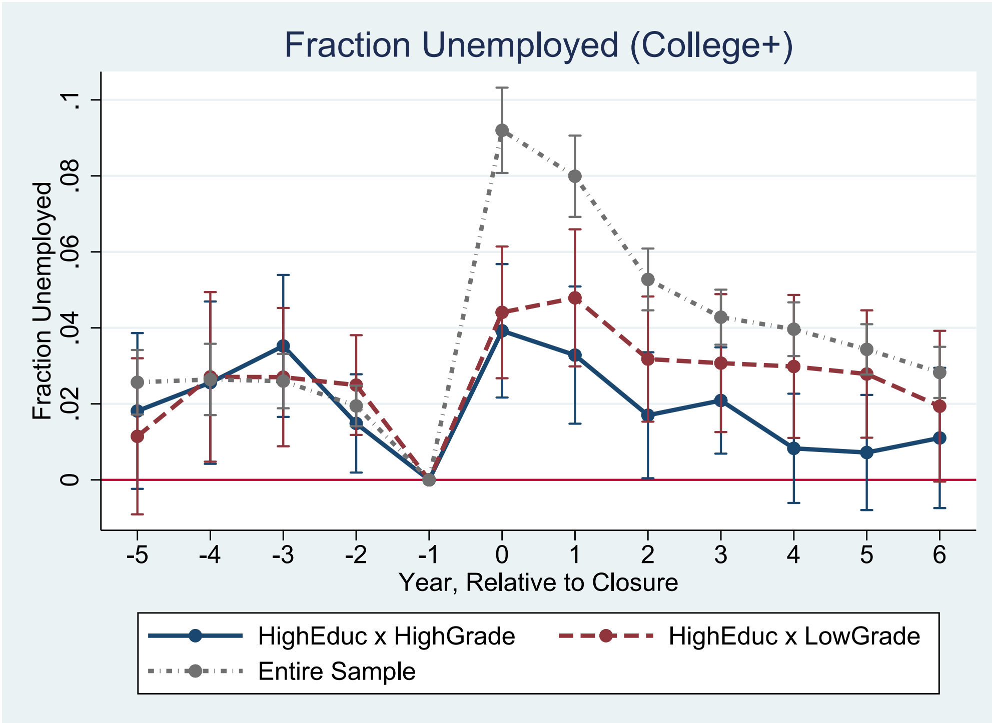


(a) Education: College and Higher

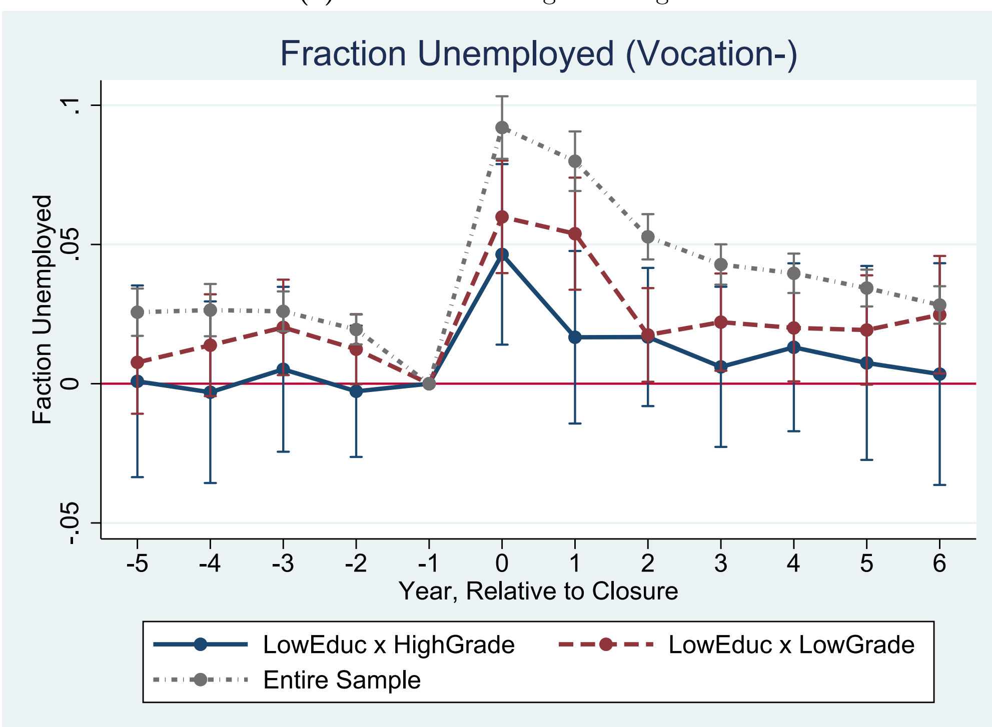


(b) Education: Vocational and Lower

Figure II
Earnings Loss after Establishment Closure



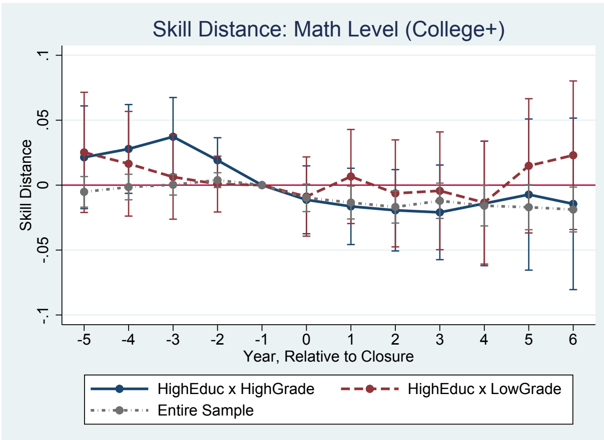
(a) Education: College and Higher



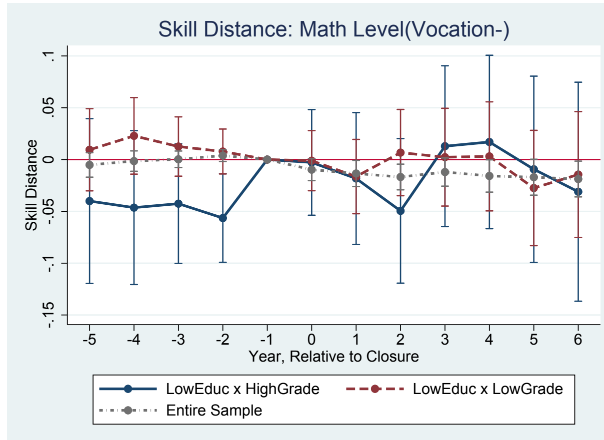
(b) Education: Vocational and Lower

Figure III

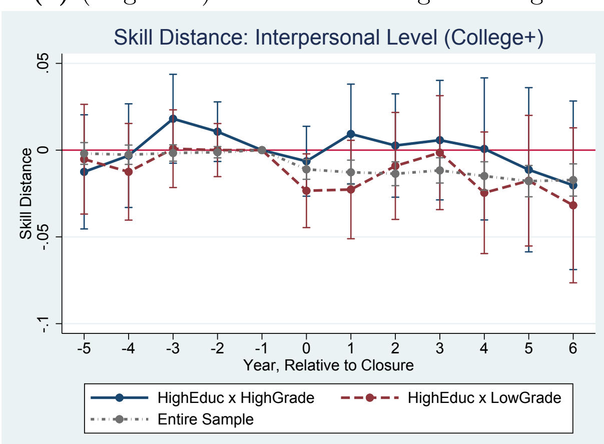
Unemployment after Establishment Closure



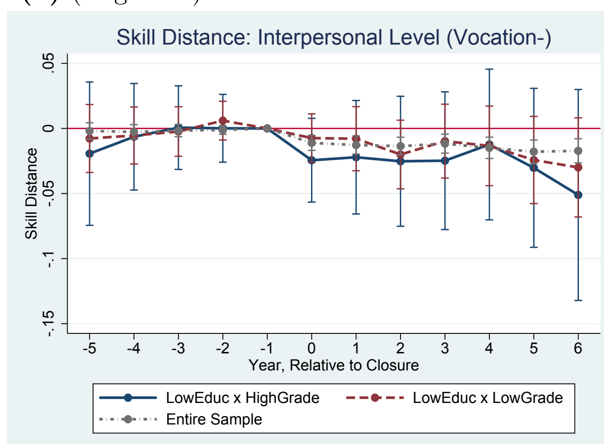
(a) (Cognitive) Education: College and Higher



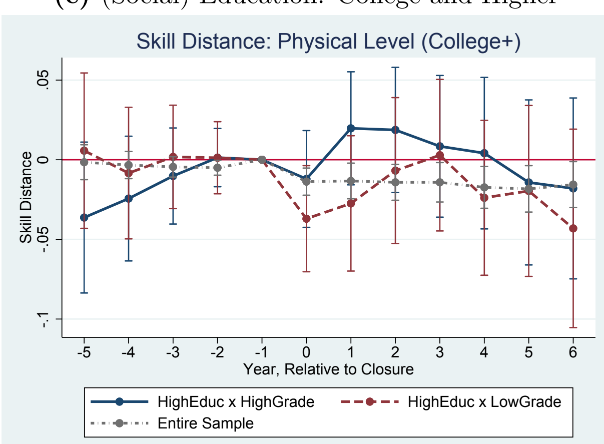
(b) (Cognitive) Education: Vocational and Lower



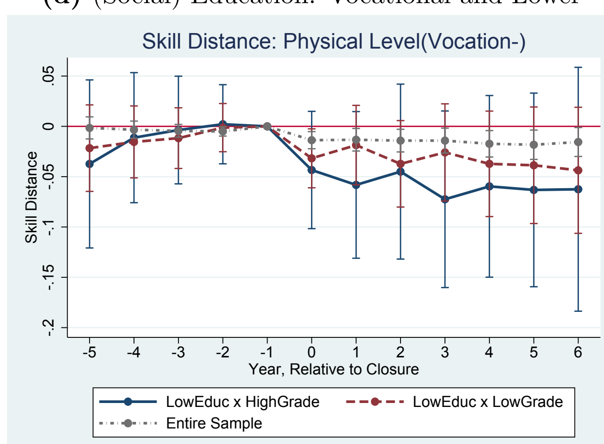
(c) (Social) Education: College and Higher



(d) (Social) Education: Vocational and Lower

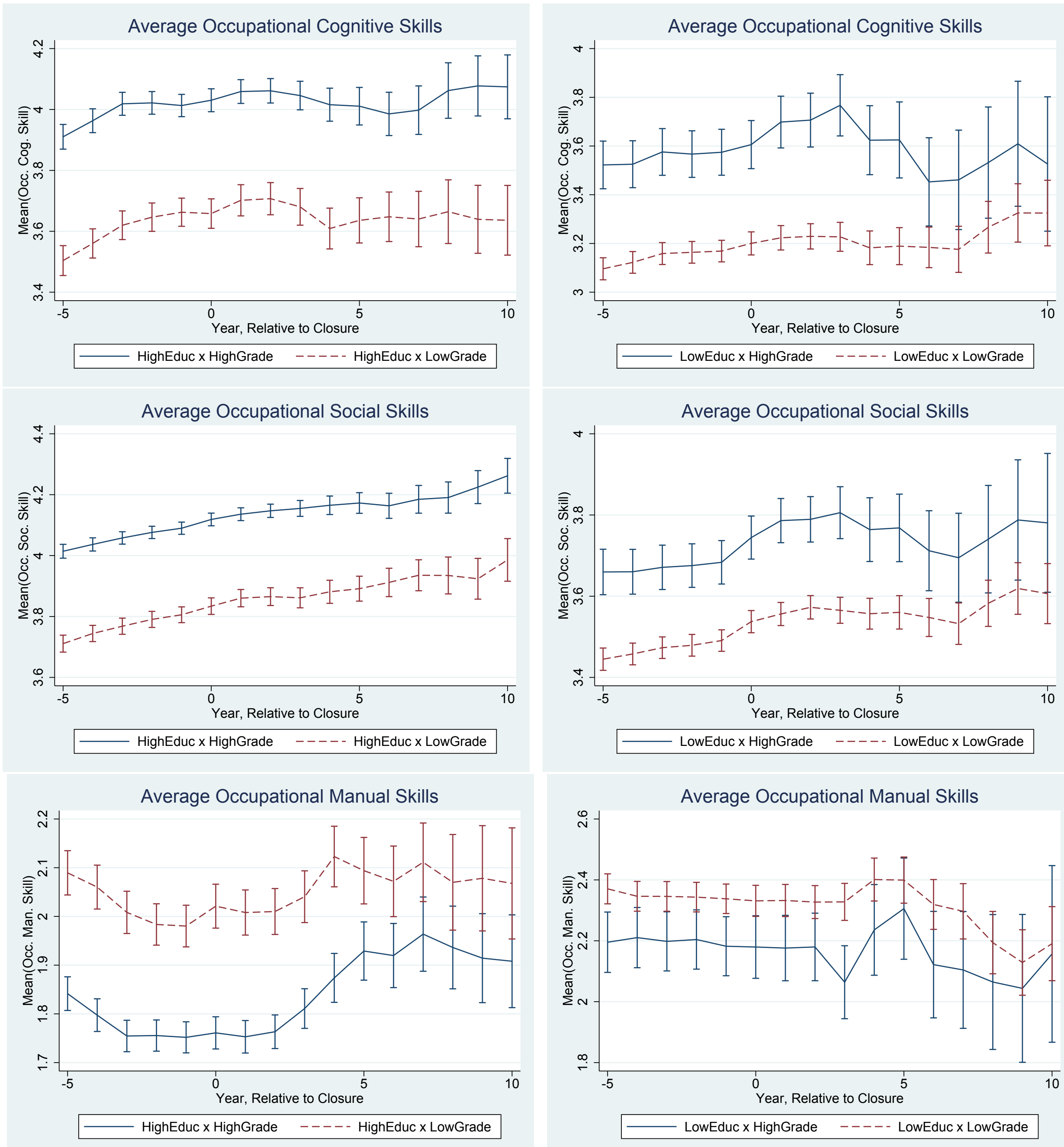


(e) (Manual) Education: College and Higher



(f) (Manual) Education: Vocational and Lower

Figure IV
Skill Choice after Establishment Closure



(a) Education: College and Higher

(b) Education: Vocational and Lower

Figure V

Average Skills in Chosen Occupations (in Levels) after Establishment Closure

IV. Theory

In this section, I develop a model of human capital accumulation and production. Having detailed those processes, I turn to the worker's problem and describe equilibrium. Finally, I use the worker's problem to analytically derive occupational switching rates, and state the propositions which will be used to benchmark the model.

IV.A. Human Capital Accumulation

Workers acquire human capital through learning-by-doing. When employed, the worker's skill set evolves to match that of their chosen occupation. Each year, workers close a fixed percentage of the distance between their own skills and the occupation's. Critically, different worker types close these gaps at different speeds. Fast learners converge to occupational skills more quickly than slow learners.

Each worker, $i \in 1, 2, ..., N$ at time $t \in 1, 2, ..., T_i$ has a human capital vector, Ξ_{it}, of length K+2. K is the number of skills, and is indexed by k. This human capital vector is given by

$$\Xi_{it} \equiv \left\{ \begin{array}{c} \xi_{it}^1 \\ \xi_{it}^2 \\ \vdots \\ \xi_{it}^K \\ Exp_{it} \\ G_i \end{array} \right\}$$

where ξ_{it}^k is skill k, Exp_{it} is experience measured by number of years working full-time, and G_i is a fixed indicator of worker i's learning type.

Each worker i begins her career by taking a draw of skills from a multivariate normal distribution,

$$\xi_{i1} \overset{iid}{\sim} N(\mu_{\Gamma_i}, P_1).$$

Define $\xi_{it} \equiv \begin{pmatrix} \xi_{it}^1 & \xi_{it}^2 & \dots & \xi_{it}^K \end{pmatrix}'$, i.e. ξ_{it} is an individual's skills vector. Note that the mean of the initial draw can vary by type, however the variance is common to all workers. I impose this assumption for model estimation.

In each period, workers choose either an occupation or unemployment. I describe the worker's problem in the next section, but for now I consider the problem using the generic occupational index $j \in 0, 1, 2, .., J$ for notational convenience. Occupation 0 represents unemployment. Each occupation demands a time-invariant set of skill levels (LV) and importance (IM), denoted by

$$S_j \equiv \begin{Bmatrix} s_{j,LV}^1 & s_{j,IM}^1 \\ s_{j,LV}^2 & s_{j,IM}^2 \\ \vdots & \vdots \\ s_{j,LV}^K & s_{j,IM}^K \end{Bmatrix}$$

where for j = 0 (unemployment), this is a matrix of zeros.

Why separate skills into a set of levels and importance? These vectors serve distinct purposes, purposes that are missed if I do not take care to differentiate them. In learning-by-doing, workers converge to the level of skills demanded by their chosen occupation. In contrast, the skill importance determines the salary penalty workers suffer as a consequence of level mismatch. It's more costly to be mismatched when a skill is central to executing a job than when the skill is ancillary.

To illustrate this difference, consider two occupations in the current O*NET database: telemarketers and lawyers. For both occupations, persuasion is central to the job. Of the 971 occupations in O*NET (V24.0), telemarketers and lawyers rank 9th and 10th in persuasion "importance," respectively. That said, the persuasive tasks in these two occupations are not equally difficult. Telemarketers must persuade potential customers to make a purchase, consequently they rank 160th among all occupations in persuasion "level." Lawyers must persuade judges, juries, clients, and peers to agree with complex, contentious, and often high-stakes arguments. They rank 3rd.

Returning to the vector of worker skills, the law of motion for these skills is given by

$$\xi_{i,t+1} = (\mathbf{I}_K - \Gamma_{G_i})\,\xi_{it} + \Gamma_{G_i} S_{j,LV} + \eta_{i,t+1} \tag{4}$$

where $\mathbf{I}_K$ is the KxK identity matrix, $\eta_{i,t+1} \overset{iid}{\sim} N(0, \Sigma_\eta)$ is a $(K$x1) multivariate normal draw, and

$$\Gamma_{G_i} = \begin{bmatrix} \gamma^1_{G_i} & 0 & \dots & 0 \\ 0 & \gamma^2_{G_i} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \gamma^K_{G_i} \end{bmatrix}$$

is a diagonal matrix of the worker type- and skill-specific learning speeds. These parameters govern how quickly a worker of type G_i absorbs skill k to suit that particular skill demand at her chosen occupation. Due to the functional form, the worker closes exactly the percentage $\gamma^k_{G_i}$ of the skill-k "mismatch" between her skill and the occupation's skill demand in every period. These $\gamma^k_{G_i}$ are the central parameters of this paper. The estimation routine detailed in section V is primarily intended to uncover these parameters and study how they differ by worker type.

In addition to the skill adjustment just described, workers accumulate general human capital when they work. General human capital, Exp_{it}, is simply measured as the total number of periods the worker has been employed full-time. Therefore, the evolution of general human capital is given by

$$Exp_{i,t+1} = \begin{cases} Exp_{it} + 1 & \text{if } i \text{ employed at time t} \\ Exp_{it} & \text{if } i \text{ unemployed at time t} \end{cases}$$

All skills and general experience evolve after a worker produces in a given period. Production is described in the following subsection.

As a final note, workers have full information. They know their own learning type and state, the occupational skill demands, and the parameters of all distributions. The same is true for global prices and parameters used throughout this model.

IV.B. The Economy, Production, and Salaries

By choosing an occupation, a worker establishes her own single-worker firm and produces that occupation's output. The worker bears a fixed cost if she is switching into that occupation. "Stayers" do not bear a fixed cost. All workers reside in a small open economy, meaning neither individual nor aggregate decisions affect goods prices. In the model, this implies that aggregate decisions do not influence labor market prices either. In short, all prices are set by the global economy. By constructing this context, I am able to study the single-agent decision in isolation.

Workers choose an occupation in each period, then produce that occupation's output, with the number of units produced given by

$$y_{ijt} = \exp\left\{\alpha_i + \alpha_j + X_{it}'\beta_{exp}^{G_i} + \beta_4 M_{ijt} + \epsilon_{ijt}\right\}$$

where α_i, α_j stand for individual and occupation fixed effects. X_{it} is a vector of the worker's experience up to a cubic term. Note that the parameters governing returns to general experience are type-specific. M_{ijt} is the worker's skill mismatch with occupation j, and $\epsilon_{ijt} \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$ is a Gaussian residual, i.i.d. across and within all indices.

Mismatch is defined as

$$M_{ijt} \equiv \sum_{k=1}^{K} s^k_{j,IM} \max\left\{s^k_{j,LV} - \xi^k_{it}, 0\right\}. \tag{5}$$

This is a weighted sum of skill *level* mismatch. The weights are skill importance, meaning it's more costly to be under-qualified in essential skills than secondary ones. There is no benefit to over-qualification, e.g. workers will not be paid for knowing calculus when only arithmetic is needed for the job. Finally, β_4 is the return of skill mismatch on salary, which I expect to be negative.

Salaries (w_{ijt}) are piece-rate, with the price of output (p_{jt}) set by the world economy:

$$\begin{aligned} w_{ijt} &= p_{jt} y_{ijt} \\ \log(w_{ijt}) &= \log(p_{jt}) + \alpha_i + \alpha_j + X'_{it}\beta_{exp} + \beta_4 M_{ijt} + \epsilon_{ijt} \end{aligned} \tag{6}$$

There is no intensive margin for a worker's labor supply within a period. Once a worker chooses an occupation, output and salary are fixed. Demand for the goods is perfectly elastic at the world price, and worker-level supply is perfectly inelastic, conditional on occupational choice.

The log of all goods prices, $\log(\vec{p_t})$ follows a random walk in the model, allowing me to easily integrate out expected future shocks from the worker's problem. To be complete, the log of a given good's price evolves according to

$$\log(p_{j,t+1}) = \log(p_{jt}) + \omega_{j,t+1}$$

where $\omega_{j,t+1} \overset{iid}{\sim} N(0, \sigma^2_\omega)$.

IV.C. The Worker's Problem

In each period, worker i maximizes her expected discounted sum of lifetime utility. Workers must retire at age 65. Then, worker i's problem at age a is to choose occupation j from the set of all occupations (or unemployment), $\mathfrak{J} = \{0, 1, ..., J\}$:

$$\max_{j \in \mathfrak{J}} \left\{ \mathbb{E} \left[\sum_{t=a}^{65} \rho^{t-a} U(j, \Phi_{it}) \right] \right\}$$

where ρ is the discount factor. Φ_{it} is the state vector, which I will fully describe shortly. The flow utility for an agent who chooses occupation j is

$$U(j, \Phi_{it}) = \mathbb{E}[\ln(w_{ijt}(\Phi_{it}))] + \psi_j + \nu_{ijt} - C_j \mathbb{I}\{j \neq \mathcal{J}(i, t-1)\} \tag{7}$$

where

- ψ_j is the time-invariant period utility payoff of occupation j, capturing compensating differentials, or the non-pecuniary benefits of occupations.
- $\nu_{ijt} \overset{iid}{\sim} T1EV(\sigma_1, \sigma_2)$. The value of σ_1 dictates the mean of the draws, but does not influence this problem, since it's an equal scaling of each utility term. The shocks themselves are important for two reasons. First, they give way to the familiar closed form switching probabilities in discrete choice models. Second, they help match the observed gross flows of workers between occupations.
- C_j is the time-invariant and unilateral cost of switching into occupation j.
- $\mathcal{J}(i, t-1)$ is worker i's chosen occupation at time $t-1$.

The expectation of wage is taken over the shock to production, ϵ_{ijt}. Neither the worker nor the econometrician observes this shock prior to the worker's decision.

With the flow utility and terminal condition specified, the worker's problem at time t and age a is given by the following Bellman equation:

$$V_a(\overbrace{\Phi_{it}}^{\text{State}}) = \begin{cases} \max_{j \in \mathfrak{J}} \begin{cases} \mathbb{E}[ln(w_{ijt}(\Phi_{it}))] + \psi_j + \nu_{ijt} - C_j \mathbb{I}\{j \neq \mathcal{J}(i,t-1)\} + \\ \rho \mathbb{E}[V_{a+1}(\Phi_{i,t+1}) | \Phi_{it}, \mathcal{J}(i,t)]\}, \end{cases} & a < 65 \\ \\ \max_{j \in \mathfrak{J}} \{\mathbb{E}[ln(w_{ijt}(\Phi_{it}))] + \psi_j + \nu_{ijt} - C_j \mathbb{I}\{j \neq \mathcal{J}(i,t-1)\}, & a = 65 \end{cases} \tag{8}$$

A few clarifying notes about this problem: the choice of $j = 0$ represents unemployment, and there is no switching cost for choosing unemployment. While there is no wage for the unemployed, $\psi_0 = b$, so there is some utility from being unemployed. The skill vector of unemployment is $S_0 = \vec{0}$, meaning the worker's skills will depreciate in unemployment. The expectation of the continuation value is taken over future shocks $\left(\{\omega_{it'}, \nu_{ijt'}\}_{t'=t+1}^{65}, \{\epsilon_{ijt'}, \eta_{it'}\}_{t'=t}^{65}\right)$. These series of shocks $\{\epsilon_{ijt'}\}_{t'=t}^{65}$ and $\{\omega_{it'}\}_{t'=t+1}^{65}$ have no influence on the worker's decision, since they are mean-zero linear Gaussian terms, and therefore integrate out to zero. Finally, there is no savings or borrowing technology.

With so many shocks that do not affect occupation choice, I should clarify which stochastic factors the workers observe before making their decision: occupational prices and the Type-I Extreme Value preference draws. I make timing assumptions in order to conveniently solve the model, assumptions that are presented visually in figure VI. First, each worker observes the vector of occupational good prices ($\vec{p_t}$) as well as the vector of T1EV draws (ν_{ijt}). Second, she chooses her period t occupation. Third, she observes the production shock (ϵ_{ijt}), produces, and receives her salary. Finally, she realizes the shock to skills evolution (η_{ijt}) and her skills evolve according to the law of motion.

To conclude the model's formal presentation, the state vector for worker i in period t is

$$\Phi_{it} = \{\Xi_{it}, \mathcal{J}(i,t-1), \nu_{ijt}, \vec{p_t}\}$$

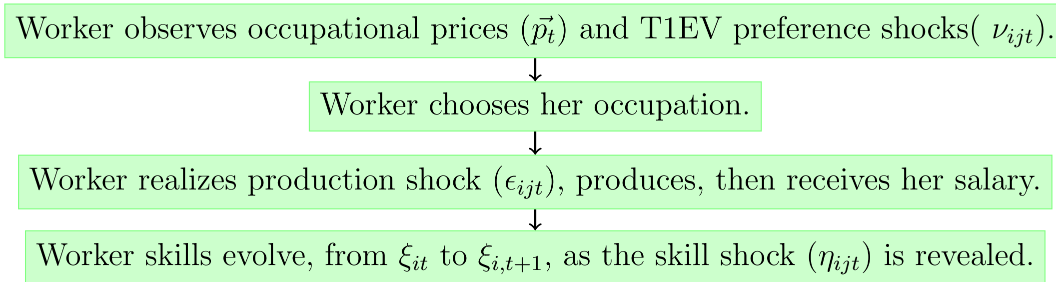


Figure VI
Timing assumptions for period t.

In order, this state vector consists of

- Ξ_{it}: The vector of human capital: the skills, experience, and worker type.
- $\mathcal{J}(i, t-1)$: The individual's chosen occupation in $t-1$. This is only relevant to the worker's decision through switching costs.
- $\nu_{ijt} \overset{iid}{\sim} T1EV(\sigma_1, \sigma_2)$, described above. These shocks are not observed by the econometrician.
- $\vec{p_t}$, the full set of occupational piece-rates in period t.

I have not included age in the state, though age is important due to the terminal condition. Note that the Value function is written with an age subscript. Equivalently, I could drop this subscript and include age in the state.

IV.D. Choice Probabilities

Two assumptions about the occupational utility shock (ν_{ijt}) are crucial to solving choice probabilities. The closed-form expression for choice is essential in solving the worker's problem computationally. These two assumptions are

1. The shocks follow the Type-I Extreme Value (Gumbel) distribution, and are i.i.d. across and within individuals, occupations, and periods.
2. Workers observe the full set of ν_{ijt} before they make their period t occupational choice. This is sometimes called the "Rust assumption," following Rust (1987).

To develop these choice probabilities, let $V_a^{\mathcal{J}(i,t)}(\Phi_{it}; \theta)$ represent the expected present value of a (potentially counter-factual) occupational choice, $\mathcal{J}(i, t)$, *without* the current period's T1EV shock. θ represents the parameters of the worker's problem. Then, the probability that worker i chooses some j in this period can be written as[8]

$$
\begin{aligned}
\Pr\{\mathcal{J}(i,t) = j|\Phi_{it}\} &= \Pr\{V_a^j(\Phi_{it};\theta) + \nu_{ijt} > V_a^{j'}(\Phi_{it};\theta) + \nu_{ij't}\} \; \forall \; j' \neq j \\
&= \Pr\{V_a^j(\Phi_{it};\theta) - V_a^{j'}(\Phi_{it};\theta) > \nu_{ij't} - \nu_{ijt}\} \; \forall \; j' \neq j \\
&= \frac{\exp\{V_a^j(\Phi_{it};\theta)\}}{\sum_{j'=0}^{J} \exp\{V_a^{j'}(\Phi_{it};\theta)\}} \qquad (9)
\end{aligned}
$$

For reference, the T1EV-less choice-specific value is

$$
\begin{aligned}
V_a^{\mathcal{J}(i,t)}(\Phi_{it};\theta) = \mathbb{E}[\log(p_{jt})] + \alpha_i + \alpha_j + X'_{it}\beta_{exp} + \beta_4 M_{ijt} + \psi_j - C_j \mathbb{I}\{j \neq \mathcal{J}(i,t-1)\} + \\
\rho \mathbb{E}[V_{a+1}(\Phi_{i,t+1})|\Phi_{it}, \mathcal{J}(i,t)]
\end{aligned}
$$

IV.E. Proposition

Proposition 1 (Option values increase in worker types)**.** *Assume* $t < T$*,* $\beta_4 < 0$ *and that* $s_{j^*,LV} \geq s_{j',LV}$*, with at least one inequality strict. Further assume that other occupations,* j*, have* $s_{j,LV} \geq s_{j^*,LV}$*. Then for* $\Phi_{i^*t} = \Phi_{i't}$*, aside from* $G_{i^*} > G_{i'}$*,*

$$
\frac{\Pr\{\mathcal{J}(i^*,t) = j^*|\Phi_{i^*t}\}}{\Pr\{\mathcal{J}(i^*,t) = j'|\Phi_{i^*t}\}} > \frac{\Pr\{\mathcal{J}(i',t) = j^*|\Phi_{i't}\}}{\Pr\{\mathcal{J}(i',t) = j'|\Phi_{i't}\}}. \qquad (10)
$$

Proof: Appendix A.

Translated: ceteris paribus, a low-type worker is less likely than a high-type worker to choose a higher-skilled occupation, relative to the probability of choosing a lower-skilled occupation. This higher choice probability comes through the increased option value for higher-skilled occupations in future periods.

[8]For a full derivation, see, e.g. Train (2009), 74.

This is the core message of the model: workers who learn more quickly are more likely to choose more difficult jobs. Accordingly, these quick learners will be more likely to adjust after an adverse shock: they are able to close skill gaps in short order, and therefore realize the benefits of the transition sooner. Further, the model points to a dynamic complementarity of occupational choice. If a high-type worker chooses a high-skill job today, then their payoff for high-skill jobs increases in the future. Importantly, the marginal returns to choosing the high-skill job are increasing in worker type.

This proposition establishes that, in the early career, faster-learning workers will choose more skilled occupations than slower-learning workers. They do so not necessarily because they are better suited to those occupations during their early careers, but rather because the skill acquisition in high-skilled occupations pays off in their mid- and late-careers. I observe this result in the structural estimates in Section V. Further, this proposition helps explain why there faster-learning workers maintain a consistently high level of cognitive skills even during adverse events such as job loss.

Having stated the model, I now turn to the estimation of learning rates and the costs of skill mismatch. I will also provide evidence for the proposition above, namely, that high-type workers initially choose more skilled jobs, suffering a more costly mismatch penalty in the short run. However, they enjoy a long run pay-off in the form of higher earnings, and by mid-career they suffer the least costly mismatch penalty of all the worker types.

V. Structural Estimation

V.A. The Problem

The model presented in Section IV was written with structural estimation in mind. In particular, it lends itself to the estimation of learning speeds for each worker and skill type – a central contribution of this paper. Additionally, I use the model to estimate the Mincer equation's parameters, including the cost of skill mismatch. I estimate these parameters using a subsample of the Danish administrative data introduced in Section II.

To orient the reader, I rewrite two central equations which pervade this section. First is the skills law of motion, equation (4). Second is the wage observation, equation (6) (i.e. the Mincer equation).

$$\xi_{i,t+1} = (\mathbf{I}_K - \Gamma_{G_i})\,\xi_{it} + \Gamma_{G_i} S_{j,LV} + \eta_{i,t+1}$$
$$\log(w_{ijt}) = \log(p_{jt}) + \alpha_i + \alpha_j + X_{it}'\beta_{exp} + \beta_4 M_{ijt} + \epsilon_{ijt}.$$

where

$$M_{ijt} \equiv \sum_{k=1}^{K} s_{j,IM}^k \max\left\{s_{j,LV}^k - \xi_{it}^k, 0\right\}.$$

Since this paper's notation is cumbersome, the reader can refer to Table V as a notational resource. For purposes of estimating the model, I set the number of skills to $K = 3$. As I discuss in Section II, I classify these three skills as cognitive, manual, and interpersonal. This classification follows Lise and Postel-Vinay (2019), and I construct worker skills in a fashion similar to that paper.

Both equations come directly from the model, though they take the convenient form of linear regressions. These regressions cannot be directly estimated since worker skills ξ_{it} are latent variables. Fortunately, these latent variables emit a signal in each period when a worker is employed. This class of problems, featuring a time-evolving latent variables which emit regular signals, are collectively known as Dynamic Bayesian Networks (DBNs). DBNs nest a variety of well-known frameworks, such as Hidden Markov Models, the Kalman Filter, and Particle Filters. There are a variety of ways to estimate these models, depending on their particulars.

To solve the problem, I propose a flexible, non-linear approach to estimate each workers' evolving skills. Specifically, I use an adapted version of the Unscented Kalman Filter originally proposed in Julier and Uhlmann (1997).[9] I use this adapted filter within an Expectations-Maximization algorithm to estimate the series of individual-level unobserved worker skills for a panel of Danish workers covering 1991-2016. The next subsection will present both the Unscented Kalman Filter and Expectations-Maximization algorithm in detail.

To what end do I develop this approach? My central challenge is simultaneously estimating unknown parameters and unknown states. I cannot estimate parameters without data, nor can I estimate data without parameters. I address this "dual" problem by using an Expectations-Maximization algorithm. The Expectation-Maximization algorithm offers a solution by tackling the unknowns in two steps, iteratively estimating one set of unknowns at a time. The iterations index is q. In the Expectation step, I fix the problem's parameters and estimate the states. I estimate the sequence of states for each individual i, conditional on the parameters as well as past states and wages:

$$\hat{\xi}_{it'}^{(q)} = \mathbb{E}[\xi_{it'} | \hat{\theta}_i^{(q-1)}, \{\xi_{it}, w_{ijt}\}_{(1:t'-1)}] \tag{11}$$

where θ_i represents the collected parameters of the problem, and $\{x_{it}\}_{(1:t')} \equiv \{x_{it}\}_{t=1}^{t'}$ as a notational convention. With that set of estimates $\hat{\xi}^{(q)}$ in hand, the Maximization step

[9] For a well-written primer on the Unscented Kalman Filter, see Wan and van der Merwe (2002). However, be warned – their presentation of the filter with additive errors is incorrect. To correct the filter, a second draw of the sigma state must be taken after projecting the state and MSE in each period.

maximizes the likelihood

$$\hat{\theta}^{(q)} = \arg\max_{\theta \in \Theta} \mathcal{L}\left(\theta | \hat{\xi}^{(q)}, Y\right) \tag{12}$$

where $\hat{\xi}^{(q)}$ is the set of all latent variable estimates from the E-Step and Y is the set of all observed data relevant to the likelihood function. In the section after next, I detail the specifics of this algorithm.

V.B. Filtering Problems and Sigma Points

My version of the Unscented Kalman Filter (UKF) estimates the unobserved worker skills, given the problem parameters. The UKF leverages the model's structure, which I illustrate in Figure VII with a Trellis diagram, which shows how the latent variables, worker choice, and wages are related within and across periods. Putting words to this diagram, a period starts with state Φ_{it}. Worker i chooses an occupation ($\mathcal{J}_{it}$) based on the state, then receives a salary (w_{it}) based on both the state and occupational choice. Then, $\Phi_{i,t+1}$ is partly driven by occupational choice in period t.

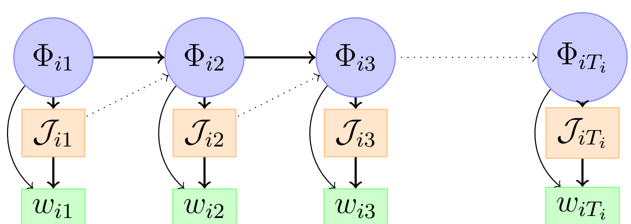


Figure VII
Trellis diagram depicting the evolution of states, occupations, and wages for worker i.

Before detailing the UKF, I state the filtering problem in general terms. To estimate states with a filter, I must already have a set of parameters $\theta = \{\alpha, \beta, \Gamma, G, \Sigma, \sigma\}$ in hand. Missing sub- and superscripts indicate the set of all parameters sharing the symbol. I will discuss estimating θ in the Expectation-Maximization (EM) algorithm, but for now I take it as given. The following describes the filtering problem for a single worker i.

To initiate the filter, I assume that each worker takes a draw of initial skills from a

TABLE V
Selected Notation

	Symbol	Notes/Data Source
Indices		
Workers index	$i \in 1, 2, ..., N$	
Occupations index	$j \in 0, 1, ..., J$	
Skills index	$k \in 1, 2, ..., K$	
Time index	$t \in 1, 2, ..., T_i$	
E-M Algorithm iteration	$q = 1, 2, ...$	
Parameters		
Worker i's learning type	G_i	Inferred through grades
Learning speed (type-skill specific)	$\gamma_{G_i}^k$	Estimated (EM Algorithm)
Matrix of learning speeds (type-specific)	Γ_{G_i}	
A variety of fixed effects	α	Estimated (Both)
Wage equation parameters	β	Estimated (EM Algorithm)
Occupation compensating differential	ψ_j	
Occupation switch cost	C_j	
Discount factor	ρ	Calibrated
All parameters of the worker's problem	θ	
Variables – Data		
Salary (measured monthly)	w_{ijt}	Danish Register
Age	a_{it}	Danish Register
Labor market experience	Exp_{it}	Danish Register
Occupational skill level	$s_{j,LV}^k$	O*NET
Occupational skill importance	$s_{j,IM}^k$	O*NET
Chosen occupation	$\mathcal{J}(i,t)$	Danish Register
Combined data	Y	
Variables – Unobserved/Inferred		
Skill state	ξ_{it}^k	Estimated (EM Algorithm)
Worker-occupation mismatch	M_{ijt}	Implied by $\{\xi_{it}, S_j\}$
Kalman filter mean squared error	$P_{i,t\|t/t-1}$	Estimated (EM Algorithm)
Stochastic Terms		
Residual: skills law of motion	$\eta_{it} \overset{iid}{\sim} N(0, \Sigma_\eta)$	
Residual: choice-specific flow utility	$\nu_{ijt} \overset{iid}{\sim} \text{T1EV}(\sigma_1, \sigma_2)$	
Residual: log production function	$\epsilon_{ijt} \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$	
Residual: occupational price random walk	$\omega_{it} \overset{iid}{\sim} N(0, \sigma_\omega^2)$	

common distribution,

$$\xi_{i1} \overset{iid}{\sim} N(\mu, P_1) \tag{13}$$

where P_1 is variance of initial skills, and $P_{i,1|0}$ is the estimated Mean Squared Error of this draw for worker i, with $P_{i,1|0} = P_1 \; \forall \; i$ initially with,

$$P_{i,1|0} = \mathbb{E}\left[(\hat{\xi}_{i,1|0} - \xi_{i1})(\hat{\xi}_{i,1|0} - \xi_{i1})'\right].$$

The initial projection of worker i's skills is $\hat{\xi}_{i,1|0} = \mu$. The initial distribution of skills is common across all types. Though I start the estimation routine with this assumption in place, it is not as restrictive as it first appears. The EM algorithm updates the estimated initial state and MSE for each individual during every algorithm iteration. This permits me to estimate a heterogeneous initial skill distribution by learning type without imposing ex-ante differences.

The unobserved skills immediately emit a signal in the form of worker i's salary in the first year of her career. Observing that salary, I update the projected estimate, $\hat{\xi}_{i,1|0}$, to the filtered estimate, $\hat{\xi}_{i,1|1}$, using Bayes rule. The initial prior distribution is given by (13), denoted by $p(\xi_{i1})$. The posterior distribution of skills, accounting for the observed salary, is

$$p(\xi_{i1}|w_{ij1}) = \frac{p(w_{ij1}|\xi_{i1})p(\xi_{i1})}{\int p(w_{ij1}|\xi_{i1})p(\xi_{i1})d\xi_{i1}}$$

The filtered state, $\hat{\xi}_{i,1|1}$, is the expectation of the state under this posterior distribution. Similarly, the filtered MSE, $\hat{P}_{i,1|1}$, is the variance of the filtered state under the distribution. Having completed this initial step, the filter proceeds through a two-step process in every period $t \in 2 : T_i$.

1. Projection Step

(a) Project the period t distribution based on the previous state,

$$p(\xi_{it}|w_{1:t-1}, \xi_{i,t-1}) = \int p(\xi_{it}|\xi_{i,t-1})p(\xi_{i,t-1}|w_{1:t-1})d\xi_{i,t-1} \tag{14}$$

(b) Calculate the state's projected mean ($\hat{\xi}_{i,t|t-1}$) and MSE ($\hat{P}_{i,t|t-1}$) on the basis of this distribution.

2. Filtering Step

(a) Filter the period t distribution on the basis of the observed salary

$$p(\xi_{it}|w_{1:t}, \xi_{i,t-1}) = \frac{p(w_{ijt}|\xi_{it})p(\xi_{it}|w_{1:t-1})}{\int p(w_{ijt}|\xi_{it})p(\xi_{it}|w_{1:t-1})d\xi_{it}} \tag{15}$$

(b) Calculate the state's filtered mean ($\hat{\xi}_{i,t|t}$) and MSE ($\hat{P}_{i,t|t}$) on the basis of this distribution.

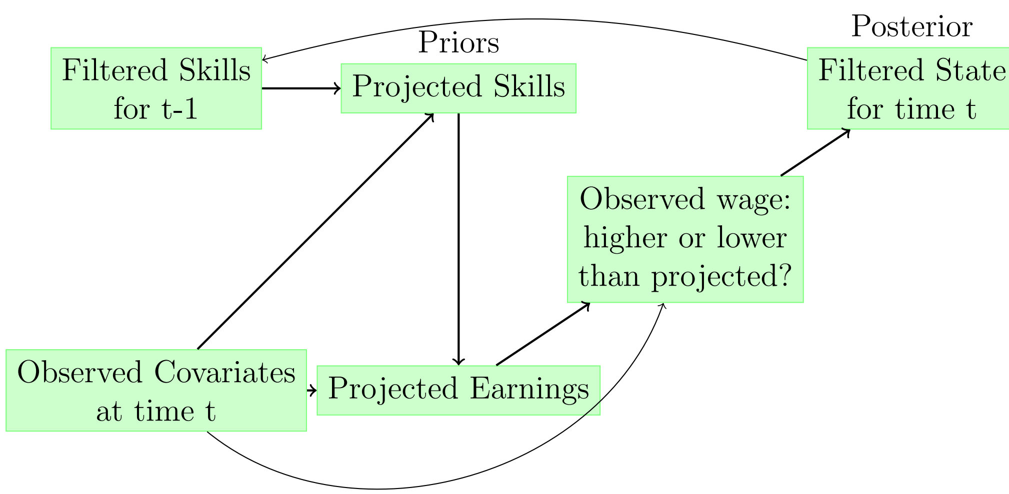


Figure VIII
The Filtering Process

In the case where both errors are Gaussian and both core equations are linear, this process becomes the standard Kalman Filter. The Kalman Filter makes repeated use of the fact that the normal distribution is its own conjugate prior, which leads to its convenient closed-form solution. The model does not lend itself to an application of the standard Kalman Filter because the skills enter the wage equation non-linearly. Since workers enjoy no additional wage returns to being overqualified, a sufficiently large shock η_{ijt} does not fully pass through to the wage equation. For this reason the first period posterior distribution $p(w_{ijt}|\xi_{i1})$ is *not* normally distributed. This non-normality propagates through the entire problem, so the expedient solution of the standard Kalman Filter does not apply.

Absent a closed form solution, the UKF estimates the mean and variance of each posterior distribution empirically. The data for those estimates are a parsimonious set of deliberately selected points from each prior distribution. These points, drawn twice in each period, are traditionally called "Sigma Points," a convention I maintain. By empirically estimating the mean and variance of each posterior distribution in this way, the UKF achieves an accuracy matching a third order Taylor approximation of the filtering problem (see Julier (2002)). In contrast, the Extended Kalman Filter (a popular alternative widely used by engineers) is a first order Taylor approximation of the filtering problem. Furthermore, the small set of points used in the UKF significantly reduces this problem's computational burden, a major advantage over another potential alternative, the Particle Filter.

Sigma Points play a crucial role in the UKF, so a brief overview is in order. Sigma Points are a set of $2K+1$ points symmetrically distributed about the mean of a distribution. Using the initial draw of skills as an example, the mean and covariance of this distribution are μ and P_1. The set of Sigma Points, χ, are calculated by

$$\begin{aligned}
\chi_0 &= \mu \\
\chi_x &= \mu + \left(\sqrt{(K+\lambda)P_1}\right)_x, \qquad x = 1, ..., K \\
\chi_x &= \mu - \left(\sqrt{(K+\lambda)P_1}\right)_{x-K}, \qquad x = K+1, ..., 2K
\end{aligned}$$

where $\lambda \in \mathbb{R}$ is a scaling parameter and $\left(\sqrt{(K+\lambda)P_1}\right)_x$ denotes the x^{th} column of the lower triangular Cholesky decomposition of the scaled covariance matrix. Though these are called Sigma "Points," each point is in fact a vector of length K, representing a potential vector of worker skills. Hence in this case, $\mathcal{X}$ is a 3x7 matrix, with each column representing one Sigma Point.

The mean and covariance of the source distribution are, empirically, precisely reflected in the Sigma Points. To estimate the mean and covariance post-transformation (read: the skill transformation into mismatch), each Sigma Point is passed through the observation equation. The result is a vector of length $2K + 1$, $\mathcal{Y}$, which reflects the worker's salary, had each Sigma Point been their true set of skills. Let the salary observation equation be represented as operator $\log(w_{ijt}) = f(\xi, Y)$, then

$$\mathcal{Y} = f(\mathcal{X}, Y)$$

where Y includes the covariates of the Mincer equation. Making use of these Sigma Points and their transformation, I detail how the UKF approximates the solution to the filtering problem.

V.C. The Unscented Kalman Filter and Expectation-Maximization Algorithm

I assume the initial distribution of states is multivariate normal, with mean μ and covariance P_1. Hence I initialize the filter by calculating Sigma Points from a distribution with mean μ and covariance P_1. The unobserved skills immediately emit a signal in the form of the first period's salary. To update the skills estimate,

1. Calculate Sigma Points around the initial distribution, $\mathcal{X}_{i,1|1}$. The mean ($\hat{\xi}_{i,1|0}$) and variance ($P_{i,1|0}$) are treated as if they are known, so I do not have to estimate them in this single instance.
2. Calculate the set of wages implied by those Sigma Points, $\mathcal{Y}_{i1} = f(\mathcal{X}_{i,1|1})$

3. Calculate the mean and variance of the *projected* wage:

$$\log(\hat{w}_{ijt}) = \sum_{x=0}^{2K} \tau_\mu^x \mathcal{Y}_{i1}^x$$
$$P_{i1}^w = \sum_{x=0}^{2K} \tau_\sigma^x \left(\mathcal{Y}_{i1,x} - \log \hat{w}_{ijt}\right)^2 + \sigma_\epsilon^2$$

   where the τ are weights, with $\sum_x \tau_\mu^x = \sum_x \tau_\sigma^x = 1$.[10] Note that the variance of the projected wage accounts for both the uncertainty in state estimation and the wage equation residual.

4. Estimate the covariance of the skills and wage:

$$P_{i1}^{\xi w} = \sum_{x=0}^{2K} \tau_\sigma^x \left[\mathcal{X}_{i,1|1}^x - \hat{\xi}_{i,1|0}\right] \left[\mathcal{Y}_{i1}^x - \log(\hat{w}_{ijt})\right]'$$

5. Compute the Kalman Gain, which is the covariance between skills and wage projections, divided by the variance of the wage estimate:

$$\kappa_{it} = P_{i1}^{\xi w} (P_{i1}^w)^{-1}$$

6. Compute the filtered state for the first period. This is the core of the algorithm, and will be discussed immediately below.

$$\hat{\xi}_{i,1|1} = \hat{\xi}_{i,1|0} + \kappa_{it} \left(\log(w_{ijt}) - \log(\hat{w}_{ijt})\right)$$

7. Compute the filtered MSE of the state for the first period:

$$P_{i,1|1} = P_{i,1|0} - \kappa_{i1} P_{i1}^w \kappa_{i1}'$$

Though technically complex, the intuition of filtering is straightforward. The econometrician makes a "best guess" regarding the value of unobserved skills, and specifies how uncertain he is about that guess. That guess implies a particular salary. The econometrician then updates the guess based on the true, observed salary. If the true salary is higher than the projected salary, his guess of the states was probably too low. If the true salary is lower than the projected one, the guess was probably too high.

But how much should the econometrician update the guess? Given the fact that there is noise throughout the process, it would be a bad idea to update it to perfectly fit the

[10]See Appendix D for details on these weights.

observed salary. Instead, he updates the guess a little more when salaries and skills are highly correlated, but a little less when salaries are noisy due to causes independent of worker skills. The mean squared error of the skills estimate is central to this process. It indicates the uncertainty surrounding the skills estimate, and thereby influences how strongly one should update the prior skills estimate once the wage is observed.

The filtering process for periods $2 : T_i$ is similar to that the first period, but not precisely the same. It starts by calculating a set of Sigma Points, then projecting the evolution of the state based on those Sigma Points. Starting with $t = 2$, in period t,

1. Calculate Sigma Points, $\mathcal{X}_{i,t-1}$ from a mean of $\hat{\xi}_{i,t-1|t-1}$ and covariance $P_{i,t-1|t-1}$.
2. Propagate those Sigma Points through the law of motion to project the evolution of a worker's skills:

$$\mathcal{X}_{i,t|t-1} = F\left(\mathcal{X}_{i,t-1}, s_{\mathcal{J}(i,t-1),LV}\right)$$

   where $F(\cdot, \cdot)$ is the law of motion, absent the noise term.
3. Calculate the projected skills for period t and its MSE:

$$\hat{\xi}_{i,t|t-1} = \sum_{x=0}^{2K} \tau_{\mu}^{x} \mathcal{X}_{i,t|t-1}^{x}$$
$$P_{i,t|t-1} = \sum_{x=0}^{2K} \tau_{\sigma}^{x} \left[\mathcal{X}_{i,t|t-1}^{x} - \hat{\xi}_{i,t|t-1}\right] \left[\mathcal{X}_{i,t|t-1}^{x} - \hat{\xi}_{i,t|t-1}\right]' + \Sigma_{\eta}$$

4. Repeat steps 1-7 detailed for the initial period, replacing "1"s with "t"s and "0"s with "$t-1$"s. Note that for step one, I generate $\mathcal{X}_{i,t|t}$ from a distribution with mean $\hat{\xi}_{i,t|t-1}$ and covariance $P_{i,t|t-1}$.

This completes the Unscented Kalman Filter. For a version with every detail, see Appendix D. See the same for details regarding the Unscented Kalman Smoother (UKS), proposed in Särkkä (2008). The details are omitted from the main text of this paper. The UKS solves the problem of updating the skills estimate, given *future* wages. The filter updates estimates based on past states and salaries, but future salaries also carry information about the current state. The smoother is an approximation of

$$p(\chi_{it}|w_{1:T_i}) = p(\chi_{it}|w_{1:t}) \int \left(\frac{p(\xi_{i,t+1}|\xi_{it}) p(\chi_{i,t+1}|w_{1:T_i})}{p(\xi_{i,t+1}|w_{1:t})} \right) d\xi_{i,t+1}$$

The properties of this smoother have not been studied in detail, and therefore I have some skepticism regarding its estimates. In Monte Carlo testing, the most precise results were found using a mixture of the smoother and filter estimates. This paper's estimates are calibrated with the mixture that performed best in Monte Carlo simulations: a $\frac{3}{4}$ weight on the filter, $\frac{1}{4}$ on the smoother.

Now that I have detailed the filtering process, I can return to the parameter estimates. While the UKF requires a fully specified model, the parameters of the model are unknown. Therefore, those parameters must be estimated jointly with the unobserved state. The Expectation-Maximization algorithm offers a solution to this dual problem. The "Expectation" step of the algorithm was just described: it is the UKF. The maximization step, fortunately, is more simple. Recall the "Maximization" step problem in equation (12):

$$\hat{\theta}^{(q)} = \arg\max_{\theta \in \Theta} \mathcal{L}\left(\theta | \hat{\xi}^{(q)}, Y\right)$$

The assumption that $\{\epsilon_{ijt}, \eta_{ijt}\}$ are mutually Gaussian and always independent from every other draw means that this problem boils down to a set of regressions. Additionally using the fact that worker skills follow a first-order Markov process, I can consequently write the likelihood function as

$$\begin{aligned} \mathcal{L}\left(\theta; \hat{\xi}, Y\right) &= p(\xi_{1:T}, Y_{1,T}|\theta) \\ &= \prod_{i=1}^{N} \left(p(\xi_{i1}) p(w_{ij1}|\xi_{i1}, Y_{i1}) \prod_{t=2}^{T_i} \left[p(\xi_{it}|\xi_{i,t-1}, Y_{i,t-1}) p(w_{ijt}|\xi_{it}, Y_{it}) \right] \right) \end{aligned} \tag{16}$$

where the conditional distributions above are distributed as

$$
\begin{aligned}
&\xi_{i1} \sim N(\mu, P_1) \\
&\xi_{it}|\xi_{i,t-1}, Y_{i,t-1} \sim N\left(\left(\mathbf{I}_K - \Gamma_{G_i}\right)\xi_{i,t-1} + \Gamma_{G_i} S_{(j,t-1,LV)}, \Sigma_\eta\right) \\
&w_{ijt}|\xi_{it}, Y_{it} \sim N\left(\log(p_{jt}) + \alpha_i + \alpha_j + X'_{it}\beta_{exp} + \beta_4 M_{ijt}, \sigma^2_\epsilon\right)
\end{aligned}
$$

Then, the log-likelihood is

$$
\log\left(\mathcal{L}\left(\theta; \hat{\xi}, Y\right)\right) = \underbrace{\sum_{i=1}^{N}\sum_{t=2}^{T_i} p(\xi_{it}|\xi_{i,t-1}, Y_{i,t-1})}_{\text{Regressions taking the form of the LOM}} + \underbrace{\sum_{i=1}^{N}\sum_{t=1}^{T_i} p(w_{ijt}|\xi_{it}, Y_{it})}_{\text{Mincer Regression}} + \sum_{i=1}^{N} p(\xi_{i1})
$$

The final term can be estimated with simple statistics, though in practice I update each worker's initial distribution using the results of the UKF. This completes the description of the Maximization-step.

The Expectations-Maximization algorithm itself is a simple iteration between the E-step and M-step. After initializing the algorithm with a set of parameters, the routine starts with the expectations step.[11] It alternates between equation (11) and equation (12) until convergence, defined as a sufficiently small change in the parameter estimates from one iteration to the next. The EM algorithm ensures an increasing log-likelihood in every iteration, and convergence is assured due to Wu (1983) and Dempster, Laird, and Rubin (1977).

At the end of each Maximization step, the following parameters are updated: $\{\beta, \Gamma, \alpha\}$. Note that the yearly occupational price and time-invariant occupational fixed effects are estimated as a set of single occupational-year fixed effects. Each iteration of the algorithm also updates $\{\hat{\xi}_{i,1|0}, P_{i,1|0}\}$ as well, using the filter/smoother mixed estimator and its associated MSE. These are estimated in the Expectations Step.

[11] I initialize the parameters of the Mincer regression, including fixed effects, by estimating the regression without a mismatch term in a pre-step. $\beta_4 = -0.05$ initially, to be roughly in line with linear returns to experience.

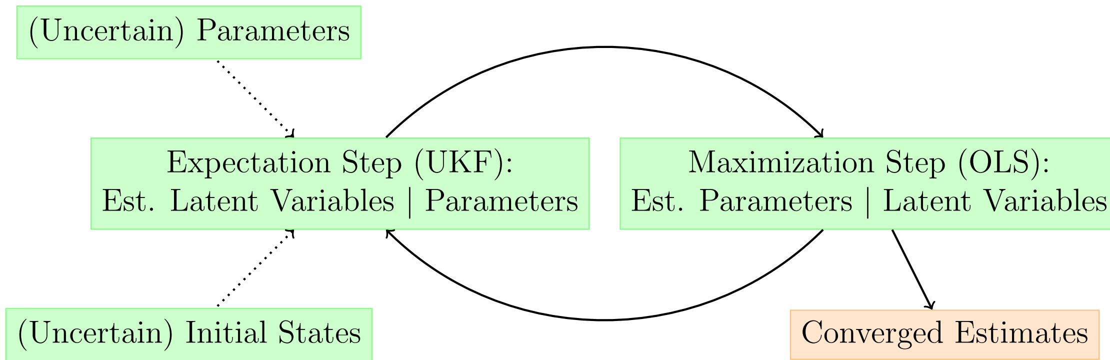


Figure IX
The Expectations-Maximization Algorithm

No ex-ante differences are imposed on the prior distribution of skills. I initialize the EM algorithm with an initial distribution of

$$\hat{\xi}_{i,1|0} \sim N\left(\begin{pmatrix}2\\2\\2\end{pmatrix}, \begin{pmatrix}0.5 & 0 & 0\\0 & 0.5 & 0\\0 & 0 & 0.5\end{pmatrix}\right)$$

for every individual's skills: cognitive, manual, and interpersonal, respectively. While technically $\xi \in \mathbb{R}^3$, in practice I bound the states by [0,7], a bound that rarely binds. All learning rates are initialized at 5%, i.e. $\gamma_{G_i}^k = 0.05 \; \forall \; \{G_i, k\}$. The only ex-ante difference imposed by type is the returns to general experience, which are estimated separately by learning type in a single Mincer Regression. I also make initial estimates of individual and occupation-year fixed effects. These are critical for projecting worker salaries in each period.

V.D. Identification

No result proves identification for the Unscented Kalman Filter. Even in the more simple case of the linear Kalman Filter, identification is conditional on the eigenvalues of the state law of motion matrix lying within the unit circle. Since the equivalent matrices for this model, Γ_{G_i}, are all diagonal and all represent the coefficients of a convex linear combination, the model trivially satisfies the identification requirement of the linear Kalman filter. I can offer no proof that this result holds up with the non-linearity in the observation equation. Having said that, the satisfaction of the linear Kalman identification condition alongside the convergence result of the EM algorithm together hint at identification here.

To address another challenge, how can three latent variables (the cognitive, manual, and social skills) be identified from one signal (the salary)? The answer is model structure. This is where I make use of the O*NET distinction between Skill Importance and Skill Level, introduced in Section IV. The mismatch term is specified so that mismatch in a skill highly important to the worker's occupation is more costly to the worker than a mismatch in a skill of lower importance. In turn, higher-importance skills and wages have a higher covariance than lower-importance skills and wage. Hence when the filter updates its skills estimates, it will put a greater weight on high-importance skills. This is intuitive: if a worker earns a lower than average salary in a manual skill-intensive occupation, it is more likely due to a manual skill deficiency than an interpersonal skill deficiency. In the results section, we will see that the filter is capable of making sharp distinctions in the development of each skill.

V.E. Data for the Structural Estimation

Though I introduce the data in Section II, more specifics are needed for the context at hand. As the demands and goals of the structural estimation differ from the establishment closure study, so do the data selected. Additionally, estimating this model using the population of Denmark is computationally infeasible. Therefore, I select a nonrandom subsample of the population that meet two useful criteria.

First, the sample's age lends itself to capturing early career wage dynamics and permits the longest possible panel. Since Statistics Denmark records worker occupations as of 1991, I select Danes born between 1968 and 1973. These cohorts are between 18 and 23 years of age when the panel starts. I observe work histories of up to 25 years, presenting an unusually long panel of this size. The second selection criterion is the observation of a grade from secondary education. I split workers into learning types using this grade. A significant advantage of this choice is that type identification is not based on labor market outcomes. I want to avoid type identification on that basis, as in that case the results would, in part, support my hypothesis by construction.

This second criterion has drawbacks. The sample does not reflect the cross-section of Danish society. Secondary education in Denmark does not function as it does in the United States, especially in the 1980s, the relevant time frame for this cohort. To start, secondary education is not compulsory in Denmark. More importantly, vocational education – not covered in the grades sample – was a popular option for teens at the time. In contrast, the individuals in this sample mostly attended "Higher Preparatory Examination" or HF, a two-year program intended to prepare students for post-secondary education. It appears to have been successful, as reflected by the fact that 84% of the sample completed some post-secondary degree, a rate far higher than Danish society in general. This is a drawback for my approach in the sense that I may not cover the full scope of the "learning speeds" hypothesis. The sample both did sufficiently well in earlier grades to qualify for HF, and may indicate a predilection for learning by choosing to attend HF.

This selection can also be an advantage. Insofar as choices like college education confound this paper's results, it mitigates such concerns. The sample is a group who made similar educational choices in their teens, but then differed in their performance in later secondary education. Further, the sample selection controls for differences in the caliber of instruction at this level. Though Danish schooling differs significantly across tracks (e.g. HF vs. vocational), it is fairly uniform within tracks. Grades are based on oral examinations and, in some core subjects, written examinations, which are drafted and graded by educators from outside a student's own school.

The sample is divided into five "Learning Types" based on within-cohort grade. Each quintile of the distribution defines one type. This is the level of essential heterogeneity, denoted by G_i. For each individual, I identify a career start based on two consecutive years of full-time employment without any educational pursuits. When that criterion is met, I mark the first year of that employment spell as the beginning of the individual's career. This definition differs from a more standard definition (see, eg. Sørensen and Vejlin (2014)), where a career starts after the worker completes her education. In the sample, it's reasonably common for workers to establish a full-time work history, then return to school for a short period in their late twenties or early thirties. Throwing out the work history prior these brief returns would entail a significant loss in panel length.

For each job a worker holds, I construct a monthly salary.[12] Using a monthly salary allows me to keep the initial and final year of a job spell, in contrast to yearly salary. Under the yearly measure, a job starting in September returns a fairly low salary due to the short timeframe. This can lead to a drastic over-statement of year-on-year salary growth. I also eschew using hourly wages, though they are observed in the data. As discussed at length in Lund and Vejlin (2016), the record of the hourly wage is often unreliable, and the authors call the measure into question despite considerable efforts to correct its extant issues. Further, wages are less regularly observed than yearly salary (by job), so this measure affords me superior coverage. Finally, in the case where a worker holds multiple jobs in a single year, I select the job with the highest salary paid in that year. This deviates from the convention of using a worker's job in November (which is the cleanest observation in these data). The job that pays the most salary in a year is also most likely to be a clean measure of the worker's true skills, if for no other reason than employment longevity.

Figure X shows how the log of the average monthly salary evolves over the course of a career. Note that this figure does not compare workers at the same age, but rather at the same point in their careers, i.e. relative to each worker's entry into the labor market. Therefore, the near equality of early salaries should not be interpreted as merely an effect of part-time work or impermanent jobs. Critically, the earnings inequality arising over these careers is a product of salary growth, not a persistently higher salary level. With earnings

[12] There are exceptions, as I do not always observe the beginning and end of jobs.

growth driving differences in labor market outcomes, as opposed to sustained level differences, I now turn my focus to estimating the heterogeneous learning rates.

Figure X
log(Monthly Salary) and Experience

V.F. Results

The estimates for learning rates are presented in Table VI. Note, importantly, that these estimates should not be interpreted as the growth rate of skills. Instead, these coefficients should be interpreted as the percent of skill mismatch that a worker closes in each period, for a given skill. For example, consider the type V (high-type learners) coefficient for cognitive skills. This should be read as "A high-type learner closes the mismatch between her own cognitive skills and that demanded by her chosen occupation at a rate of 14% per year." There are three distinct gradients of learning rates, one for each skill type. The highest types absorb the cognitive skills at 65% the rate of the lowest types, representing a 5.5 percentage point difference in learning rates. For manual skills, we observe the inverse: the highest types do not absorb manual skills at all, while the lowest types learn these skills at the rate of 4.1% percent per year. Finally, there is no discernible difference in learning speeds for interpersonal skills. Interestingly, interpersonal skills are absorbed the fastest, a result that stands in contrast to Lise and Postel-Vinay (2019).

TABLE VI
Learning Speed Estimates, by Worker Type

	Type I (Low)	Type II	Type III	Type IV	Type V (High)
Cognitive	0.085	0.087	0.102	0.115	**0.140**
Manual	**0.041**	0.029	0.021	0.008	0.002
Interpersonal	0.176	0.203	0.199	**0.211**	0.204

How do those learning rates translate into changing a workers' skill mismatch? Figure XI illustrates the evolution of average mismatch over the course of a career, split by learning type. At the outset of their career, the highest-type workers take on the highest skill mismatches. By the end of the panel, the highest types have the lowest average mismatch, reducing their average mismatch by nearly 75%. The average evolution of skills mismatch for all other types follows a similar pattern, but at a slower rate. As I predict, the rate of mismatch closure monotonically declines in learning type, though the differences between types I-III are negligible.

To fully understand the consequences of these heterogeneous learning rates, we must also discuss the evolution of the skills themselves. Figures XII-XIV show how average skills change over time, by learning type. Interpersonal skills evolve very similarly across types. Manual skills are mostly stagnant, and the strong fluctuations in later years may suggest there is a sample selection problem in this graph. Finally, the cognitive skills graph illustrates the different levels and slopes of cognitive skill growth, irrespective of occupation skill demands. High-type learners both have higher initial cognitive skills and a higher raw growth rate of skills than all other types.

Not only do the highest skilled workers close skill mismatches at the fastest rate, but

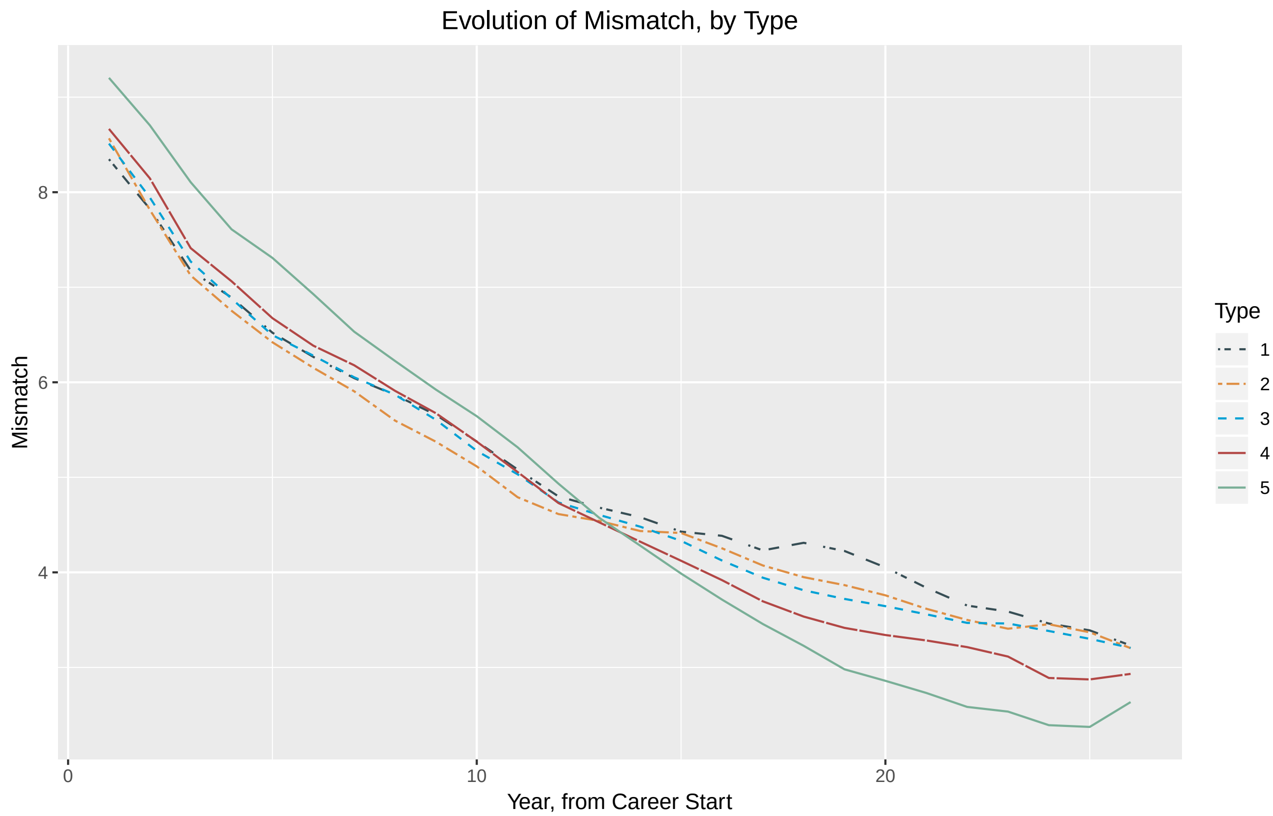


Figure XI
Evolution of Mismatch, by Type

they do so while moving into better paying occupations. Figure XV shows the average (occupational x year) fixed effect value by type. It is normalized in each year, and the values plotted are expressed in terms of standard deviations. By doing this, I illustrate occupation-level effects, as opposed to the secular wage growth that year fixed effects naturally reflect.

How much of the high-type worker's earnings over the course of a career be attributed to some innate characteristic, independent of skill growth? Very little, according to the average individual fixed effect estimates from the Mincer regression. The differences in average fixed effects across types is nearly non-existent, as illustrated in Table VII.

These results illustrate that acquiring skills is critical to driving wage growth. What is more, rapid skill acquisition – illustrated in the evolution of average skills – not only drives wage growth directly; it also allows for a quicker transition to other occupations if a worker chooses to switch. This underlying ability to acquire skills is more important to lifetime

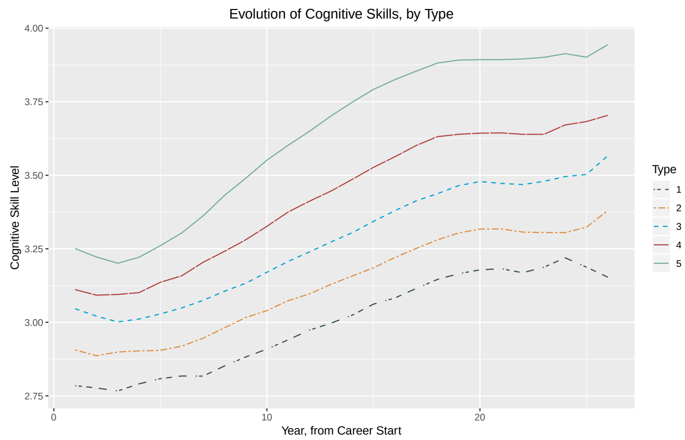


Figure XII
Cognitive Skills

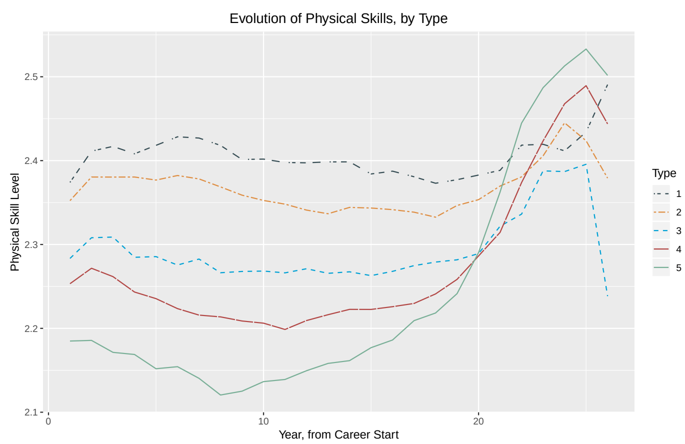


Figure XIII
Manual Skills

Figure XIV
Interpersonal Skills

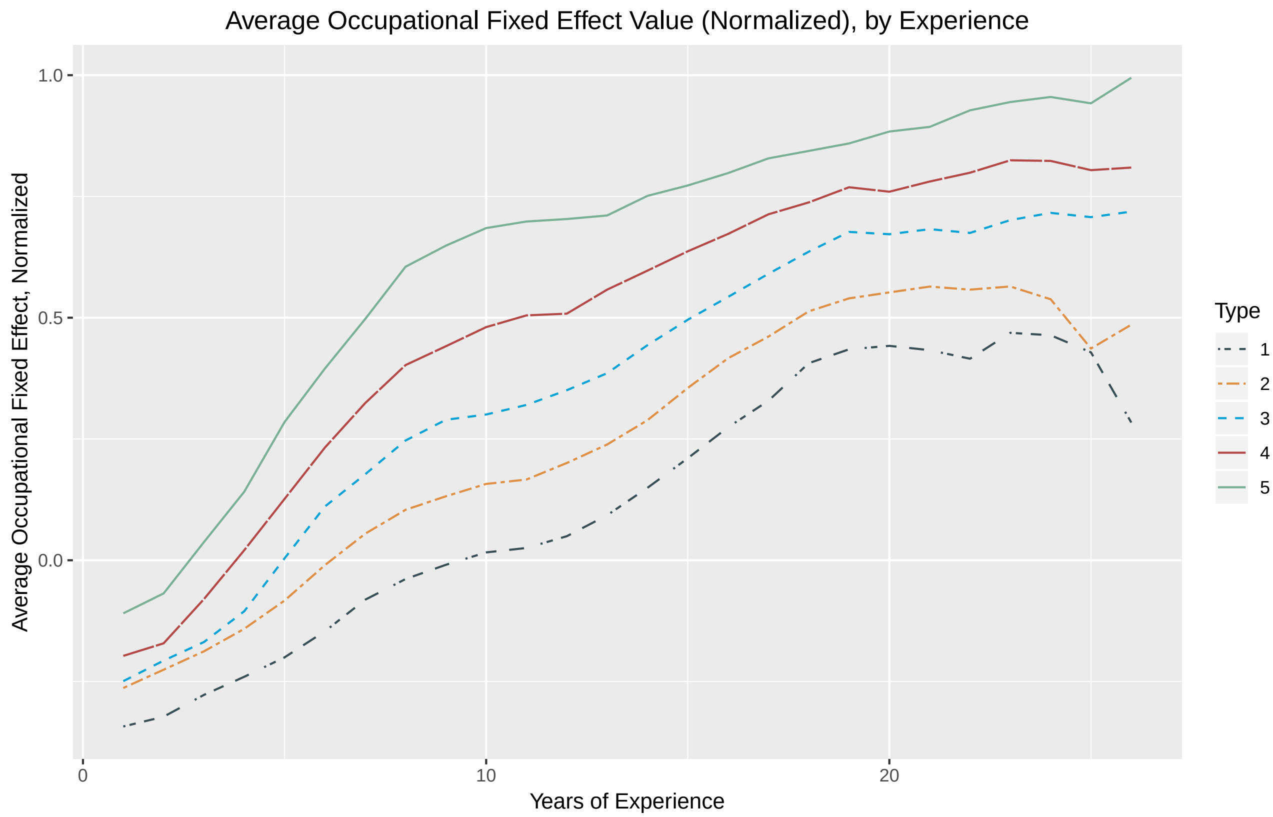


Figure XV
Average of Normalized Occupational/Year Fixed Effect

labor outcomes than the possession of the skills themselves.

In sum, I have evidence that workers differ substantially in their learning rates, particularly in acquiring cognitive skills. Further, these differences coincide with patterns of occupational choice: workers who quickly acquire cognitive skills choose skill-intensive occupations early in their career, and suffer an earnings penalty for doing so. However, as their career progresses, quicker learners reduce their realized mismatch to the lowest of all the learning types, and their earnings growth reflects the pattern. Re-enforcing the growth is the movement up the skills ladder throughout their career. This movement is enabled by the skills acquired during the early years of their career. Taken together, these results support the theoretical mechanism driving stronger recoveries to adverse shocks such as job displacement. Faster learners are quicker to adopt a new set of skills in the case that the demand for their own occupation falls, and are thus insulated from the vicissitudes of the labor market.

TABLE VII
Average Individual Fixed Effect, by Type

	Type I (Low)	Type II	Type III	Type IV	Type V (High)
Average ID Fixed Effect	11.50	11.50	11.49	11.50	11.49

TABLE VIII

Mincer Regression Coefficients

	Dependent variable:
	log(Salary)
Mismatch	-0.092^{***} (0.002)
Type I, Experience	-0.111^{***} (0.007)
Type I, $\exp^2$	0.004^{***} (0.0001)
Type I, $\exp^3$	-0.0001^{***} (0.00000)
Type II, Experience	-0.122^{***} (0.008)
Type II, $\exp^2$	0.006^{***} (0.0002)
Type II, $\exp^3$	-0.0001^{***} (0.00000)
Type III, Experience	-0.120^{***} (0.008)
Type III, $\exp^2$	0.006^{***} (0.0003)
Type III, $\exp^3$	-0.0001^{***} (0.00001)
Type IV, Experience	-0.104^{***} (0.009)
Type IV, $\exp^2$	0.006^{***} (0.0004)
Type IV, $\exp^3$	-0.0001^{***} (0.00001)
Type V, Experience	-0.082^{***} (0.009)
Type V, $\exp^2$	0.005^{***} (0.001)
Type V, $\exp^3$	-0.0001^{***} (0.00001)
Worker FE	Y
Occupation x Year FE	Y
Observations	560,386
R^2	0.648
Adjusted R^2	0.626
Residual Std. Error	0.493 (df = 526692)
Note:	*p<0.1; **p<0.05; ***p<0.01

VI. Conclusion

This paper provides evidence that rates of skill acquisition are important to labor market choices and outcomes. I do this in two ways: first, I show that workers who earned high grades in secondary schooling enjoy a long-run recovery to the salaries of their peers. Notably, this is more predictive of recovery than education. I interpret this as initial evidence that the ability to quickly acquire new skills is central to recovery from adverse labor market shocks. Second, I develop a dynamic Roy model featuring heterogeneous learning rates at its core. I directly estimate this model using a corrected Unscented Kalman filter within an Expectation-Maximization algorithm. These estimates confirm earlier predictions: the fastest learning workers choose occupations with a high mismatch during their early careers. Though they suffer an early wage penalty for this choice, they reap the benefits quickly. These workers acquire the skills appropriate to their chosen occupation, and as a result enjoy sustained wage growth. It is this growth—as opposed to a persistent level—that drives long run earnings inequality.

In the Introduction, I briefly discussed how government policies are designed to help workers hurt by adverse shocks. A theme common to active policies is retraining: the subsidization of education to update a worker's skills. The model and evidence suggests that this will fall short for some workers, and might explain why the observed take-up for these programs is often very low. What might help, in contrast, is retraining at a more fundamental level: equip workers with problem-solving tools which have general applications, instead of teaching them how to solve specific problems and complete particular tasks.

This paper leaves ample room for future research. Foremost on my mind is the broad definitions of the skills. "Cognitive" is a single skill, and perfectly transferable across occupations. While this coarse definition is still on the frontier of skills research, significant new avenues would be open if we could expand the list of skills. For example, physicians and lawyers are both cognitively demanding occupations. They are not, however, perfectly substitutable in the cognitive skills dimension. The more granular the skill definition, the better we can understand occupational choices, transitions, and how government policies can best support workers through labor market shocks. Unfortunately, expanding this set of skills remains unfeasible without stronger assumptions elsewhere in the model.

References

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

Autor, D. H., D. Dorn, and G. H. Hanson (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103(6), 2121–2168.

Autor, D. H., D. Dorn, G. H. Hanson, and J. Song (2014): "Trade Adjustment: Worker-Level Evidence," *The Quarterly Journal of Economics*, 129(4), 1799–1860.

Azoulay, P., J. S. Graff-Zivin, and J. Wang (2010): "Superstar Extinction," *The Quarterly Journal of Economics*, 125(2), 549–589.

Bobbio, E., and H. Bunzel (2018): "The Danish Matched Employer-Employee Data," *Working Paper*.

Boeri, T., and L. Calmfors (2001): *The role of unions in the twenty-first century.* Oxford University Press, Oxford, UK.

Davis, S., and T. von Wachter (2012): "Recessions and the Costs of Job Loss," *Brookings Papers on Economic Activity*.

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, 39(1), 1–38.

Dix-Carneiro, R. (2014): "Trade Liberalization and Labor Market Dynamics," *Econometrica*, 82(3), 825–885.

Eliason, M., and D. Storrie (2006): "Lasting or Latent Scars? Swedish Evidence on the Long-Term Effects of Job Displacement," *Journal of Labor Economics*, 24(4), 831–856.

Gathmann, C., and U. Schönberg (2010): "How General Is Human Capital? A Task-Based Approach," *Journal of Labor Economics*, 28(1), 1–79.

Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2019): "Multidimensional Skill Mismatch," *American Economic Journal: Macroeconomics*, Forthcoming.

Hamilton, J. D. (1994): *Time Series Analysis.* Princeton University Press, Princeton, NJ.

Hummels, D., R. Jørgensen, J. Munch, and C. Xiang (2014): "The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data," *American Economic Review*, 104(6).

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993): "Earnings Losses of Displaced Workers," *The American Economic Review*, 83(4), 685–709.

Jäger, S., and J. Heining (2019): "How Substitutable are Workers? Evidence from Worker Deaths," *R & R at American Economic Review*.

Jovanovic, B. (1979): "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87(5), 972–990.

Julier, S., and J. Uhlmann (1997): "New Extension of the Kalman Filter to Nonlinear Systems," *Signal Processing, Sensor Fusion, and Target Recognition*, 3068.

Julier, S. J. (2002): "The Scaled Unscented Transformation," in *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*, vol. 6, pp. 4555–4559.

Lindenlaub, I. (2017): "Sorting Multidimensional Types: Theory and Application," *Review of Economic Studies*, 84(3), 718–789.

Lise, J., and F. Postel-Vinay (2019): "Multidimensional Skills, Sorting, and Human Capital Accumulation," *R&R, American Economic Review*.

Lund, C., and R. Vejlin (2016): "Documenting and Improving the Hourly Wage Measure in the Danish IDA Database," *The Danish Journal of Economics*, 1(1), 1–35.

Neal, D. (1998): "The Link between Ability and Specialization: An Explanation for Observed Correlations between Wages and Mobility Rates," *The Journal of Human Resources*, 33(1), 173–200.

Pavan, R. (2011): "Career Choice and Wage Growth," *Journal of Labor Economics*, 29(3), 549–587.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55(5), 999–1033.

Särkkä, S. (2008): "Unscented Rauch-Tung-Striebel Smoother," *IEEE Transactions on Automatic Control*, 53(3), 845–849.

Seim, D. (2019): "On the incidence and effects of job displacement: Evidence from Sweden," *Labour Economics*, 57(1), 131–145.

Sørensen, K., and R. Vejlin (2014): "Return to Experience and Initial Wage Level: Do Low Wage Workers Catch Up?," *Journal of Applied Econometrics*, 29(6), 984–1006.

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.

Taber, C., and R. Vejlin (2019): "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market," *Revision requested, Econometrica*.

Topel, R. H., and M. P. Ward (1992): "Job Mobility and the Careers of Young Men," *The Quarterly Journal of Economics*, 107(2), 439–479.

Traiberman, S. (2019): "Occupations and Import Competition: Evidence from Denmark," *American Economic Review*, Forthcoming.

Train, K. E. (2009): *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, NY.

Wan, E. A., and R. van der Merwe (2002): *The Unscented Kalman Filter*chap. 7, pp. 221–280. John Wiley and Sons, Ltd.

Wu, C. F. J. (1983): "On the Convergence Properties of the EM Algorithm," *The Annals of Statistics*, 11(1), 95–103.

Yamaguchi, S. (2012): "Tasks and Heterogeneous Human Capital," *Journal of Labor Economics*, 30(1), 1–53.

Appendix A: Proposition Proof

I. Proposition I

To start, note that if

$$\frac{\Pr\{\mathcal{J}(i^*,t)=j^*|\Phi_{i^*t}\}}{\Pr\{\mathcal{J}(i^*,t)=j'|\Phi_{i^*t}\}} > \frac{\Pr\{\mathcal{J}(i',t)=j^*|\Phi_{i't}\}}{\Pr\{\mathcal{J}(i',t)=j'|\Phi_{i't}\}}$$

then

$$\log\left(\frac{\Pr\{\mathcal{J}(i^*,t)=j^*|\Phi_{i^*t}\}}{\Pr\{\mathcal{J}(i^*,t)=j'|\Phi_{i^*t}\}}\right) > \log\left(\frac{\Pr\{\mathcal{J}(i',t)=j^*|\Phi_{i't}\}}{\Pr\{\mathcal{J}(i',t)=j'|\Phi_{i't}\}}\right).$$

Recall, from equation (9), that for a given state the choice probability expression's denominator is the same for *all* choices. This means I can rewrite the inequality above as

$$\log\left(\frac{\exp\{V_a^{j^*}(\Phi_{i^*t};\theta)\}}{\exp\{V_a^{j'}(\Phi_{i^*t};\theta)\}}\right) > \log\left(\frac{\exp\{V_a^{j^*}(\Phi_{i't};\theta)\}}{\exp\{V_a^{j'}(\Phi_{i't};\theta)\}}\right).$$

To temporarily save on notation, let's consider the left hand side of this equation exclusively. Clearly the mathematical operations for the two sides will work symmetrically. Now, plugging in the choice-specific values, the LHS is

$$\overbrace{\log(p_{j^*t})+\alpha_i+\alpha_{j^*}+X'_{it}\beta_{exp}+\beta_4 M_{ij^*t}+\psi_{j^*}-C_{j^*}\mathbb{I}\{j^*\neq\mathcal{J}(i,t-1)\}}^{\text{Choice value for occupation } j^*}$$
$$+\rho\mathbb{E}[V_{a+1}(\Phi_{i,t+1})|\Phi_{it},\mathcal{J}(i,t)=j^*]-$$
$$\underbrace{\Big(\log(p_{j't})+\alpha_i+\alpha_{j'}+X'_{it}\beta_{exp}+\beta_4 M_{ij't}+\psi_{j'}-C_{j'}\mathbb{I}\{j'\neq\mathcal{J}(i,t-1)\} +\rho\mathbb{E}[V_{a+1}(\Phi_{i,t+1})|\Phi_{it},\mathcal{J}(i,t)=j']\Big)}_{\text{Choice value for occupation j'}}$$

Fortunately, this expression will simplify greatly. Immediately, the α_i and $X'_{it}\beta_{exp}$ terms cancel out within this expression. I can then subtract the occupational prices, occupation-time fixed effects, and switch costs from both sides of the inequality. Further, I can subtract the term $(\beta_4 M_{ij^*t} - \beta_4 M_{ij't})$ from both sides of the inequality due to the assumption that $\Phi_{it} = \Phi_{i't}$ (aside from the type). Taken together, that condition and the timing assumption regarding the evolution of the state mean that the two workers have the same mismatch in all occupations for period t. Therefore, I can rewrite the inequality as

$$\rho(\mathbb{E}[V_{a+1}(\Phi_{i^*,t+1})|\Phi_{i^*t}, \mathcal{J}(i,t) = j^*] - \mathbb{E}[V_{a+1}(\Phi_{i^*,t+1})|\Phi_{i^*t}, \mathcal{J}(i,t) = j']) >$$
$$\rho(\mathbb{E}[V_{a+1}(\Phi_{i',t+1})|\Phi_{i't}, \mathcal{J}(i',t) = j^*] - \mathbb{E}[V_{a+1}(\Phi_{i',t+1})|\Phi_{i't}, \mathcal{J}(i',t) = j']) \quad (17)$$

The problem has boiled down to which worker has the larger difference in the continuation values between j^* and j'. To establish the inequality, I start by considering the problem in period T-1, then proceed by recursive induction.

For $t = T-1$, the continuation value is simply the expected flow utility in period T. This expectation is taken over shocks $(\{\omega_{iT}, \nu_{ijT}, \epsilon_{ijT}, \eta_{ijT}\})$. Referring back to equation (8), the shocks $\{\epsilon_{ijT}, \omega_{iT}\}$ integrate out to zero. This leaves me with the T1EV utility shocks and the skill evolution shocks.

The flow utility for choice j in period T is similar to the "LHS" expression above, however there are important differences. The full expression is now dependent on the occupational choice in period T-1. To be concrete, the value of occupational choice j in period T, conditional on having chosen j' in T-1 is

$$V^j_{a+1}(\Phi_{i^*,t+1}|\Phi_{i^*t}, \mathcal{J}(i,t) = j') = log(p_{jT}) + \alpha_i + \alpha_j + X'_{iT}\beta_{exp}$$
$$+ \beta_4 M_{ijT}\left(\mathcal{J}(i^*, T-1) = j^*\right) + \psi_j - C_j \mathbb{I}\{j \neq j'\}$$

From here, I return to the inequality expressed in (17). I proceed by showing that this inequality holds in expectation for any period T choice aside from unemployment (where it is an equality). As the inequality holds in expectation for every choice $j \neq 0$, it also will hold for the expected maximum utility of the workers' period T choice.

As before, I can subtract all terms not dependent on i from both sides of the inequality. The individual fixed effects cancel out within each side of the inequality. This leaves me to write the inequality with only mismatch terms remaining. Doing so, while substituting in for the mismatch term, the inequality is now

$$\beta_4 \Bigg(\sum_k s^k_{j,IM} \max \left\{ s^k_{j,LV} - \xi^k_{i^*T} \left(\mathcal{J}(i^*, T-1) = j^* \right), 0 \right\} \\ - \sum_k s^k_{j,IM} \max \left\{ s^k_{j,LV} - \xi^k_{i^*T} \left(\mathcal{J}(i^*, T-1) = j' \right), 0 \right\} \Bigg) > \\ \beta_4 \Bigg(\sum_k s^k_{j,IM} \max \left\{ s^k_{j,LV} - \xi^k_{i'T} \left(\mathcal{J}(i', T-1) = j^* \right), 0 \right\} \\ - \sum_k s^k_{j,IM} \max \left\{ s^k_{j,LV} - \xi^k_{i'T} \left(\mathcal{J}(i', T-1) = j' \right), 0 \right\} \Bigg)$$

Recall the condition $\beta_4 < 0$ (i.e. mismatch is penalized rather than rewarded), so multiplying either side to remove this term flips the inequality. I now plug in the law of motion, and use the fact that η_{ijt} is normally distributed. As before, I only consider the LHS to temporarily save on notation. Then,

$$
\sum_k s^k_{j,IM} \max\left\{s^k_{j,LV} - \xi^k_{i^*T}\left(\mathcal{J}(i^*,T-1)=j^*\right),0\right\} -
$$

$$
\sum_k s^k_{j,IM} \max\left\{s^k_{j,LV} - \xi^k_{i^*T}\left(\mathcal{J}(i^*,T-1)=j'\right),0\right\}
$$

$$
= \sum_k s^k_{j,IM} \max\left\{s^k_{j,LV} - \left(\gamma^k_{G_i} s^k_{j^*,LV} + (1-\gamma^k_{G_i})\xi^k_{i^*t} + \eta_{i^*j',T}\right),0\right\} -
$$

$$
\max\left\{s^k_{j,LV} - \left(\gamma^k_{G_i} s^k_{j',LV} + (1-\gamma^k_{G_i})\xi^k_{i^*t} + \eta_{i^*j^*,T}\right),0\right\}
$$

$$
= \sum_k s^k_{j,IM} \int_{-\infty}^{Q^{i^*,k}_{j,j^*}} s^k_{j,LV} - \left(\gamma^k_{G_i} s^k_{j^*,LV} + (1-\gamma^k_{G_i})\xi^k_{i^*t}\right) f_\eta(\eta_{i^*j^*,T})\partial\eta - \mathbb{E}[\eta_{i^*j^*,T}|\eta_{i^*j^*,T} < Q^{i^*,k}_{j,j^*}] -
$$

$$
\left(\int_{-\infty}^{Q^{i^*,k}_{j,j'}} s^k_{j,LV} - \left(\gamma^k_{G_i} s^k_{j',LV} + (1-\gamma^k_{G_i})\xi^k_{i^*t}\right) f_\eta(\eta_{i^*j',T})\partial\eta - \mathbb{E}[\eta_{i^*j',T}|\eta_{i^*j',T} < Q^{i^*,k}_{j,j'}]\right)
$$

where $Q^{i^*,k}_{j,j^*} \equiv s^k_{j,LV} - \left(\gamma^k_{G_i} s^k_{j^*,LV} + (1-\gamma^k_{G_i})\xi^k_{i^*t}\right)$ and $f_\eta()$ is a normal pdf. There are two differences between the occupational mismatch expressions. First is the integral's upper bound. Second is the period T-1 skill acquisition term in the integrand ($\gamma^k_{G_i} s^k_{j',LV}$).

Intuitively, the next step is to show that the higher type decreases future mismatches at a higher rate by choosing more skilled jobs today. Multiplying both sides by -1 flips the inequality sign back to its original direction, and I can write the problem (exclusively for some element k, again to save on notation) as

$$
\Pr\{\eta < Q^{i^*}_{j,j'}\}Q^{i^*}_{j,j'} - \mathbb{E}\left[\eta_{i^*j',T}|\eta_{i^*j',T} < Q^{i^*}_{j,j'}\right] - \left(\Pr\{\eta < Q^{i^*}_{j,j^*}\}Q^{i^*}_{j,j^*} - \mathbb{E}\left[\eta_{i^*j^*,T}|\eta_{i^*j^*,T} < Q^{i^*}_{j,j^*}\right]\right) >
$$

$$
\Pr\{\eta < Q^{i'}_{j,j'}\}Q^{i'}_{j,j'} - \mathbb{E}\left[\eta_{i'j',T}|\eta_{i'j',T} < Q^{i'}_{j,j'}\right] - \left(\Pr\{\eta < Q^{i'}_{j,j'}\}Q^{i'}_{j,j'} - \mathbb{E}\left[\eta_{i'j^*,T}|\eta_{i'j^*,T} < Q^{i'}_{j,j^*}\right]\right)
$$

Finally, we can collect like terms and bound them. By bounding two separate terms by zero below, we can establish the strict inequality. These two inequalities are

$$\left(\Pr\{\eta < Q^{i^*}_{j,j'}\}Q^{i^*}_{j,j'} - \Pr\{\eta < Q^{i^*}_{j,j^*}\}Q^{i^*}_{j,j^*}\right) - \left(\Pr\{\eta < Q^{i'}_{j,j'}\}Q^{i'}_{j,j'} - \Pr\{\eta < Q^{i'}_{j,j'}\}Q^{i'}_{j,j'}\right) > 0$$

and

$$\left(\mathbb{E}\left[\eta_{i^*j^*,T} | \eta_{i^*j^*,T} < Q^{i^*}_{j,j^*}\right] - \mathbb{E}\left[\eta_{i^*j',T} | \eta_{i^*j',T} < Q^{i^*}_{j,j'}\right]\right) -$$
$$\left(\mathbb{E}\left[\eta_{i'j^*,T} | \eta_{i'j^*,T} < Q^{i'}_{j,j^*}\right] - \mathbb{E}\left[\eta_{i'j',T} | \eta_{i'j',T} < Q^{i'}_{j,j'}\right]\right) > 0$$

Both inequalities hold as a result of Jensen's Inequality. Since I have assumed that the skills in j exceed those of j^* and j', I know I am operating on the upper half of the normal cdf (i.e. all $Q > 0$). Then by the linearity in the differences in Q and the concavity of this segment of the normal cdf, the above inequalities hold. □

Appendix B: Danish ISCO Occupational Classification-1988

TABLE A.I
Danish ISCO-88 Occupations

	One-Digit Occupation		Two-Digit Occupation
1	Top-Level Management In Companies, Organizations and the Public Sector	12	Management in Companies with Ten or More Employees
		13	Management in Companies with Fewer than Ten Employees
2	Work That Requires the Highest Skills in the Field in Question	21	Research and/or application of skills in the non-biological branches of natural sciences as well as computer science, statistics, architecture and technical sciences
		22	Research and/or application of skills in medicine, pharmacy and the biological branches of science and midwifery, general nursing work, etc.
		23	Teaching in primary schools, vocational schools, colleges and colleges, as well as research organization and supervision of teaching work
		24	Research and/or application of skills in the social sciences and humanities
3	Work that requires intermediate level skills	31	Technician work in non-biological subjects
		32	Technician work in biological subjects
		33	Teaching and care work
		34	Work with sales, finance, business services, administration, etc.
4	Office work	41	Internal office work
		42	Office work with customer service
5	Sales, service and care work	51	Service and care work
		52	Sales work
6	Work in agriculture, horticulture, forestry, hunting and fishing, which requires basic skills	61	Work in agriculture, horticulture, forestry, hunting and fishing, which requires basic skills
7	Crafted work	71	Work on raw material extraction and building crafts
		72	Metal and machine work
		73	Precision crafts, graphic work and the like
		74	Other Crafted Work
8	Process and machine operator work as well as transport and civil engineering work	81	Work on stationary process plants
		82	Operation of industrial machinery
9	Other work	91	Cleaning and renovation work, courier and watch service and telephone and door sales etc.
		92	Assistance in agriculture, horticulture, fisheries and forestry
		93	Manual work in the construction, transport and manufacturing sectors

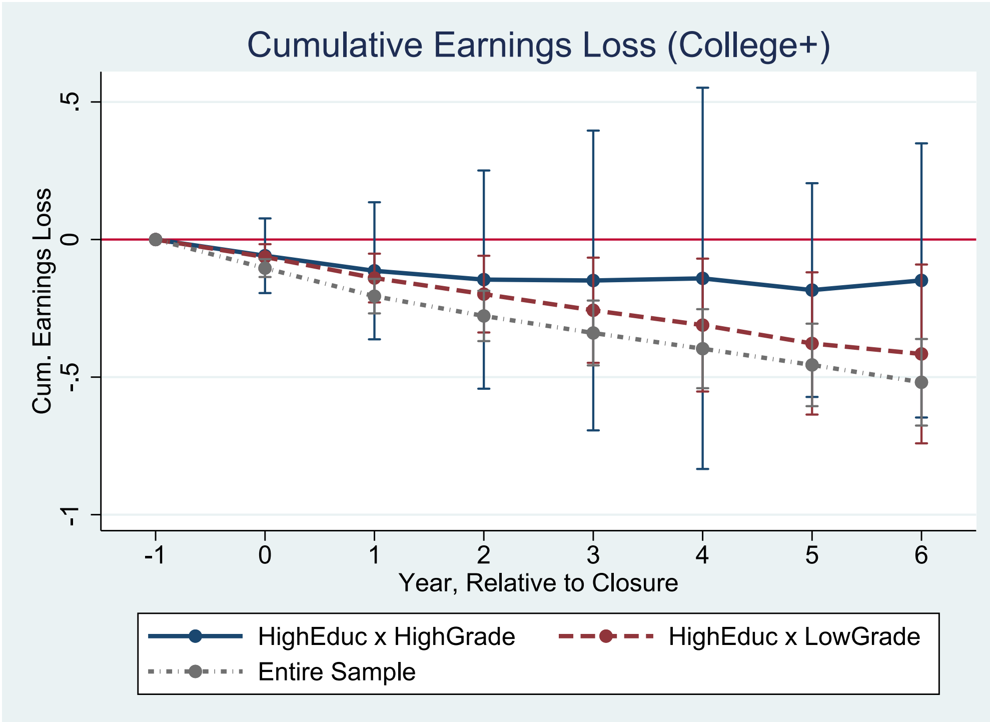
Appendix C: Further Differences-in-Differences Results

I. Additional Dependent Variables using Baseline Criteria

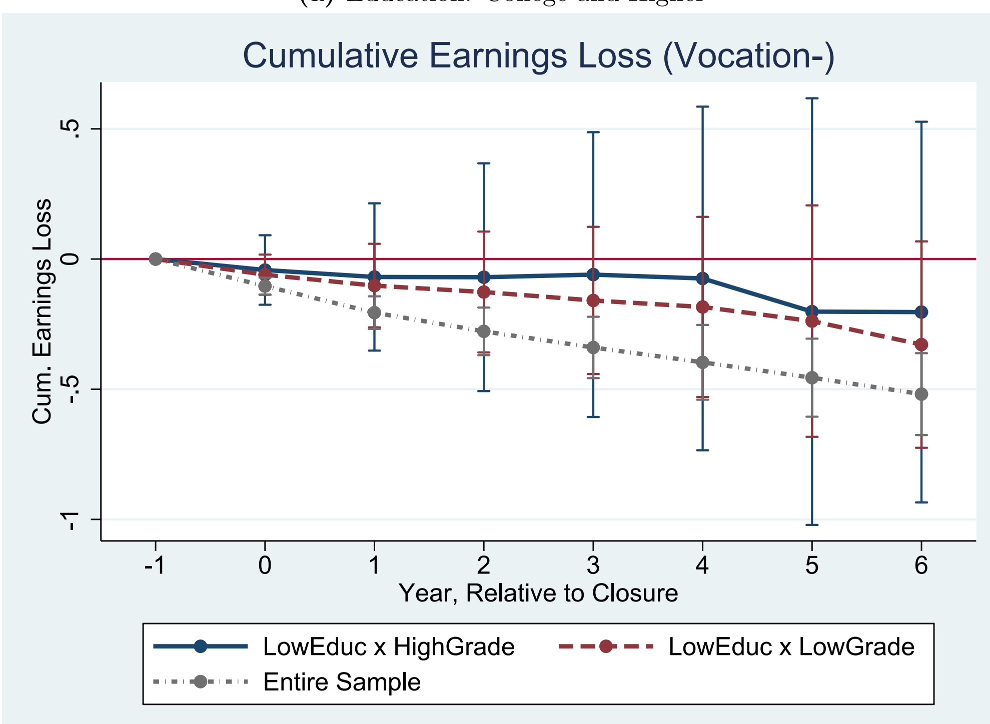
I present further differences-in-differences results. It is an extension of Section III.D, and estimated using Equation (3):

$$y_{it} = \alpha_i + \alpha_t + X'_{it}\Delta + \sum_{v=-5}^{6} \delta_v D_{it}^v + \sum_{v=-5}^{6} \delta_v^{Treat} D_{it}^v \text{ x } \mathbb{I}\{Treatment_i\} + \iota_{it} \tag{18}$$

where δ_v^{Treat} are the coefficients of interest, graphed below. The regression is estimated separately for each subsample (e.g. High Grade x Low Education).



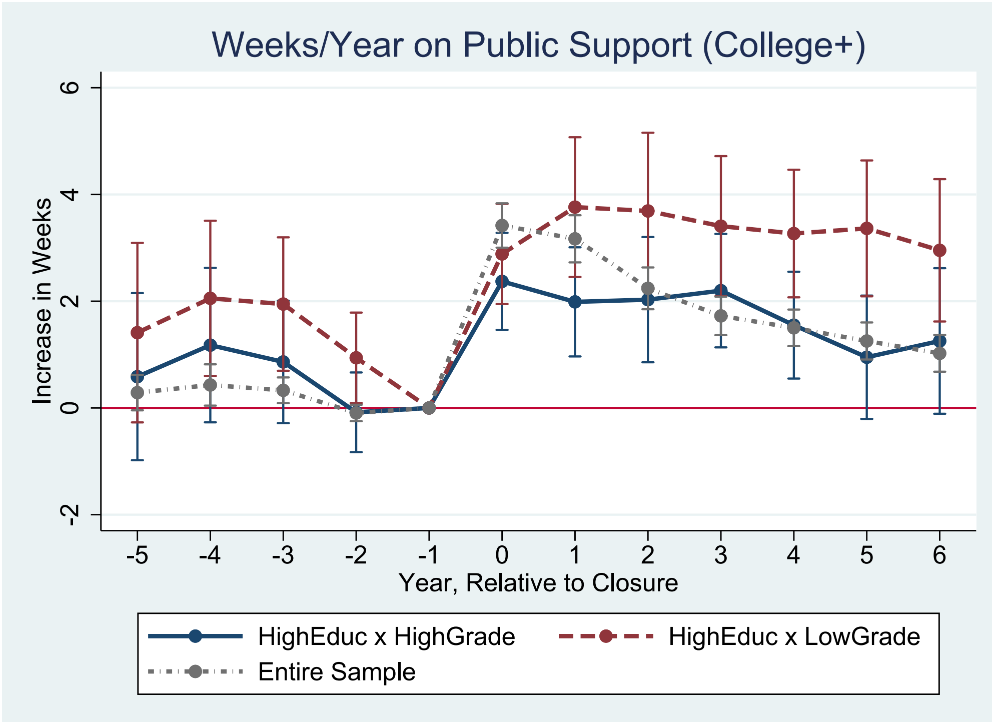
(a) Education: College and Higher



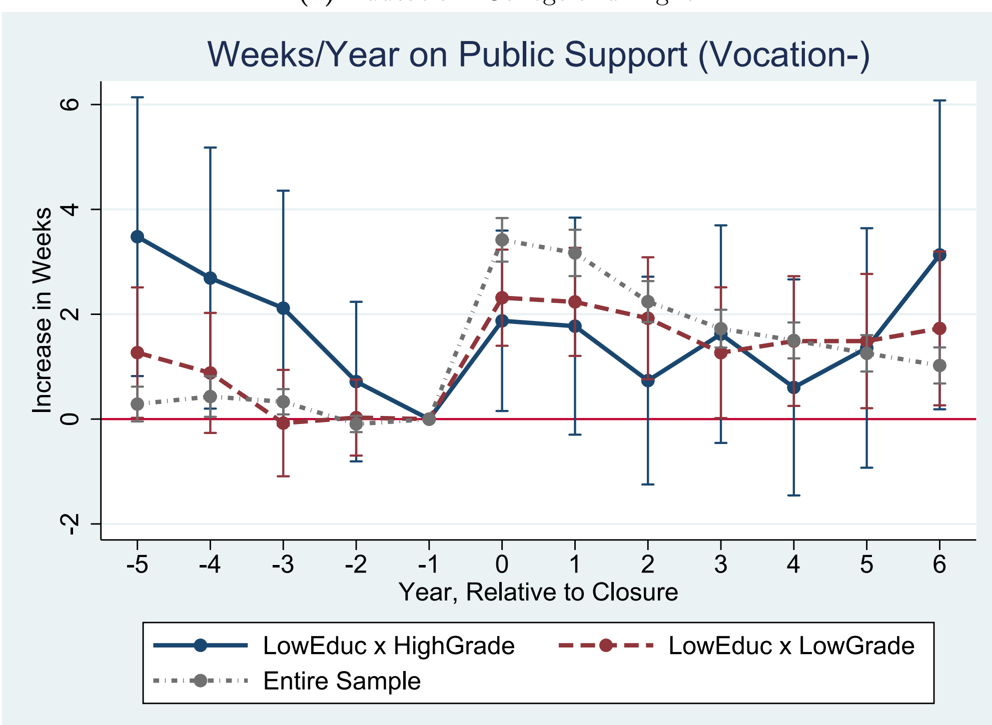
(b) Education: Vocational and Lower

Figure A.I

Cumulative Earnings Loss after Establishment Closure



(a) Education: College and Higher



(b) Education: Vocational and Lower

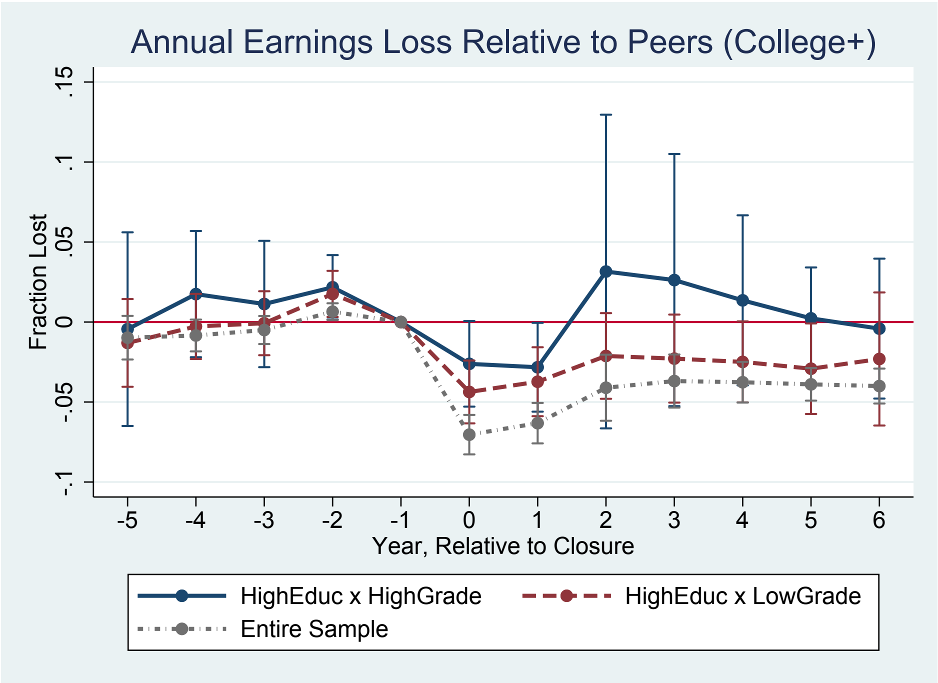
Figure A.II
Additional Government Support (in Weeks) after Establishment Closure

II. Differences-in-Differences using Relaxed Inclusion Criteria

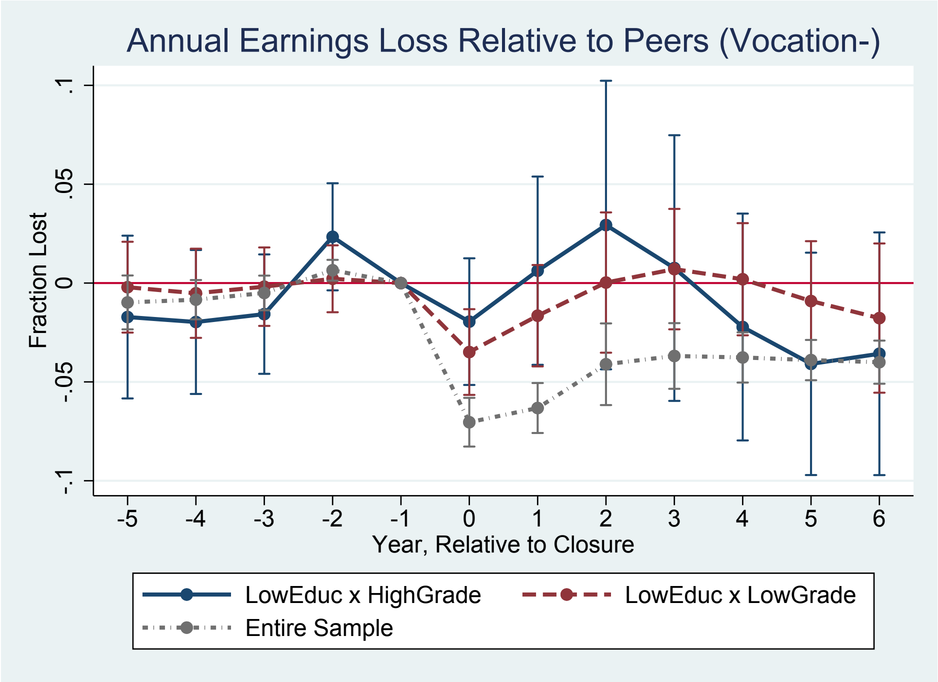
I present the same regressions, however using relaxed exclusion criteria. This increases sample size, as well as offering robustness checks. I first present the results with all relevant criteria relaxed, then results with only one relaxed criterion. The criteria under consideration here:

- **Establishment Size**: relaxed from minimum 5 FTEs to 3 FTEs.
- **Age**: relaxed from 25-55 to 25-60.
- **Tenure Requirement**: from 3 years minimum at the establishment to 1 year minimum.

II.1 Fully Relaxed Criteria



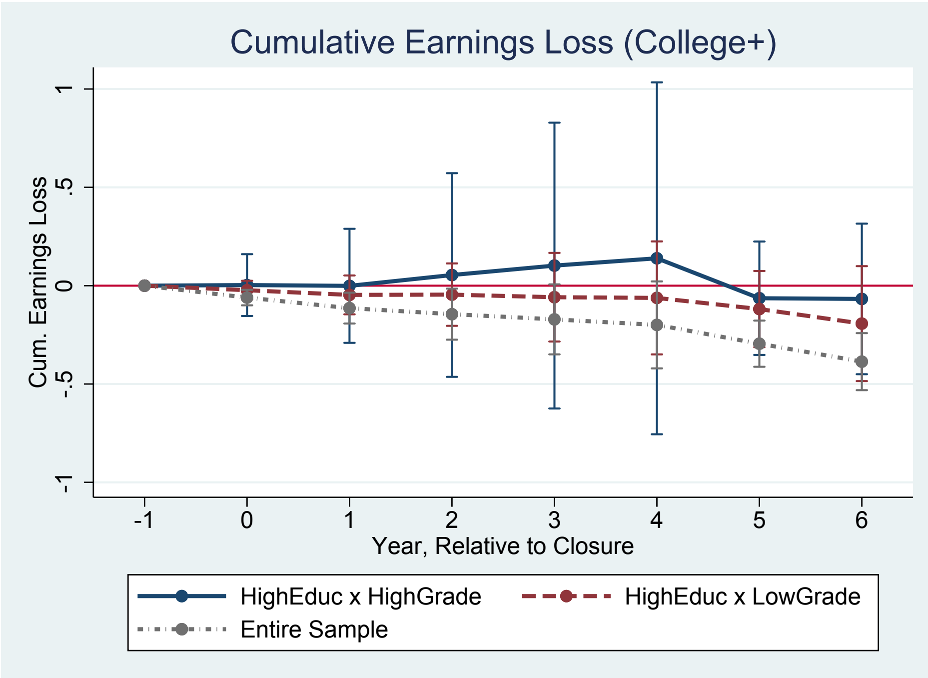
(a) Education: College and Higher



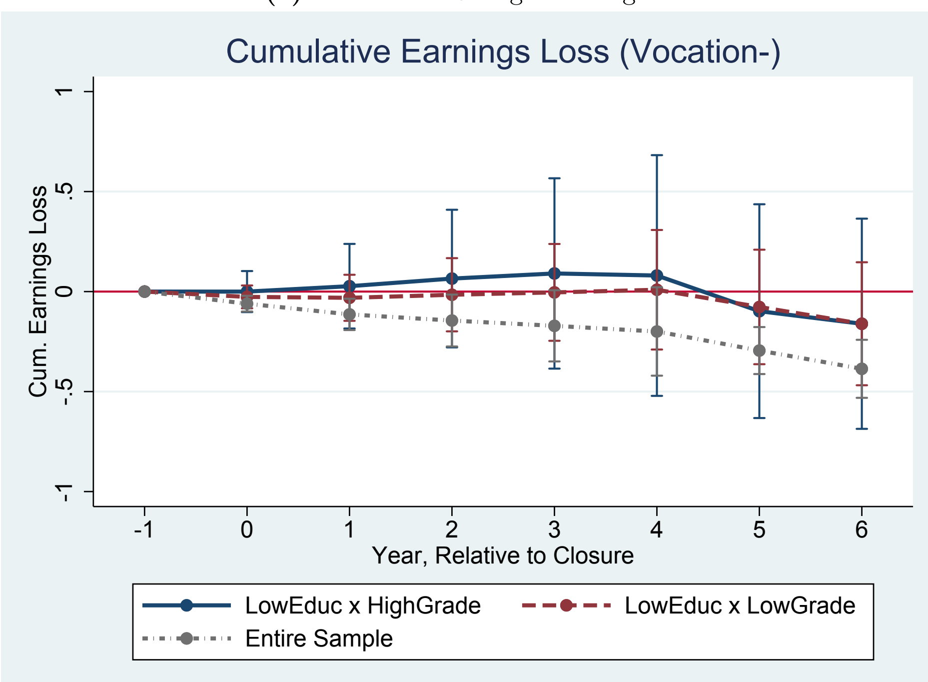
(b) Education: Vocational and Lower

Figure A.III

(Fully Relaxed) Earnings Loss after Establishment Closure in Levels



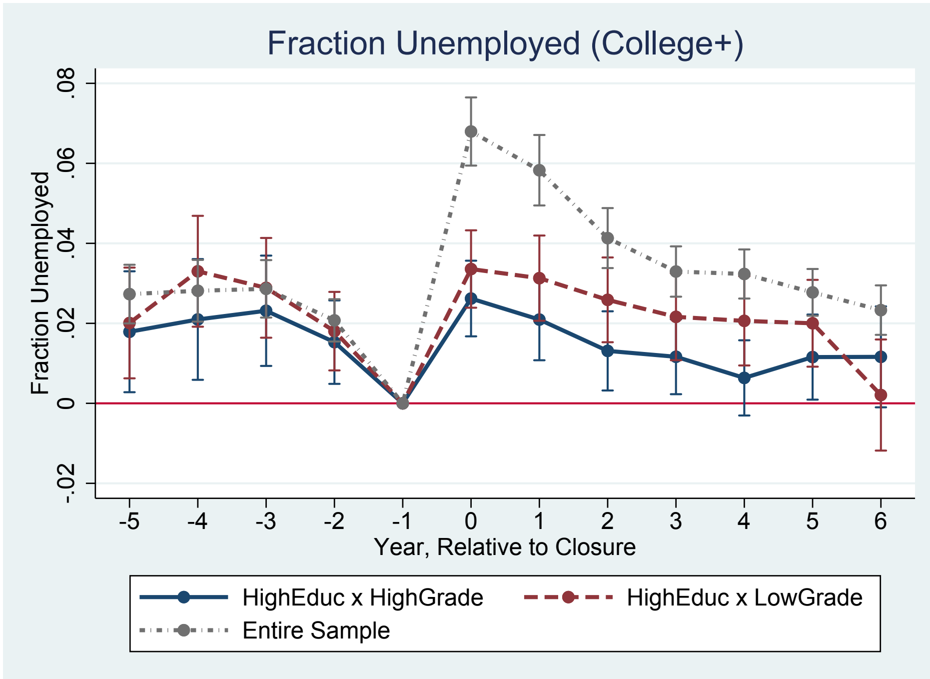
(a) Education: College and Higher



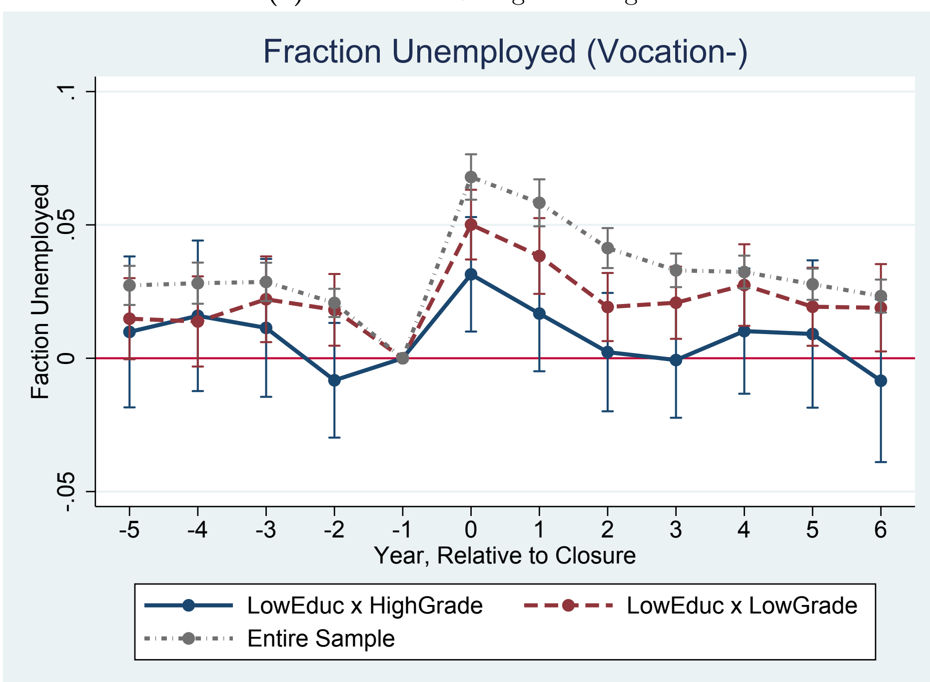
(b) Education: Vocational and Lower

Figure A.IV

(Fully Relaxed) Cumulative Earnings Loss after Establishment Closure



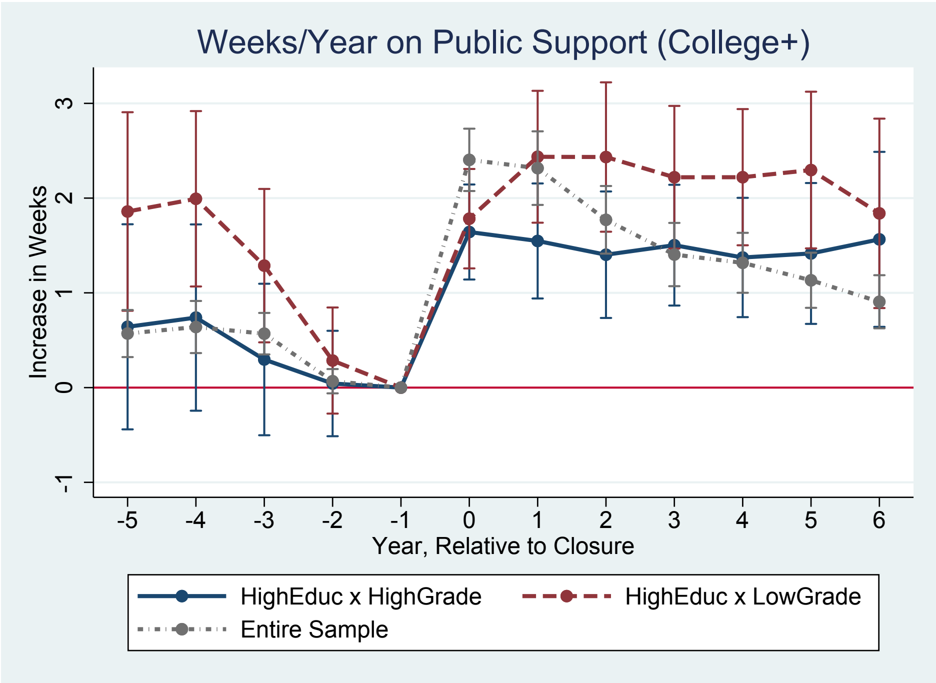
(a) Education: College and Higher



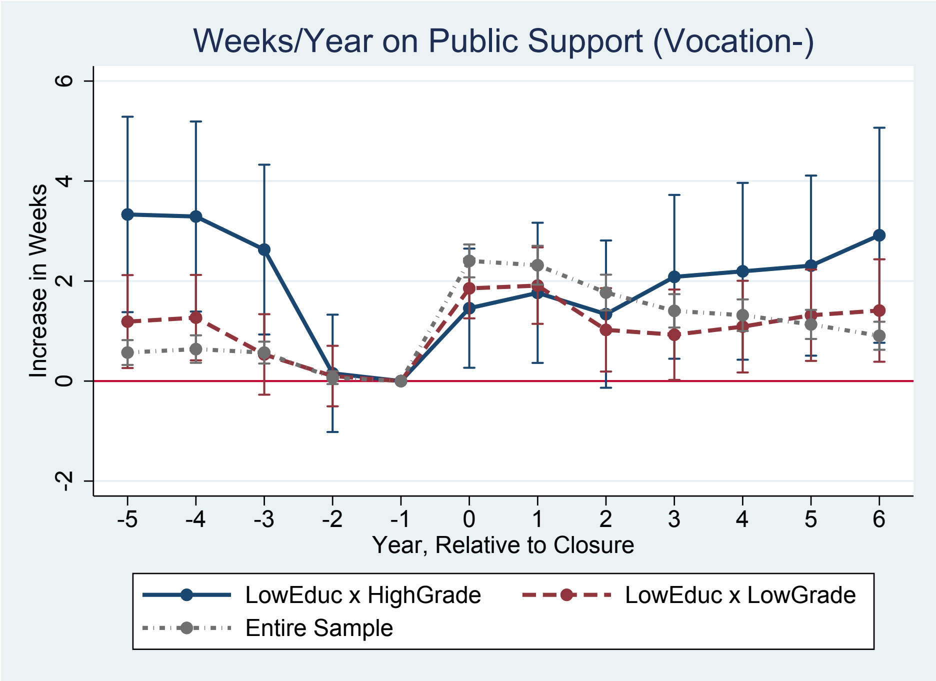
(b) Education: Vocational and Lower

Figure A.V

(Fully Relaxed) Unemployment after Establishment Closure



(a) Education: College and Higher

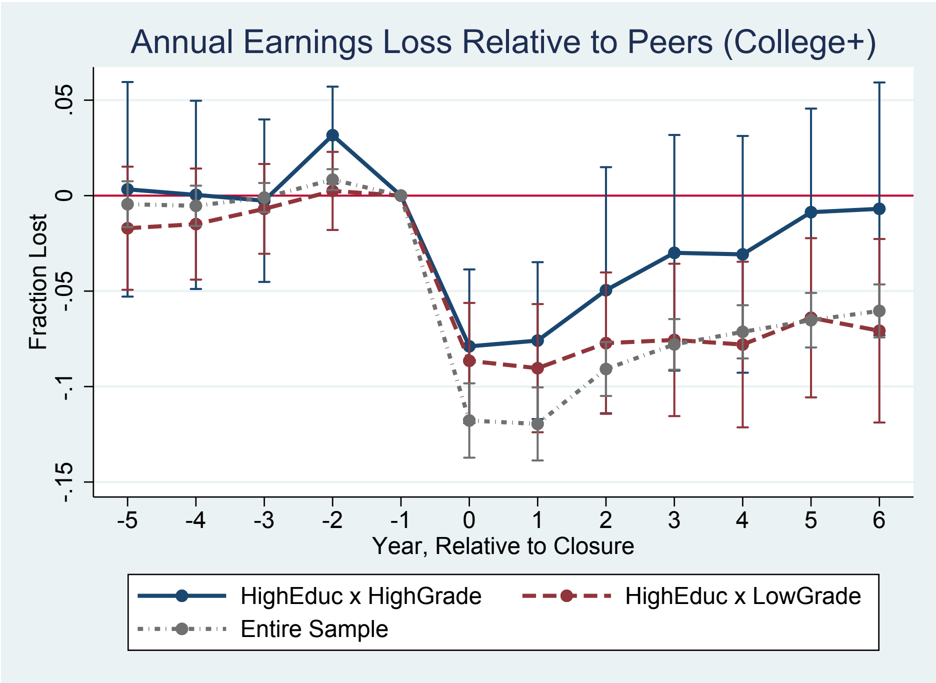


(b) Education: Vocational and Lower

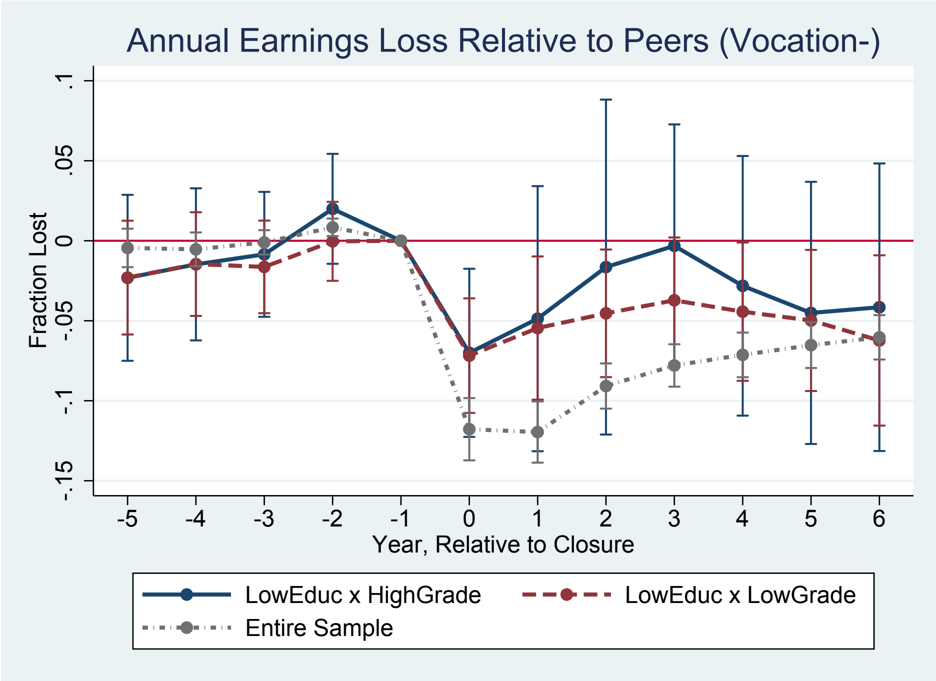
Figure A.VI

(Fully Relaxed) Additional Government Support (in Weeks) after Establishment Closure

II.2 Relaxed Age Criteria



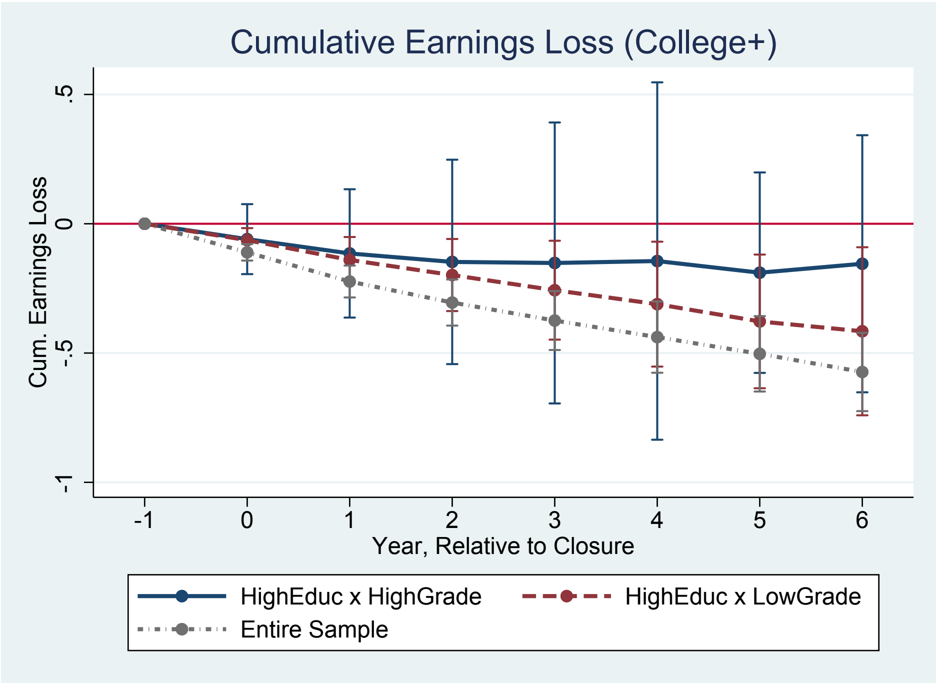
(a) Education: College and Higher



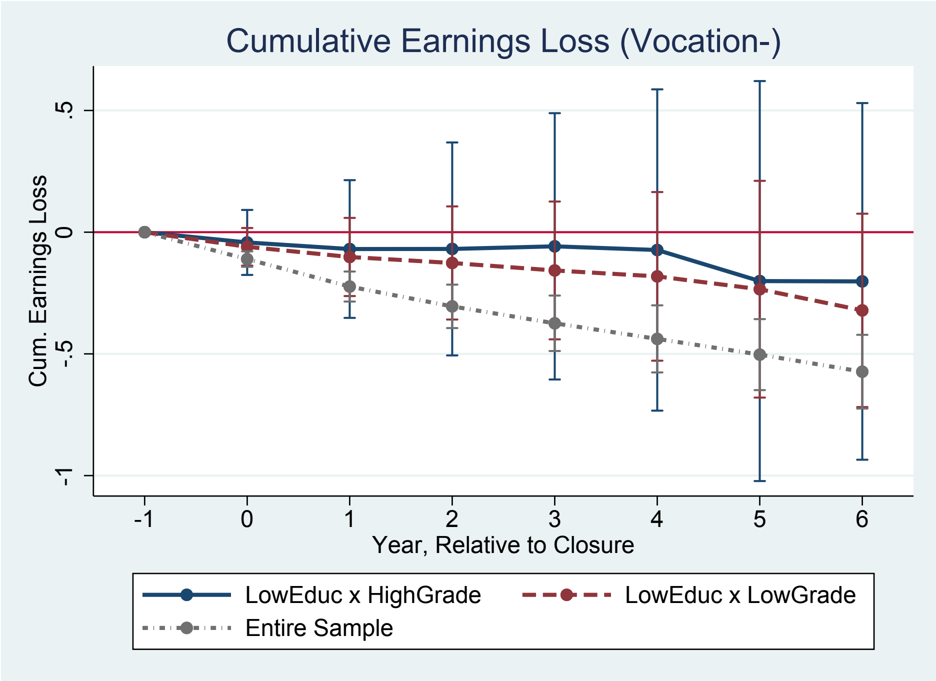
(b) Education: Vocational and Lower

Figure A.VII

(Age Relaxed) Earnings Loss after Establishment Closure in Levels



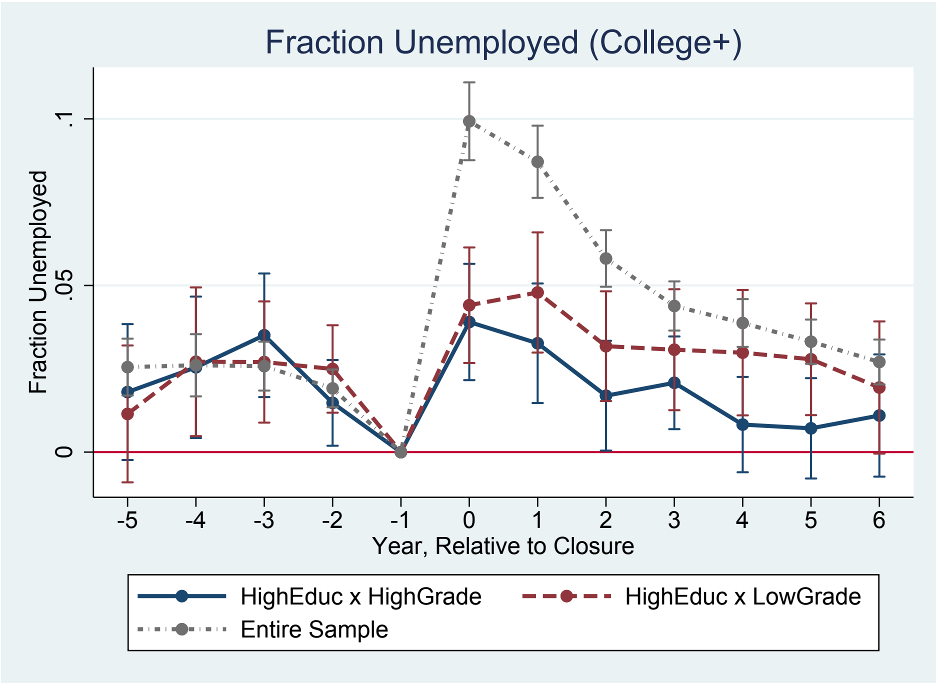
(a) Education: College and Higher



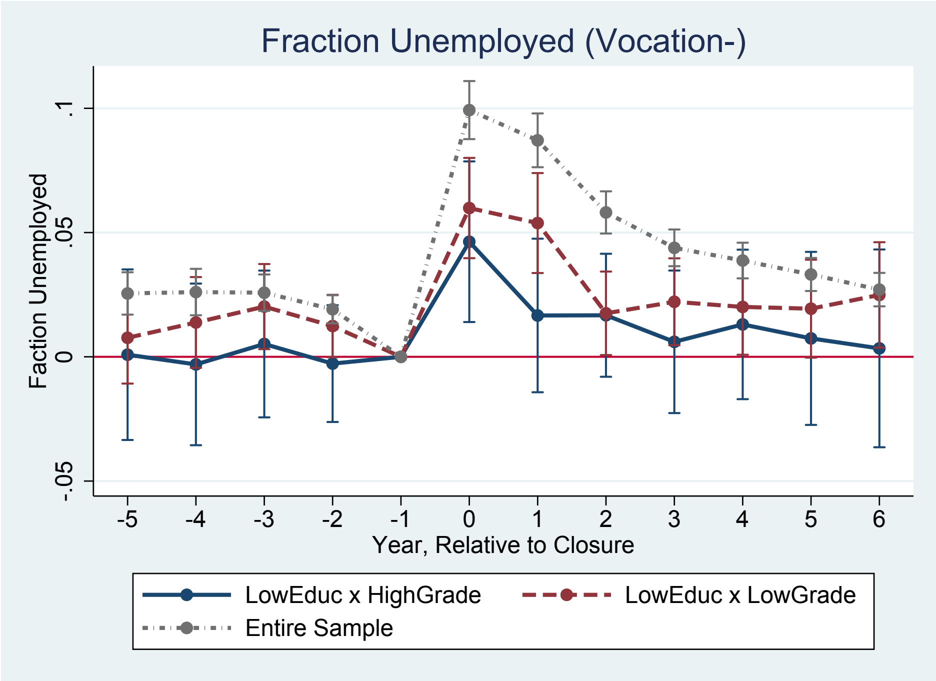
(b) Education: Vocational and Lower

Figure A.VIII

(Age Relaxed) Cumulative Earnings Loss after Establishment Closure

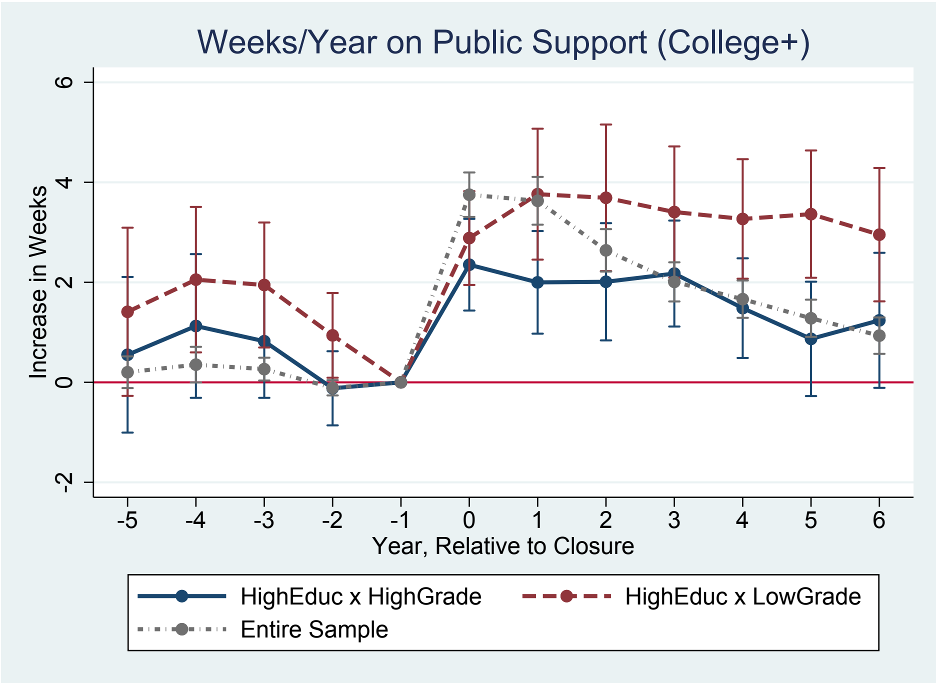


(a) Education: College and Higher

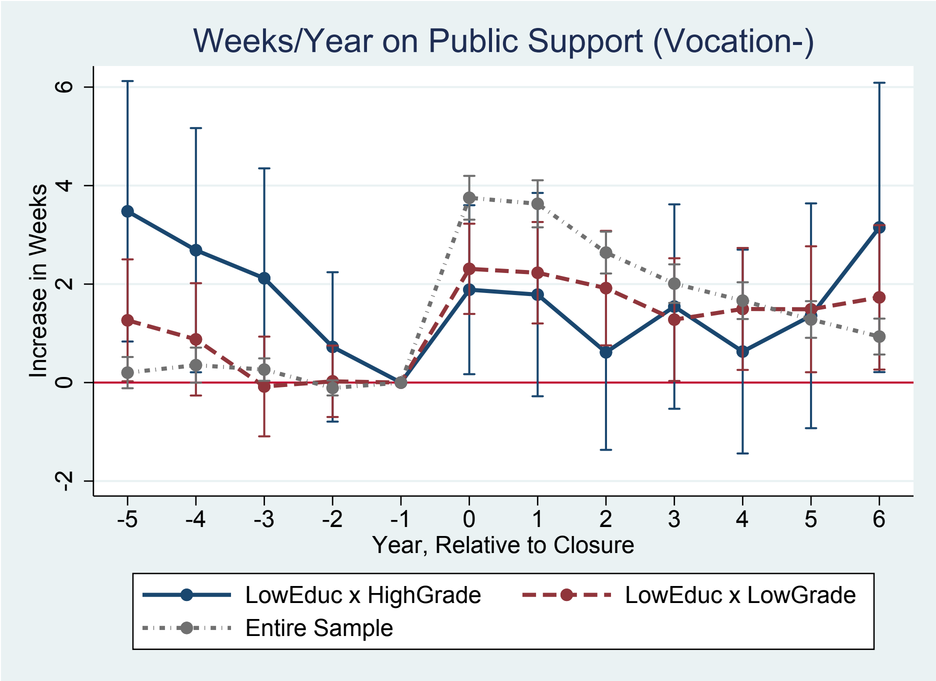


(b) Education: Vocational and Lower

Figure A.IX
(Age Relaxed) Unemployment after Establishment Closure



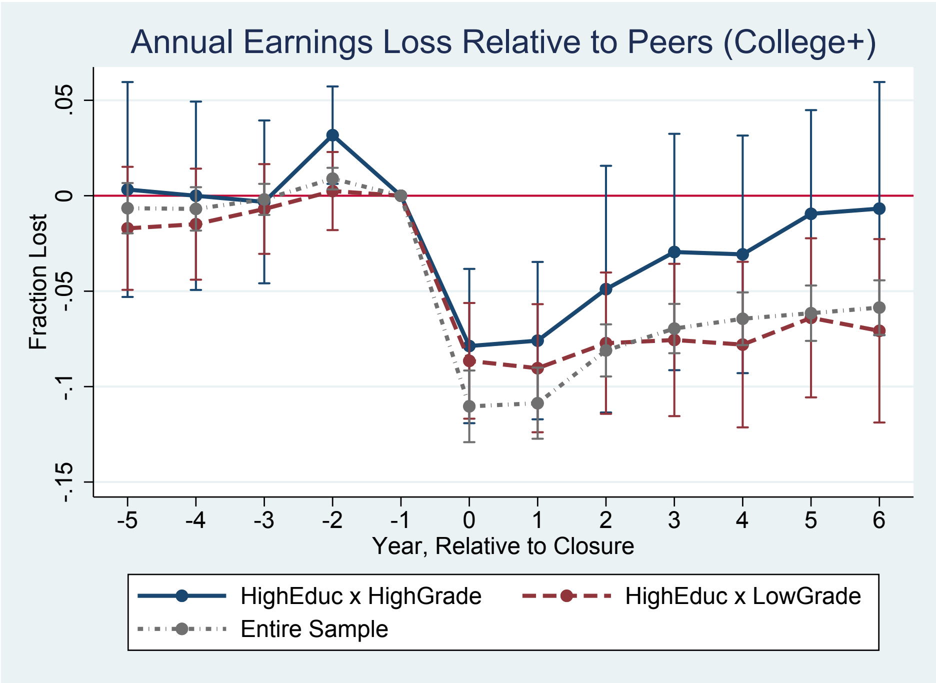
(a) Education: College and Higher



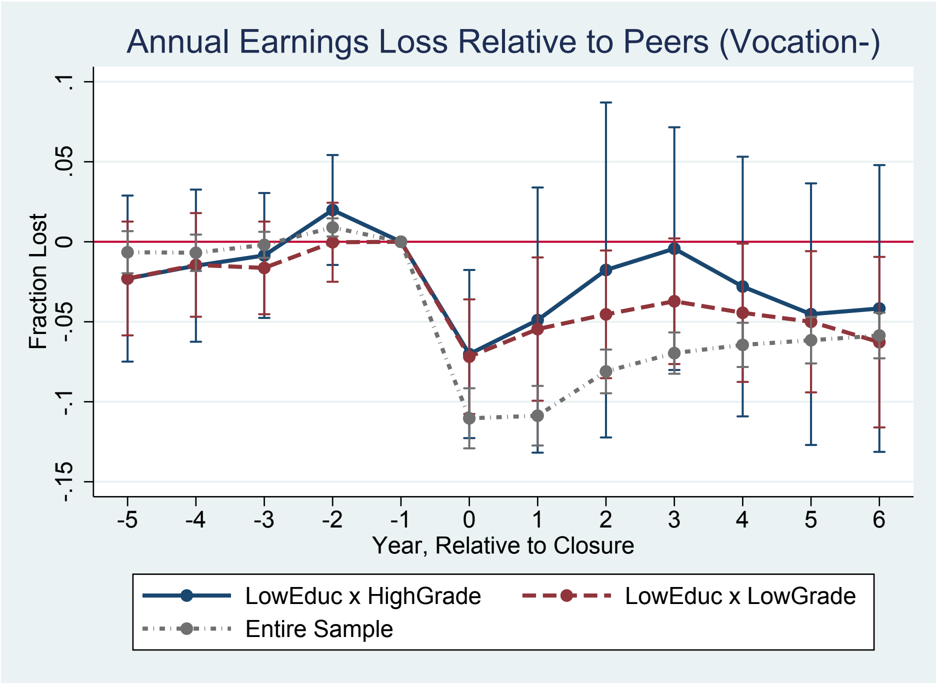
(b) Education: Vocational and Lower

Figure A.X
(Age Relaxed) Additional Government Support (in Weeks) after Establishment Closure

II.3 Relaxed Establishment Size Criteria



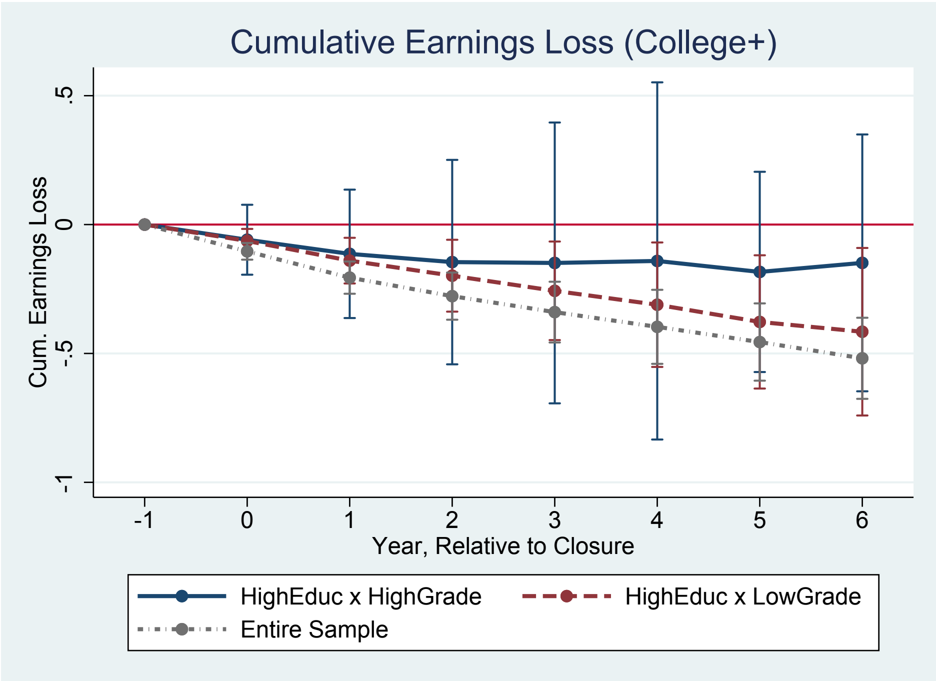
(a) Education: College and Higher



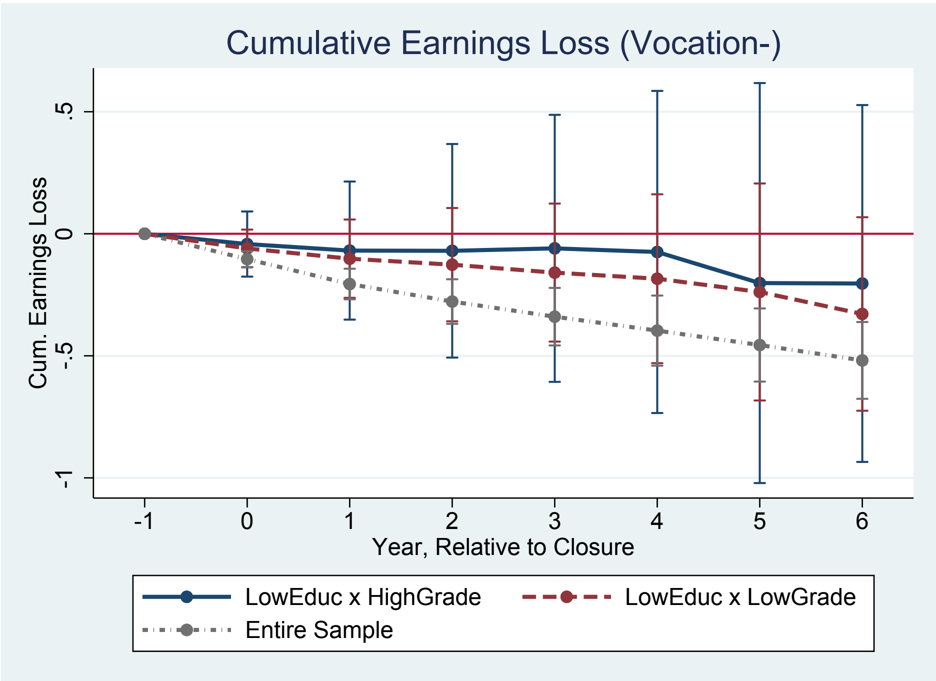
(b) Education: Vocational and Lower

Figure A.XI

(Estab. Size Relaxed) Earnings Loss after Establishment Closure in Levels



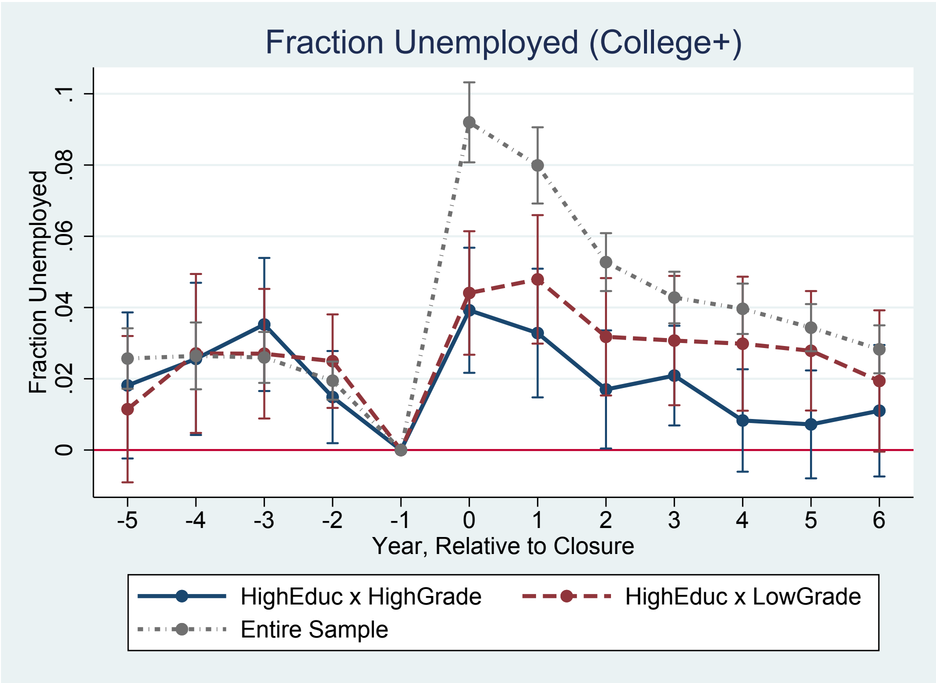
(a) Education: College and Higher



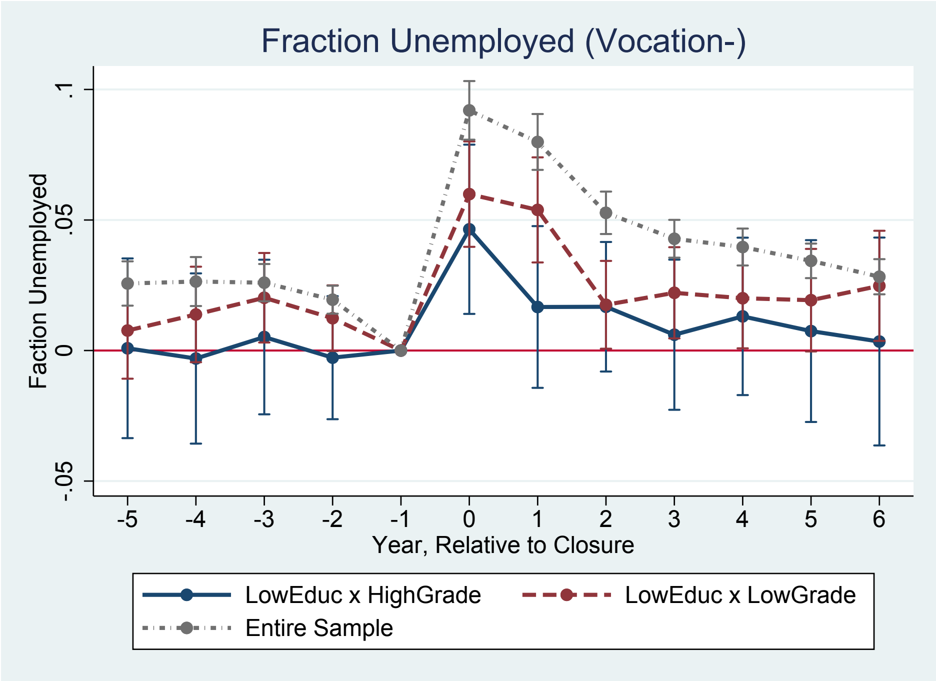
(b) Education: Vocational and Lower

Figure A.XII

(Estab. Size Relaxed) Cumulative Earnings Loss after Establishment Closure



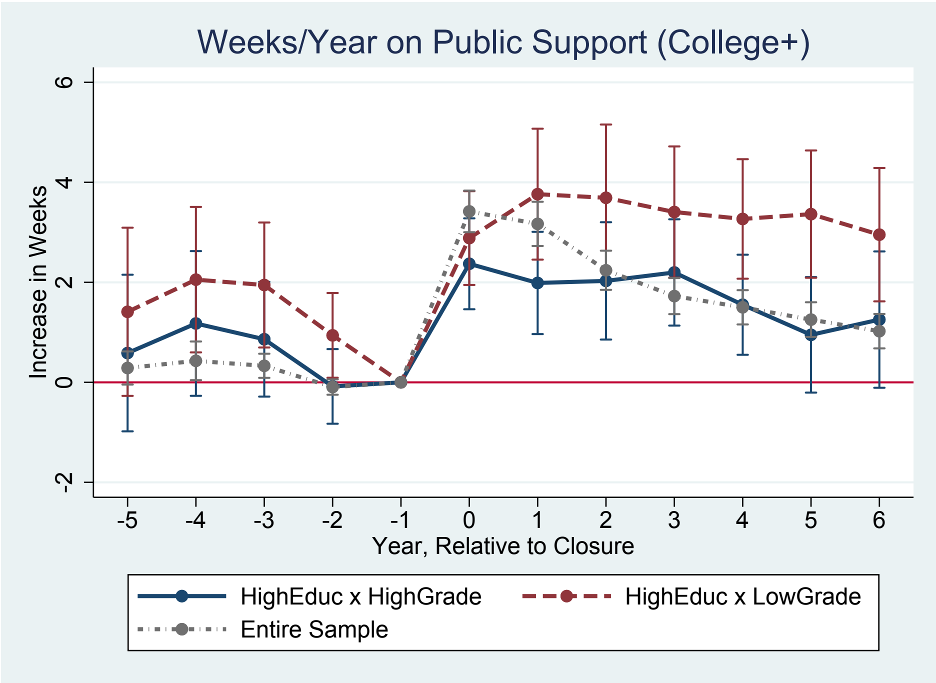
(a) Education: College and Higher



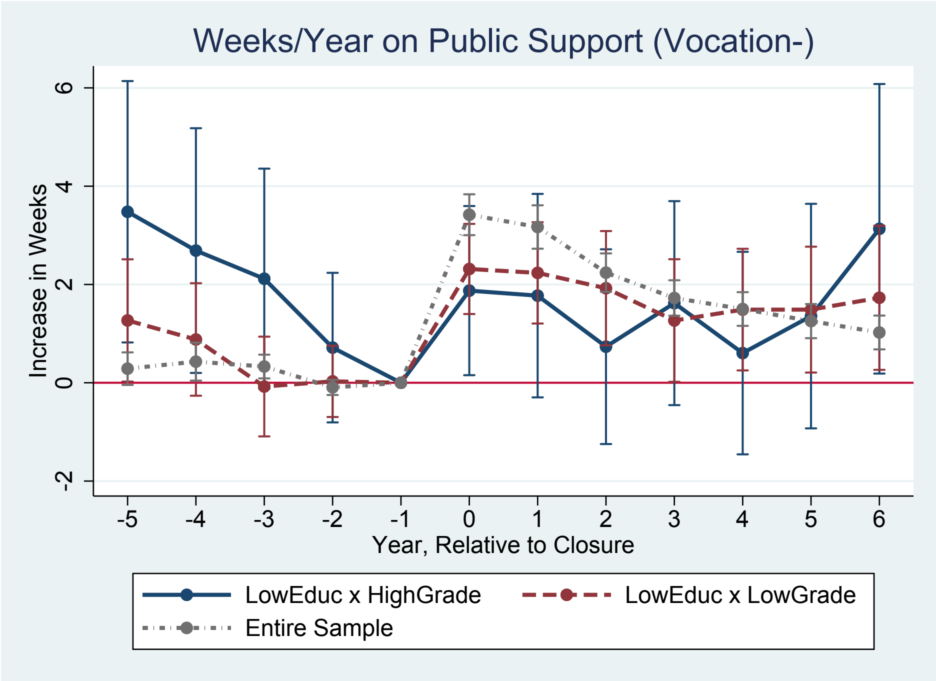
(b) Education: Vocational and Lower

Figure A.XIII

(Estab. Size Relaxed) Unemployment after Establishment Closure



(a) Education: College and Higher



(b) Education: Vocational and Lower

Figure A.XIV

(Estab. Size Relaxed) Additional Government Support (in Weeks) after Establishment Closure

Appendix D: Structural Estimation

I. *Hidden State Estimation*

This paper's model prominently features a hidden state. This hidden state is the vector of human capital (ξ), which evolves over time. The way I write this model, the hidden state and its "emission (yearly earnings) are a dynamic Bayesian network. Examples of dynamic Bayesian networks include Hidden Markov Models and Kalman Filter.

Consider the (k+2)-dimensional vector of human capital for worker i at time t:

$$\Xi_{it} = \{\xi^1_{it}, ..., \xi^k_{it}, Exp_{it}, G_i\}$$

For notational ease, define the law of motion for this part of the state as

$$\xi_{i,t+1} = f\left(xi_{it}, \mathcal{J}_{it}\right) + \eta_{it}$$

Recall that $\mathcal{J}_{it}$ is the observed chosen occupation. The first k entries of η_{it} are iid Gaussian errors, while the last two are non-stochastic and equal to 0.

In dynamic Bayesian terminology, the wage is the "emission" of the state, or the observation equation is

$$w_{it} = g(\Phi_{it}, \mathcal{J}_{it}) + \epsilon_{ijt}$$

Where ϵ_{ijt} is also iid Gaussian errors. Recall that $\xi_{it} \subset \Phi_{it}$.

Assumptions:

1. First-order Markov:

$$\Pr\{\xi_{it}|\xi_{i,1:t-1}, \mathcal{J}_{it}\} = \Pr\{\xi_{it}|\xi_{i,t-1}, \mathcal{J}_{it}\},$$

where $(1 : t-1) \equiv \{1, 2, ..., t-1\}$.

2. Wages are determined by this period's state and occupational choice:

$$\Pr\{w_{it}|w_{i,1:t-1}, \Phi_{i,1:t}, \mathcal{J}_{i,1:t}\} = \Pr\{w_{it}|\Phi_{it}, \mathcal{J}_{it}\}$$

Based on these assumptions, the likelihood function for a single worker (suppressing the i index here) is the joint probability

$$\mathcal{L} = Pr\{w_{1:T}, \Phi_{1:T}\} = Pr\{\Phi_1\}Pr\{w_1|\Phi_1, \mathcal{J}_1\}\prod_{t=2}^{T} Pr\{\Phi_t|\Phi_{t-1}, \mathcal{J}_{t-1}\}Pr\{w_t|\Phi_t\,\mathcal{J}_t\}$$

II. The Unscented Kalman Filter and Expectation-Maximization Algorithm

In this appendix section, I describe the Unscented Kalman Filter (UKF) as I use it in this paper, as well as the Expectations-Maximization (EM) algorithm used to estimate linear parameters.

With these combined algorithms, I jointly estimate the unobserved states (specifically, worker's skill vectors) and certain linear parameters of the model. The UKF estimates the hidden states of a particular variety of Bayesian dynamic networks. Specifically, the UKF is tailored towards non-linear Bayesian dynamic networks when the error terms are Gaussian. In contrast, the well known Kalman Filter addresses the same problem, however is only suited to linear models with Gaussian errors.

The standard Kalman Filter is unsuited to address non-linear models, since the Gaussian errors must be propagated through the non-linearities of the model. The equations of the

Kalman Filter are only suited to handle the propagation of the error term through a model which is jointly linear in the unobserved state and any covariates. The UKF addresses the non-linear propagation of the error term by selecting a small set of deliberately chosen points around the mean, then estimating the relevant covariances by passing them through the non-linear model. Similarly to the standard Kalman Filter, the researcher treats the model as fully known when running the UKF algorithm.

While the UKF treats the model as fully known, the reality is that I do not know the parameters of the model. Those parameters must be estimated jointly with the unobserved state. I accomplish the dual task by using the two-step EM algorithm. The two steps correspond with the two sets of unknowns to be estimated. The first step ("Expectation") estimates the values of the unobserved states, while treating the model as fully known. This first step is the Unscented Kalman Filter, combined with Unscented Rauch-Tung-Striebel Smoother. The second step ("Maximization") treats the unobserved states as known, and estimates the model's linear parameters by maximizing the log-likelihood, which I write explicitly below. The EM algorithm iterates over these two steps until both the unobserved state estimates and parameter estimates have both coverged to satisfy a pre-specified tolerance.

With that short motivation, I now present the algorithm in explicit detail.

First, start by restating the relevant segment of this paper's model: the Law of Motion for the unobserved worker skills, alongside the wage equation. Translated to traditional Bayesian dynamic networks vocabulary, these are the State Equation and Observation Equation, respectively. These are:

$$\xi_{i,t+1} = F(\xi_{it}, s_{ijt}) + u_{ijt} \tag{19}$$

$$= (\mathbb{I}_K - \Gamma_{\gamma_i})\xi_{i,t} + \Gamma_{\gamma_i} s_{ijt} + u_{it}$$

$$\log(w_{ijt}) = H(\xi_{it}, z_{it}) + \epsilon_{ijt} \tag{20}$$

$$\log(w_{ijt}) = \beta_0^i + \beta_1^{\mathcal{J}(i),t} + \beta_2' \vec{\Omega}_{it} + \beta_3 \left(\sum_{k=1}^{K} s_{ijt}^{k,(IM)} \max\left\{0, s_{ijt}^{k,(LV)} - \xi_{i,t}^k \right\} \right)$$
$$+ \beta_4 \left(\sum_{k=1}^{K} s_{ijt}^{k,(IM)} \max\left\{0, s_{ijt}^{k,(LV)} - \xi_{i,t}^k \right\}^2 \right) + \epsilon_{ijt}$$

Where $\mathbb{I}_K$ is the (KxK) identity matrix and the parameters to be estimated are in the diagonal matrix:

$$\Gamma_{\gamma_i} = \begin{pmatrix} \Gamma_{\gamma_i}^{(1)} & 0 & 0 \\ 0 & \Gamma_{\gamma_i}^{(2)} & 0 \\ 0 & 0 & \Gamma_{\gamma_i}^{(3)} \end{pmatrix}$$

I assume that $u_{ijt} \overset{\text{iid}}{\sim} N(0, \Sigma_u)$ and $\epsilon_{ijt} \overset{\text{iid}}{\sim} N(0, \sigma_\epsilon^2)$. Also, $\vec{\Omega}_{it}$ is a vector representing the third-degree polynomial terms of total labor market experience.

II.1 Expectations Step

The expectations step is the Unscented Kalman Filter, with the "smoothing" operation performed using the Unscented Rauch-Tung-Striebel Smoother. The iteration of the EM is indexed by d. The UKF proceeds as follows:

1. **Initialize** the unobserved state and its associated Mean Squared Error (MSE) estimates

(a) Initialize the unobserved state vector for each individual i:

$$\hat{\xi}^{(d)}_{i,0|0} = \begin{cases} \mathbb{E}\left[\hat{\xi}^{(0)}_{i,0}\right], & \text{for d} = 0 \\ \hat{\xi}^{(d-1)}_{i,0}, & \text{for d} > 0 \end{cases}$$

In the first iteration of the EM algorithm, I simply set all skills to be equal, however they are scaled up or down according to the individual's highest education achieved. In subsequent iterations, the initial vector is the smoothed estimate from the previous iteration.

(b) Initialize the MSE of the unobserved state vector for each individual i:

$$P^{(d)}_{i,0|0} = \begin{cases} \mathbb{E}\left[(\xi_{i,0} - \hat{\xi}^{(d)}_{i,0})(\xi_{i,0} - \hat{\xi}^{(d)}_{i,0})'\right], & \text{for d} = 0 \\ P^{(d-1)}_{i0}, & \text{for d} > 0 \end{cases}$$

In practice, I initialize the MSE in the first iteration with a diagonal matrix, entries of 0.5.

After this initialization step, for every worker i, and for each t = 1,2,...,T_i:

2. **Calculate the matrix of sigma points** around the state estimate:

$$\mathcal{X}^{(d)}_{i,t-1} = \begin{bmatrix} \hat{\xi}^{(d)}_{i,t-1} & \hat{\xi}^{(d)}_{i,t-1} + \aleph\sqrt{P^{(d)}_{i,t-1}} & \hat{\xi}^{(d)}_{i,t-1} - \aleph\sqrt{P^{(d)}_{i,t-1}} \end{bmatrix}$$

Where $\aleph$ is a constant, which I set to 0.0003 throughout (following the guidance of Wan and van der Merwe (2002)) and $\sqrt{P^{(d)}_{i,t-1}}$ represents the lower Cholesky factorization of $P^{(d)}_{i,t-1}$.

3. **Project** the state, outcome (i.e. salary), and MSE in the next period.

(a) Project the a sigma matrix of states, calculated using the columns of the sigma points from step 2:

$$\mathcal{X}^{(d)}_{i,t|t-1} = F\left(\mathcal{X}^{(d)}_{i,t-1}, s_{ijt}\right)$$

(b) Project the state based on a weighted average of the above state sigma matrix:

$$\hat{\xi}^{(d)}_{i,t|t-1} = \sum_{q=0}^{2K} W^m_q \mathcal{X}^{(d),(q)}_{i,t|t-1}$$

where $\mathcal{X}^{(d),(q)}_{i,t|t-1}$ is the q^{th} column of the sigma matrix of states, and W^m_q is a constant weight, which I specify below.

(c) Project the MSE of the state in the next period, again using a weighted average of the sigma matrix of the states' variance:

$$P^{(d)}_{i,t|t-1} = \sum_{q=0}^{2K} W^c_q \left[\mathcal{X}^{(d),(q)}_{i,t|t-1} - \hat{\xi}^{(d)}_{i,t|t-1}\right]\left[\mathcal{X}^{(d),(q)}_{i,t|t-1} - \hat{\xi}^{(d)}_{i,t|t-1}\right]' + \Sigma_u$$

Again, W^c_q is a constant weight, which I specify below.

(d) Project the sigma matrix of the outcome (i.e. salary):

$$\mathcal{Y}^{(d)}_{i,t|t-1} = H\left(\mathcal{X}^{(d)}_{i,t|t-1}, z_{it}\right)$$

(e) Project the outcome based on a weighted average of the above outcome sigma matrix:

$$\hat{y}^{(d)}_{i,t|t-1} = \sum_{q=0}^{2K} W^m_q \mathcal{Y}^{(d),(q)}_{i,t|t-1}$$

as above, $\mathcal{Y}^{(d),(q)}_{i,t|t-1}$ is the q^{th} column of the sigma matrix of outcomes.

4. **Filter** the state estimate and covariance estimates.

(a) Estimate the MSE of the outcome estimate.

$$P^{i,(d)}_{y_t y_t} = \sum_{q=0}^{2K} W^c_q \left[\mathcal{Y}^{(d),(q)}_{i,t|t-1} - \hat{y}^{(d)}_{i,t|t-1}\right]\left[\mathcal{Y}^{(d),(q)}_{i,t|t-1} - \hat{y}^{(d)}_{i,t|t-1}\right]' + \sigma^2_\epsilon$$

(b) Estimate the covariance of the state and the outcome:

$$P^{i,(d)}_{\xi_t y_t} = \sum_{q=0}^{2K} W^c_q \left[\mathcal{X}^{(d),(q)}_{i,t|t-1} - \hat{\xi}^{(d)}_{i,t|t-1}\right]\left[\mathcal{Y}^{(d),(q)}_{i,t|t-1} - \hat{y}^{(d)}_{i,t|t-1}\right]'$$

(c) Estimate the MSE of both non-linear terms in the observation equation which are functions of the state. Compute these for t-1, and add an extra step at the end of the algorithm to compute these for T_i (alongside the requisite sigma points).

$$P^{(d)}_{M_1,it} = \sum_{q=0}^{2K} W^c_q \left[M_1\left(\mathcal{X}^{(d)}_{i,t-1}, s_{ijt}) - \bar{M}_1(\mathcal{X}^{(d)}_{i,t-1}, s_{ijt}\right)\right]\left[M_1\left(\mathcal{X}^{(d)}_{i,t-1}, s_{ijt}) - \bar{M}_1(\mathcal{X}^{(d)}_{i,t-1}, s_{ijt}\right)\right]'$$

$$P^{(d)}_{M_2,it} = \sum_{q=0}^{2K} W^c_q \left[M_2\left(\mathcal{X}^{(d)}_{i,t-1}, s_{ijt}) - \bar{M}_2(\mathcal{X}^{(d)}_{i,t-1}, s_{ijt}\right)\right]\left[M_2\left(\mathcal{X}^{(d)}_{i,t-1}, s_{ijt}\right) - \bar{M}_2\left(\mathcal{X}^{(d)}_{i,t-1}, s_{ijt}\right)\right]'$$

$$P^{(d)}_{M_1 M_2,it} = \sum_{q=0}^{2K} W^c_q \left[M_1\left(\mathcal{X}^{(d)}_{i,t-1}, s_{ijt}) - \bar{M}_1(\mathcal{X}^{(d)}_{i,t-1}, s_{ijt}\right)\right]\left[M_2\left(\mathcal{X}^{(d)}_{i,t-1}, s_{ijt}\right) - \bar{M}_2\left(\mathcal{X}^{(d)}_{i,t-1}, s_{ijt}\right)\right]'$$

(d) Compute the Kalman Gain:

$$\kappa^{(q)}_{it} = P^{i,(d)}_{\xi_t y_t}\left(P^{i,(d)}_{y_t y_t}\right)^{-1}$$

(e) Compute the filtered state for time t:

$$\hat{\xi}^{(q)}_{i,t|t} = \hat{\xi}^{(q)}_{i,t|t-1} + \kappa^{(q)}_{it}\left(\log(w_{ijt}) - \hat{y}^{(d)}_{i,t|t-1}\right)$$

(f) Compute the filtered MSE of the state for time t:

$$P^{(d)}_{i,t|t} = P^{(d)}_{i,t|t-1} - \kappa^{(q)}_{it} P^{i,(d)}_{y_t y_t}\left(\kappa^{(q)}_{it}\right)'$$

This fully describes the Kalman Filter algorithm used in the estimation routine. The next step is to smooth these estimates using some sort of smoother.

As noted in the main body of the text, the Unscented Kalman Smoother that I employ is called the Unscented Rauch-Tung-Striebel Smoother (see, e.g. Särkkä (2008)). This version of the Unscented Smoother is analogous to the Kalman Smoother standard in the Macroeconomic literature, e.g. in Hamilton (1994). This is in contrast to the Unscented Smoother that is standard to the engineering literature, which is a linear combination of two separately conducted Unscented Kalman Filters: the first run forwards in time, the second backwards.

This Unscented Kalman Smoother proceeds as follows, again separately for each worker. There are three critical outputs of this Smoother, denoted as follows:

1. $\hat{\xi}^{(d)}_{i,t}$, the smoothed estimate of ξ_{it} in iteration d of the EM algorithm.
2. $\hat{P}^{(d)}_{it} \equiv \mathbb{E}\left[\left(\xi_{it} - \hat{\xi}^{(d)}_{it}\right)\left(\xi_{it} - \hat{\xi}^{(d)}_{it}\right)'\right]$, the smoothed estimate of the state's MSE.
3. $\hat{P}^{(d)}_{i,[t,t-1]} \equiv \mathbb{E}\left[\left(\xi_{i,t-1} - \hat{\xi}^{(d)}_{i,t-1}\right)\left(\xi_{it} - \hat{\xi}^{(d)}_{it}\right)'\right]$, the smoothed estimate of the period-to-period cross-covariance of the state.

The Smoother is initialized with the time T_i filtered estimates used directly as the time T_i smoothed estimates: $\hat{\xi}^{(d)}_{iT} = \xi^{(q)}_{i,T|T}$, and $\hat{P}^{(d)}_{iT} = P^{(d)}_{i,T|T}$. With these initial values, I start the smoother at $t = (T_i - 1)$, and proceed recursively to $t = 0$.

1. Calculate the cross-covariance term using filtered values from time (t-1) and projected values from time t:

$$\hat{P}^{(d)}_{i,[t+1,t]} = \sum_{q=0}^{2K} W^c_q \left[\mathcal{X}^{(d),(q)}_{i,t|t} - \hat{\xi}^{(d)}_{i,t|t}\right]\left[\mathcal{X}^{(d),(q)}_{i,t+1|t} - \hat{\xi}^{(d)}_{i,t+1|t}\right]'$$

2. Calculate the smoother gain:

$$D_{i,t}^{(q)} = \hat{P}_{i,[t+1,t]}^{(d)} \left(P_{i,t+1|t}^{(d)} \right)^{-1}$$

3. Calculate the smoothed state estimate:

$$\hat{\xi}_{i,t}^{(d)} = \hat{\xi}_{i,t|t}^{(d)} + D_{i,t}^{(q)} \left(\hat{\xi}_{i,t+1}^{(d)} - \hat{\xi}_{i,t+1|t}^{(d)} \right)$$

4. Calculate the smoothed estimate for the state MSE:

$$\hat{P}_{it}^{(d)} = P_{i,t|t}^{(d)} + D_{i,t}^{(q)} \left(\hat{P}_{i,t+1}^{(d)} - \hat{P}_{i,t+1|t}^{(d)} \right) \left(D_{i,t}^{(q)} \right)'$$

Which fully specifies the smoothing algorithm used.

II.2 Maximization Step

The maximization step estimates the unknown linear parameters. It estimates those parameters by treating the unobserved state variables as if they were observed – using the state estimates obtained in the Expectations step. To be specific, the parameters to be estimated are

$$\Theta_{EM} = \begin{bmatrix} \beta & \Gamma \end{bmatrix}$$

By rearranging the model in equations (19) and (20), we gain a useful pair of linear expression, to be estimated:

$$u_{ij,t+1} = (\xi_{i,t+1} - \xi_{it}) - \Gamma_{\gamma_i}(s_{ijt} - \xi_{it})$$
$$\epsilon_{ijt} = \log(w_{ijt}) - \beta X_{ijt}$$

where $\beta \equiv \begin{bmatrix} \beta_0^i & \beta_1^{\mathcal{J}(i),t} & \vec{\beta}_2 & \beta_3 & \beta_4 \end{bmatrix}$ and
$X_{ijt} \equiv \begin{bmatrix} 1 & 1 & \vec{\Omega}_{it} & \left(\sum_{k=1}^K s_{ijt}^{k,(IM)} \max\left\{0, s_{ijt}^{k,(LV)} - \xi_{i,t}^k\right\}\right) & \left(\sum_{k=1}^K s_{ijt}^{k,(IM)} \max\left\{0, s_{ijt}^{k,(LV)} - \xi_{i,t}^k\right\}^2\right) \end{bmatrix}'$.

Using these convenient expressions, defining $\Delta\xi_{i,t+1} \equiv \xi_{i,t+1} - \xi_{it}$, and recalling that both error terms are Gaussian, I can write the joint log-likelihood for a given individual as

$$\begin{aligned} log(\mathcal{L}_{\xi,w}(\Theta_{EM})) = -\frac{1}{2}\Bigg[\sum_{i=1}^N \Bigg(& \sum_{t=1}^{T_i} \left[u'_{ijt}\Sigma_u^{-1}u_{ijt}\right] + \left[\epsilon'_{ijt}\left(\sigma^2\right)^{-1}\epsilon_{ijt}\right] \\ & + T_i\left(\log|\sigma^2| + \log|\Sigma_u|\right) \\ & + [\pi_{i0} - \xi_{i0}]'\left(\Sigma_u^0\right)^{-1}[\pi_{i0} - \xi_{i0}] + \log|\Sigma_u^0|\Bigg) + (T_i(K+1)+1)\log(2\pi)\Bigg] \end{aligned} \tag{21}$$

$$\begin{aligned} log(\mathcal{L}_{\xi,w}(\Theta_{EM})) = -\frac{1}{2}\Bigg[\sum_{i=1}^N \Bigg(& \sum_{t=1}^{T_i} \left[[\Delta\xi_{it} - \Gamma_{\gamma_i}(s_{ij,t-1} - \xi_{i,t-1})]'\,\Sigma_u^{-1}\,[\Delta\xi_{it} - \Gamma_{\gamma_i}(s_{ij,t-1} - \xi_{i,t-1})]\right] \\ & + \sum_{t=1}^{T_i}\left[[\log(w_{ijt}) - \beta X_{ijt}]'\left(\sigma^2\right)^{-1}[\log(w_{ijt}) - \beta X_{ijt}]\right] \\ & + T_i\left(\log|\sigma^2| + \log|\Sigma_u|\right) + \\ & [\pi_{i0} - \xi_{i0}]'\left(\Sigma_u^0\right)^{-1}[\pi_{i0} - \xi_{i0}] + \log|\Sigma_u^0|\Bigg) + T_i(K+1)\log(2\pi)\Bigg] \end{aligned} \tag{22}$$

In practice, I do not make any inference from the initial distribution of the state (i.e. $\xi_{i,0}$).

This log likelihood is convenient, because maximization boils down to a linear regression. In practice, I estimate these parameters by regression with fixed effects at the individual and occupation-year levels. Importantly, note that the likelihood is written using the *true* state, whereas I use the estimated state. I cannot simply plug the estimated state in for the true state, since I must account for the MSE of the state estimate. If the state estimate is $\hat{\omega}_{it}$, then I can write

$$\xi_{it} = \hat{\xi}_{it} + (\xi_{it} - \hat{\xi}_{it})$$

also,

$$\xi_{it}\xi_{it}' = \underbrace{\hat{\xi}_{it}\hat{\xi}_{it}' + \underbrace{(\xi_{it} - \hat{\xi}_{it})(\xi_{it} - \hat{\xi}_{it})'}_{\equiv P_{it}}}_{\equiv V_i^t}$$

where the second term is the MSE of the state estimate (its estimate is denoted by $\hat{P}_{it}$) – an integral part of the Kalman Filter and Smoother. The estimator $\hat{V}_i^t$ is exactly the term above, with the estimator for P_{it} substituted in. However, also note that this MSE is uncorrelated with the observation or covariates, so I only have to worry about the extra term on $\xi_{it}\xi_{it}'$.

Given this, I can take the derivative of this log likelihood with respect to each of the unknown parameters, $\begin{bmatrix} \beta & \Gamma & \sigma_\epsilon^2 & \Sigma_u \end{bmatrix}$. As alluded to above, I make no attempt at inference for π_{i0} or Σ_u.

$$\frac{\partial log(\mathcal{L}_{\xi,w}(\Theta_{EM}))}{\partial \beta} = \sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(\sigma_\epsilon^{-1}\tilde{\hat{X}}_{ijt}'\left(\log(w_{ijt}) - \beta\tilde{\hat{X}}_{ijt}\right)\right)$$

$$\hat{\beta}^{(d)} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T_i}\tilde{\hat{X}}_{ijt}'\tilde{\hat{X}}_{ijt}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T_i}\tilde{\hat{X}}_{ijt}'\log(w_{ijt})\right) \quad (23)$$

where I define

$$\hat{\hat{V}}_i^t \equiv \begin{pmatrix} x_{it}x_{it}' & x_{it}\hat{\omega}_{it}' \\ \hat{\omega}_{it}x_{it}' & \hat{V}_i^t \end{pmatrix}$$

$$\frac{\partial log(\mathcal{L}_{\xi,w}(\Theta_{EM}))}{\partial R^{-1}} = -\frac{1}{2}\left(\sum_{i=1}^{N}\left[-T_iR' + \sum_{t=1}^{T_i}[w_{it} - \tilde{A}'\tilde{x_{it}}][w_{it} - \tilde{A}'\tilde{x_{it}}]'\right]\right)$$

$$R^* = \sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(\frac{1}{T_i}\left([w_{it} - \tilde{A}'\hat{\hat{x}}_{it}][w_{it} - \tilde{A}'\hat{\hat{x}}_{it}]' + H'\hat{P}_{it}H\right)\right) \tag{24}$$

$$\frac{\partial log(\mathcal{L}_{\xi,w}(\Theta_{EM}))}{\partial \tilde{F}} = \sum_{i=1}^{N}\sum_{t=2}^{T_i}\left(Q^{-1}[\omega_{it} - \tilde{F}\tilde{z}_{i,t-1}]\tilde{z}_{i,t-1}'\right)$$

$$\tilde{F}^* = \left(\sum_{i=1}^{N}\sum_{t=2}^{T_i}\hat{\hat{V}}_i^{t,t-1}\right)\left(\sum_{i=1}^{N}\sum_{t=2}^{T_i}\hat{\hat{V}}_i^{t-1,t-1}\right)^{-1} \tag{25}$$

where

$$\hat{\hat{V}}_i^{t,t-1} \equiv \begin{pmatrix} \hat{\omega}_{it}z_{i,t-1}' & \hat{V}_i^{t,t-1} \end{pmatrix}$$

$$\hat{\hat{V}}_i^{t-1,t-1} \equiv \begin{pmatrix} z_{i,t-1}z_{i,t-1}' & z_{i,t-1}\hat{\omega}_{i,t-1}' \\ \hat{\omega}_{i,t-1}z_{i,t-1}' & \hat{V}_i^{t-1} \end{pmatrix}$$

where $\hat{V}_i^{t,t-1}$ and $\hat{V}_i^{t-1}$ come from the Kalman smoother. They are estimates of the terms $\omega_{it}\omega_{i,t-1}'$ and $\omega_{i,t-1}\omega_{i,t-1}'$, respectively, accounting for the MSE of the state estimates.

$$\frac{\partial log(\mathcal{L}_{\xi,w}(\Theta_{EM}))}{\partial Q^{-1}} = -\frac{1}{2}\left(\sum_{i=1}^{N}\left[-(T_i-1)Q' + \sum_{t=2}^{T_i}[\omega_{it} - \tilde{F}\tilde{z}_{i,t-1}][\omega_{it} - \tilde{F}\tilde{z}_{i,t-1}]'\right]\right)$$

$$Q^* = \sum_{i=1}^{N}\sum_{t=2}^{T_i}\left(\frac{1}{(T_i-1)}\left(\hat{V}_{it} - \hat{\hat{V}}_i^{t,t-1}\left(\hat{\hat{V}}_i^{t-1,t-1}\right)^{-1}\left(\hat{\hat{V}}_i^{t,t-1}\right)'\right)\right) \tag{26}$$