

Reiremo, Tore Adam

Working Paper

The anatomy of mismatch in the college market

Memorandum, No. 02/2024

Provided in Cooperation with:

Department of Economics, University of Oslo

Suggested Citation: Reiremo, Tore Adam (2024) : The anatomy of mismatch in the college market, Memorandum, No. 02/2024, University of Oslo, Department of Economics, Oslo

This Version is available at:

<https://hdl.handle.net/10419/303262>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

MEMORANDUM

No 02/2024
September 2024

The Anatomy of Mismatch in the College Market

The seal of the University of Oslo is a circular emblem. It features a central figure of a woman in classical attire, holding a lyre. The text "UNIVERSITAS OSLOENSIS" is inscribed around the top half of the circle, and "MDCCCXXXIII" is at the bottom. The seal is rendered in a light gray tone.

Tore Adam Reiremo

ISSN: 0809-8832

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: <http://www.sv.uio.no/econ>
e-mail: econdep@econ.uio.no

In co-operation with
**The Frisch Centre for Economic
Research**

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: <http://www.frisch.uio.no>
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 01/2024	Gunnar Bårdsen and Ragnar Nymoen U.S. wage-price dynamics, before, during and after COVID-19, through the lens of an empirical econometric model
No 04/2023	Andreas Bjerre-Nielsen, Lykke Sterll Christensen, Mikkel Høst Gandil and Hans Henrik Sievertsen Playing the system: address manipulation and access to schools
No 03/2023	Gunnar Bårdsen and Ragnar Nymoen Dynamic time series modelling and forecasting of COVID-19 in Norway
No 02/2023	Hans A. Holter, Serhiy Stepanchuk and Yicheng Wang Taxation and Entrepreneurship in the United States
No 01/2023	Pedro Brinca, João B. Duarte, Hans A. Holter and João G. Oliveira Technological Change and Earnings Inequality in the U.S.: Implications for Optimal Taxation
No 05/2022	Bård Harstad A theory of pledge-and-review bargaining
No 04/2022	Michael Graber, Magne Mogstad, Gaute Torsvik and Ola Vestad Behavioural responses to income taxation in Norway
No 03/2022	Eskil Heinesen, Christian Hvid, Lars Kirkebøen, Edwin Leuven and Magne Mogstad Instrumental variables with unordered treatments: Theory and evidence from returns to fields of study
No 02/2022	Mikkel Gandil and Edwin Leuven College admission as a screening and sorting device
No 01/2022	Askill H. Halse, Karen E. Hauge, Elisabeth T. Isaksen, Bjørn G. Johansen and Oddbjørn Raaum Local Incentives and Electric Vehicle Adoption

Previous issues of the memo-series are available in a PDF® format
at: <http://www.sv.uio.no/econ/english/research/unpublished-works/working-papers/>

The Anatomy of Mismatch in the College Market*

Tore Adam Reiremo[†]

FEBRUARY 2024

Abstract

Both the availability of college education and the demand for college skills have expanded rapidly over the past decades in most countries. However, it remains an open question whether the increasing number of college-educated workers are able to successfully match with jobs where college skills are rewarded. Using rich Norwegian administrative data, I provide evidence on the evolution of mismatch in the college market. I further decompose the overall college mismatch into (i) a structural mismatch that captures misalignment between underlying demand and supply, (ii) a geographical component, and (iii) a within-location assignment component. Over the past two decades, overall mismatch in the college market has remained relatively stable, affecting about 20% of workers, despite large shifts in both the demand and supply of college workers. I further show that within-location assignment mismatch can account for more than two-thirds of the overall mismatch and its relative contribution has remained relatively unchanged. This evidence points to the vital role of reassignment policies that improve matching in the college market.

*The Norwegian Research Council supported this research under project no. 275906.

[†]Department of Economics, University of Oslo. Email: t.a.reiremo@econ.uio.no

1 Introduction

Access to college education has expanded rapidly over the past decades. Among the OECD countries, the average share of workers with a college education rose from 12 to 39% between 1996 and 2020 (OECD, 1996, 2021).¹ At the same time, college wage premiums rose sharply through the 1990s (Autor et al., 2020), and remain above 30% in a majority of OECD countries (OECD, 2021).

The success of college expansion policies relies on the ability of college-educated workers to match with jobs where college skills are rewarded. Mismatch in the college market has adverse consequences for individual earnings as well as productivity and overall efficiency. This motivates my main research question: *How prevalent is mismatch in the college market?* Furthermore, identifying effective policies to reduce mismatch requires knowledge about the underlying sources of mismatch. Firstly, one has to know *what portion of mismatch is due to a misalignment between underlying demand and supply of college workers?* And secondly, *what portion of mismatch can be explained by misalignment between the location of college workers and jobs?* While structural misalignment points towards policies that aim to equalize supply and demand for college workers, geographical mismatch may suggest that policies that aim to increase geographic mobility are more effective. The final source of mismatch is a within-location misalignment of college workers and jobs, which points to policies that improve matching in the labor market.

To quantify mismatch in the college market, I start my analysis by classifying college jobs. I classify college jobs as jobs in occupations that hold relatively high returns to a college degree. Next, I measure mismatch in the college market as the number of college workers in non-college jobs and non-college workers in college jobs. My approach extends the one by Gottschalk and Hansen (2003) by allowing for unobserved individual heterogeneity and addressing endogenous selection into occupations. The availability of rich Norwegian data sources allows for estimation of college premiums that change over time, while still preserving the granularity of the four-digit classification of occupations.

¹For Norway, this share increases from 16 to 45%, while for the US and the UK, the share increases from 24 to 50% and 12 to 50%, respectively.

To decompose overall mismatch into three distinct components, I develop a novel accounting framework that captures the part of mismatch that can be explained by misalignment between supply and demand for college workers in the aggregate economy or a corresponding misalignment between supply and demand within geographical regions. To achieve this decomposition, I propose a counterfactual reassignment of college workers to college jobs, where the overall supply and demand of college jobs are taken as given, either in the aggregate economy or within each geographical region. This reassignment procedure can be viewed as an analytical tool that allows the researcher to identify the lowest attainable mismatch given the supply and demand of college jobs. The decomposition I propose has policy relevance as it provides clear implications for the effectiveness of alternative policies aimed at reducing the overall mismatch in the college market.

Applying these methods to Norwegian administrative data, I estimate that educational mismatch affects around 20% among all workers between 2003 and 2018, as well as similar and stable levels of mismatch among college and non-college workers. Despite a rise in the share of college workers from 30 to around 40%, I find little contribution from misalignment between overall supply and demand for college workers. This result is driven by large positive shifts in college premiums across a wide set of occupations, suggesting a rising demand for college skills within these occupations. Additionally, I find that college workers and college jobs are largely concentrated in the same areas, and almost no contribution to mismatch from geographical misalignment between college workers and jobs. The within-location assignment component can account for more than two-thirds of the overall mismatch, which is also relatively unchanged over time. This evidence points to the vital role of reassignment policies that improve matching in the college market.

This paper makes several contributions to the literature. Firstly, the rich Norwegian data records allow me to control for unobserved heterogeneity and non-random selection into occupations by controlling for individual fixed effects. Comparing these results to OLS regression used in [Gottschalk and Hansen \(2003\)](#) suggests that OLS estimates have a positive bias that increases with college premiums. Adding controls for ability test scores and father's income to the OLS regressions alleviates a large portion of the bias,

suggesting that non-random selection into occupations drives the difference between my specification and OLS regression and that this bias is likely to affect mismatch measures. Secondly, studying the contribution of misalignment in overall supply and demand forces and the geographical location of workers and jobs brings new evidence of the sources of mismatch.

Educational mismatch has a long history in labor economics. Consequences of mismatch were first studied in a seminal paper by [Duncan and Hoffman \(1981\)](#), which led to a large collection of papers commonly referred to as the "overeducation"-literature. More recently, [Gottschalk and Hansen \(2003\)](#) documents falling rates of mismatch among college-educated workers in the US between 1982 and 1995. Two papers apply their method to UK data with somewhat mixed results: [Grazier et al. \(2008\)](#) finds uniformly falling rates of mismatch among college workers between 1994 and 2004, while [O'Leary and Sloane \(2016\)](#) finds mismatch rates growing from 2001 to 2006 and falling from 2006 to 2010. This paper improves on the estimation procedure by controlling for unobserved individual heterogeneity and extends the analysis by decomposing mismatch into three different components.

Several papers have studied related notions of mismatch and their implications for policy. [Lise and Postel-Vinay \(2020\)](#) investigates mismatch in cognitive, manual, and interpersonal skills. They find that cognitive skills yield high returns, but adjust slowly through on-the-job learning, illustrating the importance of alleviating mismatch in jobs requiring cognitive skills. [Fredriksson et al. \(2018\)](#) studies mismatch along a large set of cognitive and non-cognitive skills. They find that mismatched workers tend to improve match quality by searching for new jobs, and that mismatch is most prevalent among young workers. Their results shows that policies that improve matching in the labor market are especially important for young workers. [Liu et al. \(2016\)](#) show that a significant portion of the long-lasting wage penalty associated with graduating in a recession can be attributed to a mismatch between field of study and early career jobs. This illustrates that recent graduates are negatively affected by misalignment between supply and demand forces early in their career and the importance of match quality in the college market.

Compared to their paper, I abstract away from fields of study and focus on more structural sources of mismatch.

The role of the geographical location of workers and jobs is a growing literature pointing to policies that aim to increase geographical mobility. [Marinescu and Rathelot \(2018\)](#) finds that US job seekers are 35% less likely to apply for jobs 10 miles (16 kilometers) away from home, implying a strong preference for working close to home and a reluctance to move for a job. This result points to geographical misalignment as a potential source of mismatch in the college market.

The remainder of this paper is organized as follows: Section 2 presents a conceptual framework used to classify jobs as college or non-college, as well as the decomposition of mismatch into structural and assignment mismatch. Section 3 discusses identification of college premiums and presents the estimation procedure. The various data sources are described in section 4 together with descriptive statistics. In section 5 I present the main results, including the estimated premiums, observed mismatch, and the decomposition of mismatch into three components. Section 6 concludes.

2 Conceptual Framework

This section outlines a conceptual framework building on [Gottschalk and Hansen \(2003\)](#). The framework is used to motivate the classification of occupations to college and non-college occupations, and give a framework for discussing the identification of college premiums (see Section 3)

The framework abstracts away from several potentially important sources of mismatch, such as search and information frictions. The model treats education as exogenous and allows workers to self-select into jobs by considering wages and non-pay attributes. The model abstracts away from friction and therefore predicts an efficient allocation of workers to jobs.

2.1 Demand for College Workers

In this section I outline a simple model of labor demand building on [Gottschalk and Hansen \(2003\)](#). The model assumes that college and non-college workers are perfect substitutes, exogenously fixing college premiums. This part of the model motivates the process of classifying occupations into college and non-college occupations using estimated college premiums.

Start by considering an economy with J employers in a single period t . Employers represent groups of uniform jobs, and can be substituted with aggregate industries (as in [Liu et al., 2016](#)), sectors, or occupations. Each employer is endowed with a production function, characterizing the production technology²:

$$Q^j = F^j(K^j, L_0^j + \alpha^j L_1^j), \quad j = \{1, \dots, J\}$$

α^j is the efficiency units of college workers relative to non-college workers in each sector j , and represents skill-biased technological change. Label these occupations from lowest to highest α , and normalize the total number of workers to unity:

$$\alpha^1 \leq \alpha^2 \leq \dots \leq \alpha^{J-1} \leq \alpha^J$$

$$\sum_{j=1}^J (L_0^j + L_1^j) = 1$$

Firms are price takers and maximize profits by choosing capital and employment. Because college and non-college workers are perfect substitutes, the equilibrium wage of college workers relative to non-college workers is exogenously given by

$$\frac{W_1^j}{W_0^j} = \alpha^j, \quad j = 1, \dots, J$$

For notational convenience, let ω^j denote the log wage of non-college workers and δ^j the

²Time subscripts are suppressed for clarity.

difference in log wage between college and non-college workers in job j :

$$\omega^j = \ln(W_0^j)$$

$$\delta^j = \ln(W_1^j) - \ln(W_0^j) = \ln(\alpha^j) \approx \frac{W_1^j - W_0^j}{W_0^j}$$

Assume that there is a premium α^* that perfectly balances the costs and returns to college education, and label every job with $\alpha_t^j \geq \alpha^*$ as college jobs.

2.2 Job choice

This section outlines a model of job choice incorporating occupational preferences and earnings, extending the model of [Gottschalk and Hansen \(2003\)](#). This part of the model is used to investigate sources of educational mismatch and to discuss the identification of college premiums (see Section 3.2). The model does not incorporate information- or search frictions, potentially important sources of mismatch.

Earnings: Workers have either graduated from college or not ($c_i \in \{0, 1\}$), self select into jobs ($j_i \in \{1, \dots, J\}$), and are endowed with individual productivity terms λ_i . Worker i 's log wage is a function of realized job choice, college education, and individual productivity:

$$w_i(j_i, c) = \sum_{k=1}^J \mathbb{1}(k = j_i) [\omega^k + c_i \delta^k] + \lambda_i$$

Utility: Let utility from working in job j be a function of wages and preferences for non-pecuniary attributes. The utility worker i derives from non-wage attributes of job j is characterized by $\gamma_{c_i}^j + \epsilon_i$. The first term, $\gamma_{c_i}^j$, captures correlation in preferences within educational groups, and ϵ_i^j are iid. random utility draws from a type-I extreme value distributed with scale parameter 1 and location parameter 0. Individuals choose jobs to

maximize the following utility function:

$$U_i^j = \underbrace{\omega^j + c_i \delta^j + \lambda_i}_{\text{Log wage}} + \underbrace{\gamma_i^j + \epsilon_i^j}_{\text{Preferences for non-pay characteristics}}$$

2.3 Allocation of Workers

Denote the number of college (non-college) workers as p_1 (p_0). Given the distribution of the random utility draws, the expected number of college (non-college) workers in job j is given by

$$L_1^j = p_1 \frac{\exp(w^j + \delta^j + \gamma_1^j)}{\sum_{k=1}^J \exp(w^k + \delta^k + \gamma_1^k)} \quad (1)$$

$$L_0^j = p_0 \frac{\exp(w^j + \gamma_0^j)}{\sum_{k=1}^J \exp(w^k + \gamma_0^k)} \quad (2)$$

Note that increased wage level (w^j) attracts both college and non-college workers from other jobs, while increased premium (δ^j) only attracts college workers. The total number of college and non-college workers in college jobs are calculated by summing over all jobs:

$$L_1^c = \sum_{k=1}^J \mathbb{1}(\alpha^k \geq \alpha^*) L_1^k = p_1 \frac{\sum_{k=1}^J \mathbb{1}(\alpha^k \geq \alpha^*) \exp(w^k + \delta^k + \gamma_1^k)}{\sum_{k=1}^J \exp(w^k + \delta^k + \gamma_1^k)} \quad (3)$$

$$L_0^c = \sum_{k=1}^J \mathbb{1}(\alpha^k \geq \alpha^*) L_0^k = p_0 \frac{\sum_{k=1}^J \mathbb{1}(\alpha^k \geq \alpha^*) \exp(w^k + \gamma_0^k)}{\sum_{k=1}^J \exp(w^k + \gamma_0^k)} \quad (4)$$

Equations (1) and (2) above highlight a problem of classifying jobs using the modal worker, as pioneered by [Verdugo and Verdugo \(1989\)](#): The number of jobs classified as college increases mechanically as the share of college worker grows. The equations above suggest a simple correction: classifying using overrepresentation in jobs, rather than the modal worker³.

³In Appendix [A.1](#) I investigate the sensitivity of the main results to using these two alternative classification methods.

2.4 Decomposing Mismatch

Realized mismatch is the number of non-college (college) workers in college (non-college) jobs:

$$mismatch^{nc} = L_1^{nc}$$

$$mismatch^c = L_0^c$$

Overall mismatch is the sum of mismatched non-college and college workers:

$$Mismatch = L_1^{nc} + L_0^c$$

What is the role of overall supply and demand for determining mismatch in this model? To fix ideas, consider the lower bound of educational mismatch, taking the number of college and non-college workers and jobs as given. Table 1 provides a stylized example of an economy with 100 workers and 100 jobs. Here, 30 non-college workers work in college jobs and 10 college workers work in non-college jobs, i.e. 40% of workers are mismatched.

Table 1: Labor Market Mismatch, Stylized Example

		Required education		
		Non-college	College	
Actual education	Non-college	30	30	60
	College	10	30	40
		40	60	

Note: Stylized example of the allocation of workers to jobs.

The lower bound if mismatch is realized by moving 20 workers into a better match: Move 10 mismatched college workers into college jobs, and compensate by moving the same number of mismatched non-college workers into non-college jobs. After doing this reassignment, only 20% of workers are mismatched, while the overall number of college and non-college jobs (workers) are unaffected.

In the main analysis, I use this method to decompose mismatch into a structural component capturing overall demand and supply forces. I recover the geographical component of mismatch by restricting the reassignment of workers to be within geographical units.

3 Identification and Estimation

This section describes the conceptual and empirical challenges in taking the simple model outlined in Section 2 to the data. First, I discuss how to use estimated college premiums to distinguish between college and non-college occupations. Second, Section 3.2 and 3.3 discuss the empirical challenges associated with estimating college premiums.

3.1 Classifying Jobs as College or Non-College

In a model with only two jobs, as in the model used by [Gottschalk and Hansen \(2003\)](#), a simple classification would be to classify the jobs with the highest college premium as the college occupation. In an economy with more than two jobs, however, it is unclear where one would draw the line between non-college and college jobs.

One possible solution is to classify every job with a positive college premium as a college job. That way, we would label any worker with an absolute disadvantage as mismatched.

A policymaker, on the other hand, would be interested in increasing access to college education to a level equating the marginal cost and return to college education. Since we do not observe the social cost, this approach is infeasible.

As a baseline, I will use the threshold that minimizes pooled mismatch. Using the minimum mismatch ensures that the choice of threshold does not create an excess mismatch. Additionally, this implies that the classification of jobs depends only on the ordering by estimated college premiums and not their magnitude, which is helpful because relative premiums are better identified than magnitudes.

3.2 Identification

This section discusses identification of job specific college premiums in light of the model outlined in Section 2.

Assume we have panel data on $i \in \{1, \dots, I\}$ workers, covering the period $t \in \{1, \dots, T\}$. This extends the model discussed previously by allowing for more than one

time period. Each realization of worker i 's wage is given by the potential outcome wage equation:

$$w_{it} = \sum_{k=1}^J \mathbb{1}(j_{it} = k)(\omega_t^k + c_i \delta_t^k) + \lambda_i \quad (5)$$

The parameters of interest, δ_t^j , is the college premium associated with job j in year t . The realizations of λ_i are allowed to correlate with c_i and j_{it} due to self-selection into education and jobs.

Estimating college premium using within year comparisons of college and non-college workers in job j likely gives biased estimates, due to endogenous selection into education and jobs:

$$\begin{aligned} & \mathbb{E}(w|j_{it} = j, c = 1) - \mathbb{E}(w|j_{it} = j, c = 0) \\ &= \delta_t^j + \underbrace{\mathbb{E}(\lambda_i|j_{it} = j, c = 1) - \mathbb{E}(\lambda_i|j_{it} = j, c = 0)}_{\text{Selection bias}} \\ &= \delta_t^j + \underbrace{\mathbb{E}(\lambda_i|c = 1) - \mathbb{E}(\lambda_i|c = 0)}_{\text{Selection into education}} \\ & \quad + \underbrace{[\mathbb{E}(\lambda_i|j_{it} = j, c = 1) - \mathbb{E}(\lambda_i|c = 1)] - [\mathbb{E}(\lambda_i|j_{it} = j, c = 0) - \mathbb{E}(\lambda_i|c = 0)]}_{\text{Selection into jobs}} \end{aligned}$$

The bias coming from non-random selection into education is the well documented difference in average ability between college and non-college workers, and is likely to bias estimated premiums upwards. Notice that this bias term is independent occupation and less problematic because I use relative premiums to classify occupations as college or non-college.

The last term comes from systematic differences in assignment of college and non-college workers into jobs. To illustrate this bias, consider an economy where college education is randomized. This implies that college and non-college workers are comparable in the overall economy. However, if job j attracts relatively productive college workers and relatively unproductive non-college workers, within j comparisons still yield biased inference. Furthermore, the sign of the bias is ambiguous and can vary from job to job.

Hence, selection into jobs likely affects the classification of jobs using OLS to estimate college premiums.

The panel future of the data can be used to recover unbiased estimates of relative wage levels and relative premiums. This is accomplished by comparing the wage level of worker I across different jobs:

$$\mathbb{E}_{t'}(w|j_{it'} = j', c = 0, i = I) - \mathbb{E}_t(w|j_{it} = j, c = 0, i = I) = (\omega_{t'}^{j'} - \omega_t^j) \quad (6)$$

$$\mathbb{E}_{t'}(w|j_{it'} = j', c = 1, i = I) - \mathbb{E}_t(w|j_{it} = j, c = 1, i = I) = (\omega_{t'}^{j'} - \omega_t^j) + (\delta_{t'}^{j'} - \delta_t^j) \quad (7)$$

Relative premiums are identified by taking the difference. Note that the relative wage levels include time trends in wages ($j = j'$ and $t' \neq t$), while the relative premiums capture how the premiums change over time.

The full set of relative wage levels is identified when jobs form a complete graph among non-college workers⁴. Similarly, relative premiums are identified when jobs also form a complete graph among college workers. Premium levels are identified by individuals observed both before and after graduating from college.

This identification strategy relies on the strong assumption of separability between jobs and unobserved heterogeneity, i.e. that the only source of comparative advantage across jobs is college education. However, because we condition on realized matches, it is sufficient that workers move between jobs where they have a similar advantage. To see this, allow individuals to be endowed with job-specific productivity terms, λ_i^j . For the within-individual comparisons in equations 6 and 7 to give unbiased results, we need to assume that individuals do not systematically switch between jobs where they are relatively productive and relatively unproductive:

$$\mathbb{E}_{t'}(\lambda_i^{j'} | j_{it'} = j', c_i, i = I) = \mathbb{E}_t(\lambda_i^j | j_{it} = j, c_i, i = I) \quad (8)$$

There are multiple ways this assumption can be violated. If for example college workers

⁴I.e. every pair of jobs are connected by a sequence of workers.

are more likely to work in retail jobs after experiencing a negative productivity shock, this will bias premium estimates. Another way this assumption can break down is if workers are forced to switch from occupations where they are relatively productive to occupations where they are relatively unproductive. Say that a large group of workers lose their job due to a negative industry shock, and that many take on jobs where they are relatively unproductive.

The model also rules out on-the-job learning and workers improving the match quality over time by engaging in on-the-job search. In the empirical section, I address these sources of bias by controlling for tenure and labor market experience.

Identification and estimation used in the analysis are similar to the specifications used in the literature initiated by [Abowd et al. \(1999\)](#) (AKM hereafter). AKM aims to separate the contribution of unobserved worker productivity and differences in employer pay policy on the distribution of wages by including worker and employer fixed effects in a Mincer wage regression. Estimation of the AKM model suffers from a potential "limited mobility bias", originating from employers being weakly linked by having few workers moving between them ([Bonhomme et al., 2023](#)). There are two reasons why limited mobility bias is less of a concern in this paper. First, limited mobility bias in the AKM setting arises when decomposing the variance of low wages into worker and employer components using estimated fixed effects. The estimated fixed effects are unbiased but noisy. Secondly, the number of unique employers commonly used when estimating the AKM-model ($\sim 100,000$ for the Norwegian data in [Bonhomme et al. \(2023\)](#)) by far outweighs the number of unique occupation-year combinations used in this paper ($\sim 4,600$), limiting the concern of limited mobility between occupations-years.

3.3 Estimation

I estimate college premiums using a Mincer wage regressions⁵. The main specification includes individual fixed effects, which perform within-individual across occupation-year

⁵All estimations are implemented using the user-written Stata command `reghdfe`, described in [Correia \(2016\)](#).

comparison as highlighted in the previous section:

$$\underbrace{\text{Log wage}}_{w_{ijt}} = \underbrace{\text{Experience, tenure, age}}_{X'_{it}} \beta + \sum_{k=1}^J \underbrace{\text{Occupation indicators}}_{\mathbb{1}(j_{it} = k)} \omega_t^k + \sum_{k=1}^J \underbrace{\text{Occupation} \times \text{college indicators}}_{\mathbb{1}(j_{it} = k) c_{it}} \delta_t^k + \underbrace{\text{indiv. fe.}}_{\lambda_i} + \varepsilon_{ijt} \quad (9)$$

In the above specification, X_{it} contains age squared, and a second-order polynomial in labor market experience and firm tenure. ε_{it} is the residual.

Unbiased estimation requires the following mean independence assumption to hold (Bonhomme et al., 2023):

$$\mathbb{E}[\varepsilon_{ijt} | X_{11}, \dots, X_{IT}, j_{11}, \dots, j_{IT}, c_{11}, \dots, c_{IT}, \lambda_i, \dots, \lambda_I, \omega_1^1, \dots, \omega_T^J, \delta_1^1, \dots, \delta_T^J] = 0$$

This assumption allows selection into occupations to be based on individual productivity, college education, occupation wage levels, and premiums (e.g. college workers in high-premium jobs). However, it rules out selection on time-varying occupation-specific productivity shocks captured by ε_{ijt} . This assumption captures the separability between individual heterogeneity and occupations discussed in the last section, as variation in unobserved productivity across occupations (λ_i^j) would be captured by the residual in Equation (9).

Notice that if all individuals are only observed either with or without college education, there would be perfect multicollinearity between the college-occupation indicators and a combination of occupation and individual indicators. In that case, we could specify a base occupation and estimate college premiums relative to that. As the relative premiums are what I will be using to distinguish between college and non-college jobs, this is sufficient for the rest of the analysis. However, the model fixes the premium levels because we observe some individuals both before and after graduating from college. Estimated levels should be interpreted with caution because many of the job spells used before individuals graduate from college are very early in their careers and side jobs while studying, and therefore less likely to capture variation in individual productivity.

To shed light on selection bias I compare estimates using two-way fixed effects (FE)

to estimates from four different OLS specifications: without additional controls, with controls for socioeconomic background, and with ability test score controls.

4 Matched Employer-Employee Data

This section describes the matched employer-employee data used in the empirical analysis, presents descriptive statistics of the main variables, and describes the Norwegian higher education and labor market.

4.1 Data

The analysis employs administrative registry data from Norway, linked with unique identifiers for individuals, firms and establishments. The linked employer-employee data comprise the near-universe of employments between 1995 and 2018, including information on occupation, earnings, and hours per week. This data is supplemented with data on education, parental income and education from administrative records, as well as data on ability test scores from Norwegian military conscriptions. The analysis will be restricted to the years following 2003, as occupations were not reported before that. Together, the data gives detailed information on labor market outcomes, socioeconomic background and ability for the near universe of employments in Norway.

4.1.1 Registry Data Sources

The linked employer-employee data is built from two different data sources, covering the periods 1995-2014 and 2015-2018. Until 2014, employers were required to report employments once a year. Starting in 2015, employers were required to report more detailed data each month. Both data sets consist of multiple within-year records of the same workers and have been aggregated to unique yearly combinations of workers, establishments, and occupations.

The registry data prior to 2015 covers all jobs, with the exception of spells with less than four hours per week, annual duration less than a week, or annual earnings below

10,000 NOK ($\approx 1\,100$ USD) per year. Similar restrictions have been imposed on the data following 2015 to ensure consistency across the two data sets.

Total hours, used in recovering hourly wage rates, are constructed using reported start and stop date of spells together with (average) weekly hours. Due to measurement errors in hours, yearly observations with a total number of hours less than 1,5 weeks of full-time work are removed. This is done to omit outliers in calculated wages ⁶. Firm tenure and labor market experience are calculated using the complete data set starting in 1995, meaning that experience and tenure are censored, especially early in the period.

The change in reporting between 2014 and 2015 happened around the same time as a substantial decline in oil prices, negatively affecting oil and gas-related sectors, and increasing unemployment (Hvinden and Nordbø, 2016; Norwegian Ministry of Finance, 2015). This co-occurrence makes it somewhat difficult to distinguish real changes in observed earnings and employment and differences in reporting practice.

The occupational identifiers used in the analysis are four-digit occupation codes based on ISCO 88. Some employers (mainly public) were allowed to report occupations following different standards than the main occupational standard. To overcome this issue, I construct a mapping from reported occupations to the main classification. The crosswalk relies on mapping occupations from before and after employers switch to the main standard. Although the crosswalk is likely to be imperfect, the narrow definitions of occupations increase the likelihood that imputed occupation does not differ to greatly from actual occupation. In total, 4,9 million observations (12%) have imputed occupations.

There are 289 unique occupations in the data set. To reduce noise in estimated premiums, I follow Gottschalk and Hansen (2003) in aggregating occupations with less than 50 college or non-college workers in any year together with similar occupations. After this I am left with 289 unique occupations in the main data set, giving a total of 4,624 unique combinations of occupations and years. The grouping of these occupations changes the occupation of 0.5 million observations (1.4%).

In a supplementary part of the analysis, I include data on socioeconomic background

⁶A substantial amount of these observations are associated with extraordinarily high wage levels

and ability test scores for a subset of observations. Socioeconomic background is measured using paternal income rank. Fathers' income is constructed using a unique data set on yearly income (wage and capital income), covering the period between 1967 and 2017. To remove the time component of income, I demean earnings within years. As a proxy for parental lifetime income, I use the average income between ages 51 and 53, shown to be a good proxy for lifetime income in [Markussen and Røed \(2016\)](#). Income ranks are constructed within cohorts and excludes women. Father's income rank is missing for 25% of observations, either because I don't observe father's id, or because some fathers are born before 1916 or after 1964. This data set excludes almost all immigrants and second-generation immigrants.

Ability test scores were gathered by the Norwegian Armed Forces as part of mandatory conscription to Norwegian military service after 1950. The tests combine a measure of arithmetic ability, and figure and word comparison, usually measured when individuals are between 17 and 20 years old. These tests are used to give each individual a score on a discrete scale from one to nine, which are the scores I have access to. This aggregation is likely to remove a substantial amount of information relative to the score on each test. Participation was compulsory for men and voluntary for females prior to 2014, after which it became mandatory for both men and women. Ability test score is therefore observed for most men, but few women and immigrants.

In the part of the analysis that investigates the geographical component of mismatch, I utilize the regional partition of Norway into 46 commuting zones developed in [Bhuller \(2009\)](#). The classification is based on existing regional partitions of Norway, adjusted to minimize commuting across commuting zones. Subsection [A.2](#) documents the commuting zones and the number of workers in each commuting zone.

4.2 Sample of College and Non-college Workers

Table [2](#) describes the sample used in the main analysis. The complete data set has almost 42 million observations⁷. The data consists 3.5 million workers in a total of 400,000

⁷Yearly jobs, i.e. yearly combinations of workers, establishments and occupations.

establishments. The share of college graduates in the labor market increased from 30.6% percent in 2003 to 42.5% percent in 2018.

Table 2: Descriptive Statistics

	Full Sample	College	Non-college
Wage	240.86	277.77	218.26
Log wage	5.38	5.54	5.29
Age	40.69	41.17	40.39
Woman	0.50	0.56	0.46
College graduates	0.38	1.00	0.00
In education	0.12	0.12	0.12
Immigration background	0.11	0.12	0.10
Tenure	4.71	4.63	4.75
Experience	11.03	11.76	10.58
Hours (yearly)	1324.00	1379.03	1290.31
Father's income rank (51-53)	51.31	58.35	46.70
Ability test score	5.14	6.19	4.60
Observations	41,888,574	15,906,939	25,981,635
Workers	3,543,506	1,349,589	2,612,160
Workers in more than one occupation	2,365,831	1,009,477	1,757,112
Establishments	400,983	253,745	353,717
Firms	336,742	193,136	294,863
Occupations	289	289	289
Occupation-year combinations	4,624	4,616	4,623

Note: Observation count is lower for fathers income and ability test scores, which is non-missing for 31 million and 13 million observations respectively. The number of observations with observed father's income rank *and* ability test scores is 11,327,796.

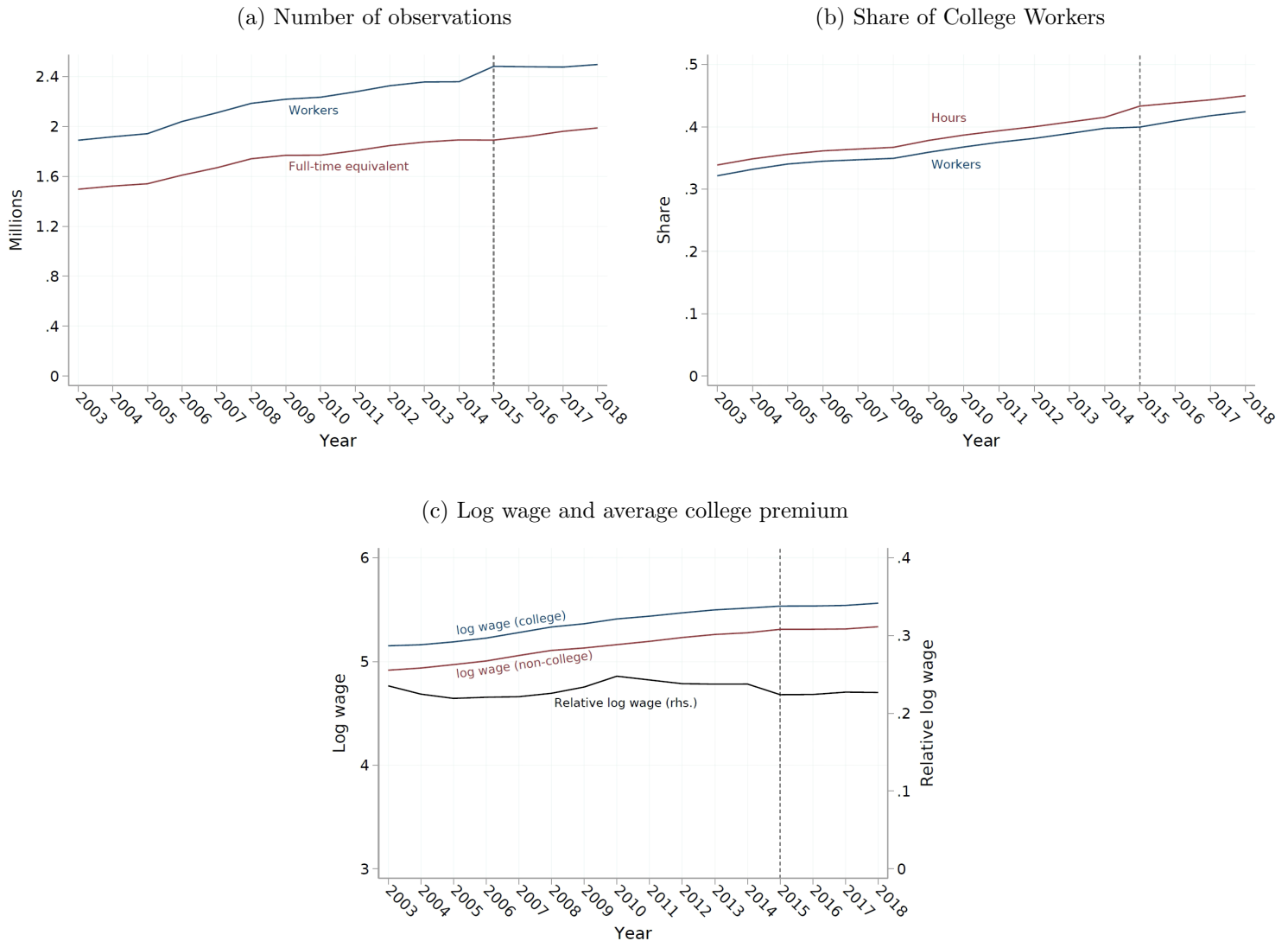
4.3 Inconsistency Across Registry Data Sources

Figure 1 documents some indication of inconsistency between the registry data sources used before and after the beginning of 2015. Panel (a) displays the yearly number of observations and full-time equivalent positions⁸. In particular, the data collected after 2015 contains more workers than before, but not a significant increase in the total number of hours worked. Panel (b) documents the increased share of college workers measured in both workers and hours. Table (c) shows a steady and parallel increase in average log wages for college and non-college workers. Taken together, these results suggest differences

⁸hours/1960

in sampling that primarily affect spells with few hours, and similarity in the inconsistencies across college and non-college workers.

Figure 1: Descriptive Statistics



Note: This figure documents time trends in (a) the number of workers (full-time equivalent jobs), (b) share of college workers (college-worker hours), and (c) log wage by education. The vertical line is in 2015 when employer’s reporting changed from annually to monthly.

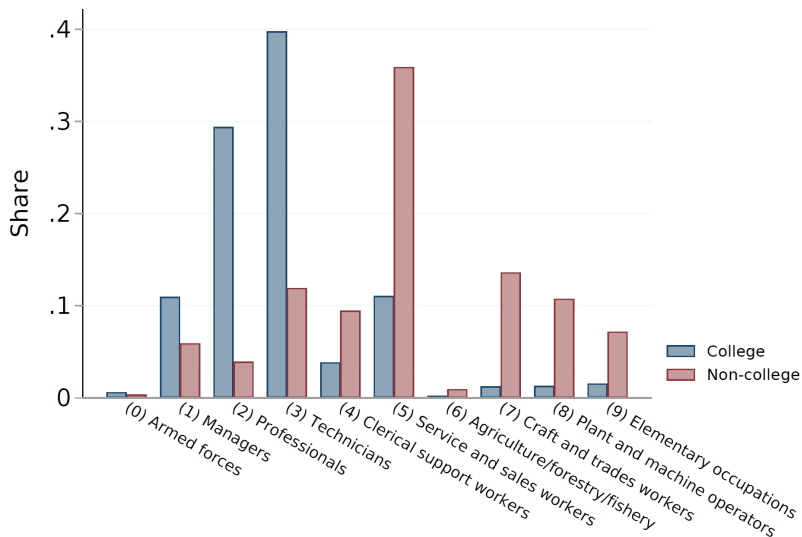
4.4 Descriptive Evidence of Sorting Into Occupations

Figure 2 documents the selection into one-digit occupations. This figure indicates that aggregating jobs within occupations is likely to capture a large portion of the differences in educational requirements across occupations. College workers are strongly selected into managerial, professional and technical occupations, while non-college workers are

selected into occupations having a first digit between 5 and 9. This selection mirrors ILO’s mapping of one-digit occupations into the required level of education (ILO, 2006), classifying occupations with a first digit between 1 and 3 as requiring at least some higher education, and occupations starting on the digits 4 to 9 as requiring high school or lower. Jobs in the armed forces are not mapped to required level of education.

A major weakness of ILO’s mapping of occupations to required levels of education is that it does not capture within-occupation changes in educational requirements. Additionally, the mapping is silent about variation within one-digit occupation.

Figure 2: Selection into Occupations (1-digit)



Note: This figure shows the distribution of workers across major occupations. Major occupations are one-digit occupations (shown in parentheses).

5 Main results

This section presents the main empirical result. The distribution of premiums estimated with fixed effects are presented in Section 5.1, with additional comparison between the fixed effects and OLS specifications. Section 5.2 documents how occupations are classified into college and non-college occupations. Section 5.4 decomposes of mismatch into a structural component, a geographical component, and a within-location assignment component. Lastly, section 5.5 shows that the observed increase in the share of college jobs is driven by an upward trend in estimated premiums.

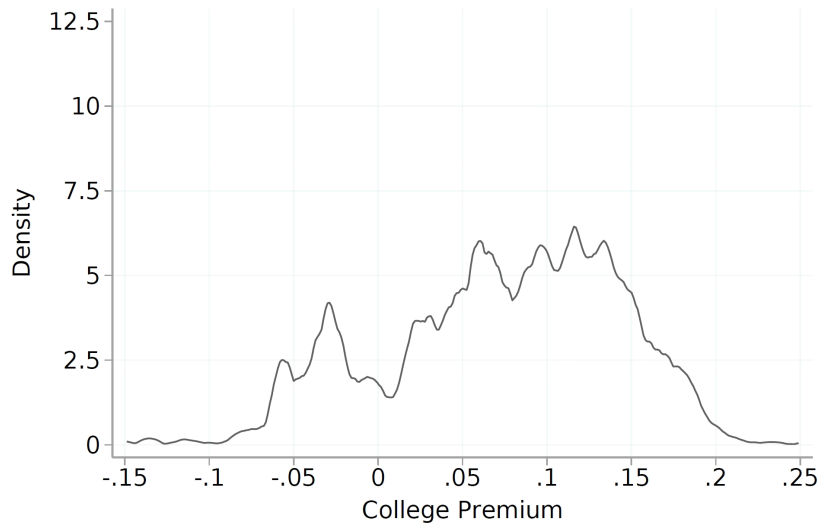
5.1 Estimated Premiums

The distribution of estimated yearly college premiums across jobs are reported in Figure 3. The figure pools the estimates from all years, including occupations with a share of college workers below 10 and above 90 percent for completeness⁹. The densities are weighted by the number of hours worked in each job, meaning that the high density around 0.7 represents a large number of hours worked in jobs with premiums ≈ 0.7 , not the number of occupations.

As documented in Figure 3, Around 2/3 of hours worked are performed in occupations with an estimated premium above 5%, while occupations with a negative premium account for around 1/5 of hours worked.

Table 3 documents the occupations that account for the largest shares of hours worked in negative-, medium, and high-premium occupations. These occupations make intuitive sense. Almost 40% of negative premium occupations are accounted for by shop assistants, while large premium occupations are populated with specialized professional occupations and CEOs.

Figure 3: Distribution of College Premiums Across Jobs (pooled)



Note: This figure documents a kernel density plot of estimated yearly college premiums, weighted by hours worked (4,624 parameters in total). The figure includes occupations with a share of college workers between 10% and 90%.

⁹These occupations are classified based on the share of college workers, and not the estimated college premium.

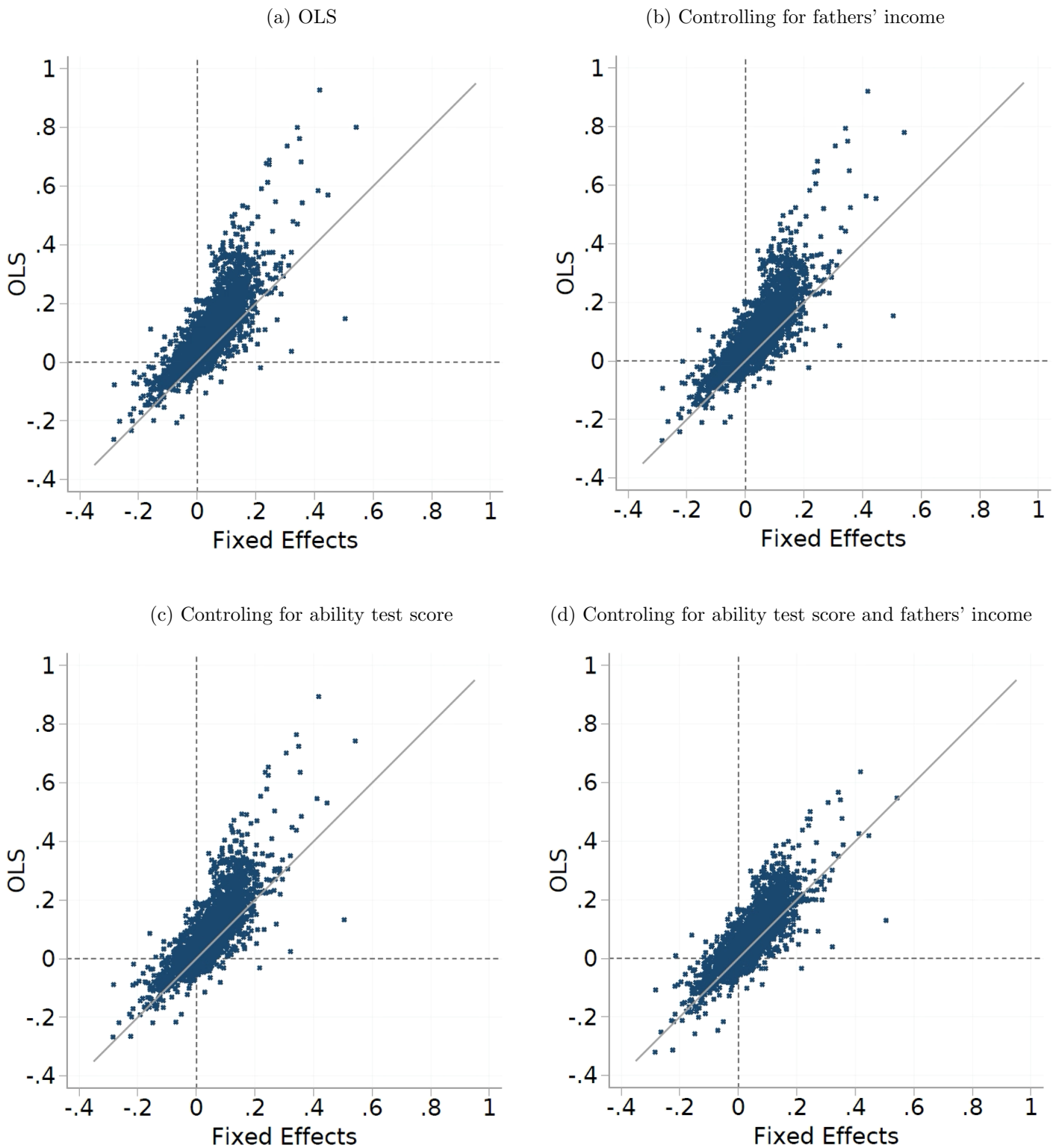
Table 3: Largest Low- Medium- and High Premium Occupations

Occupation (1)	Share (2)	Premium (3)
A: Estimated premium below 0%		
Shop assistants	0.39	−0.06 to −0.02
Company cleaners	0.08	−0.05 to +0.02
Personal care and related workers	0.07	−0.02 to +0.03
Waiters, waitresses and bartenders	0.05	−0.06 to −0.04
Car, taxi and van drivers	0.03	−0.05 to −0.03
B: Estimated premium between 0 and 10%		
Institution-based personal care workers	0.13	+0.04 to +0.08
Child-care workers	0.11	+0.02 to +0.07
Office clerks	0.06	+0.00 to +0.07
Secretaries	0.05	+0.03 to +0.07
Personal care and related workers	0.04	−0.03 to +0.03
C: Estimated premium above 10%		
Computing professionals	0.06	+0.07 to +0.17
Directors and chief executives	0.06	+0.12 to +0.20
Public administration professionals	0.05	+0.08 to +0.13
Technical and commercial sales representatives	0.05	+0.09 to +0.16
Personnel and careers professionals	0.03	+0.09 to +0.15

Note: This table shows the largest occupations among low, medium, and high-premium occupations. The premiums change over time, meaning that some occupations change groups. Shares are only calculated using occupation-year combinations where estimated premiums are in the specified range. Premium ranges show the maximum and minimum estimated premiums across all years.

The premium estimates presented above are the baseline regression, controlling for unobserved heterogeneity using individual fixed effects. Figure 4 compares these results with OLS estimates with and without controlling for observed IQ-test scores and paternal income rank. Panel (a) documents that estimated college premiums are biased upwards in high-premium occupations relative to the fixed effects estimation. This bias remains after controlling for fathers' income and/or ability test scores, but are somewhat less pronounced when I control for both. One possible reason for upward bias in high-premium occupations is if human capital-intensive occupations tend to attract particularly productive college workers without simultaneously attracting productive non-college workers.

Figure 4: Comparison Between Baseline and OLS Estimates of College Premiums



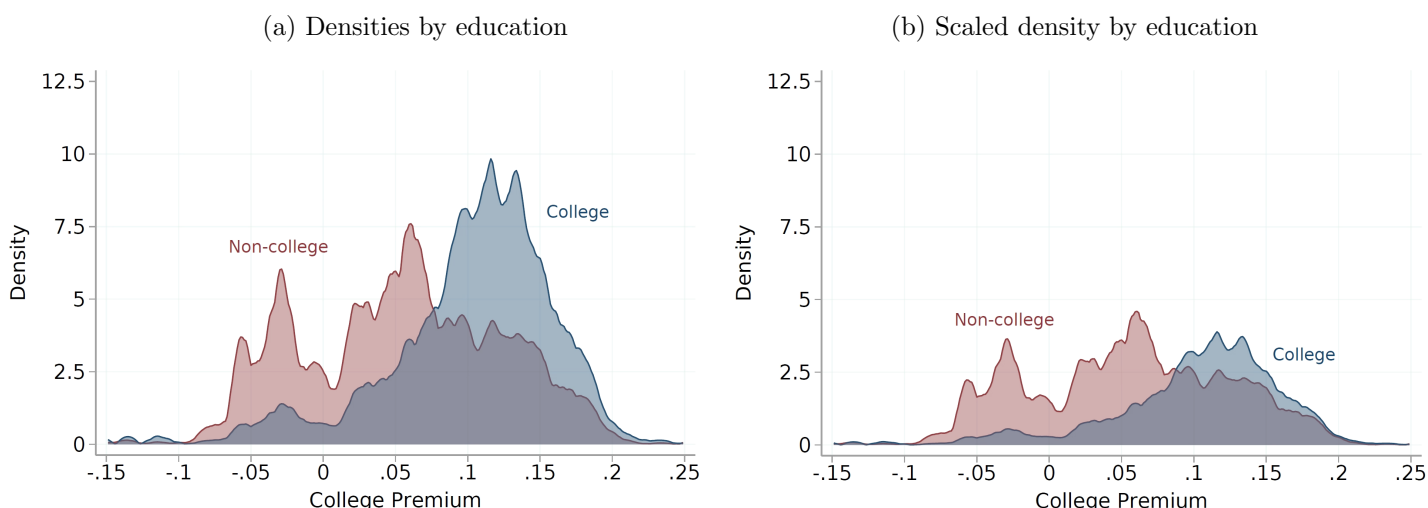
Note: This figure compares premiums estimated using the baseline specification (x-axes) to (a) OLS, (b) OLS with controls for father's income rank, (c) OLS with controls for ability test scores, and (d) OLS with both sets of controls. The figure shows parameters for occupations where the share of college workers is between 10 and 90 %, meaning that the graph contains multiple data points for the same occupation (across different years). All regressions are performed on the sample with observed ability test scores and father's income rank (i.e. 30% of the full sample).

5.2 College and Non-college Jobs

Figure 5 documents selection into high and low premium occupations by education, which will be used to classify occupations into college and non-college. Selection is particularly strong in low-premium occupations, which employ almost exclusively non-college workers. Although college-educated workers are overrepresented in occupations with estimated premiums above approximately 8 percent, there is still a considerable amount of non-college workers in these occupations.

Since the number of non-college workers is larger than the number of college workers, the relative height of the densities in panel (a) does not represent the relative number of workers. Panel (b) shows the same densities scaled to make the height densities comparable in hours worked¹⁰.

Figure 5: College Premiums by Education (pooled)



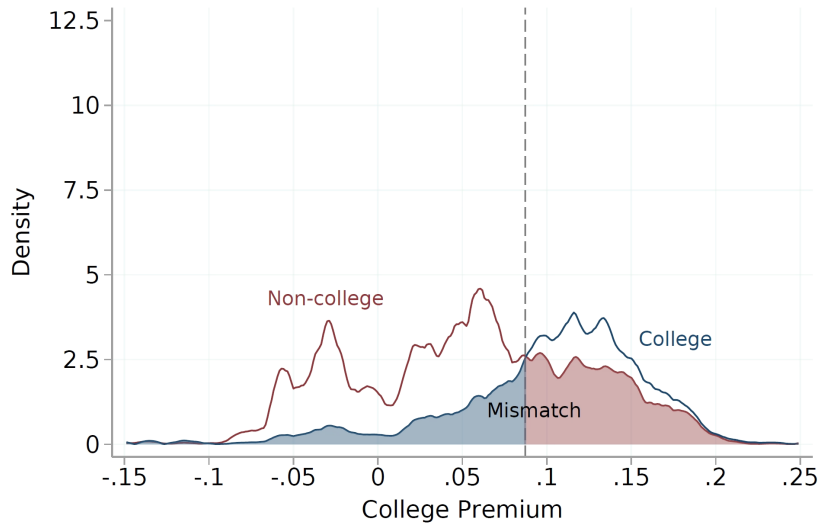
Note: This figure documents Kernel density plots of 4,624 estimated yearly college premiums, weighted by hours worked. Panel (a) shows the densities separately for non-college and college workers. Panel (b) displays the same densities scaled using the share of college and non-college workers. Panel (b) implies that there are approximately the same number of hours worked by college and non-college workers in jobs with around 9 percent college premium.

The scaled densities cross at 9.2%, meaning that occupations with a higher estimated premium employ more college workers than non-college workers. Classifying occupations with estimated premiums above 9.2% as college occupations, therefore, gives the lowest pooled level of mismatch. To see this, consider Figure 6 where I have augmented Figure 3

¹⁰I.e. scaled by the relative number of college and non-college hours.

(c) to indicate mismatch using this threshold (colored regions). If we move the threshold upwards (downwards), we move more college (non-college) workers into mismatch than non-college (college) workers out of mismatch, showing that using this threshold minimizes pooled mismatch. For the remainder of the analysis, 9.2% will be used as a baseline threshold for classifying occupations into college and non-college¹¹.

Figure 6: Premium Density With Mismatch



Note: This figure documents kernel density plots of college premium across occupation by education, rescaled to make the densities comparable in absolute hours. The dashed vertical line represents the premium that minimizes overall (pooled) mismatch. The blue (red) region indicate mismatched college (non-college) workers using this threshold.

5.3 Mismatch

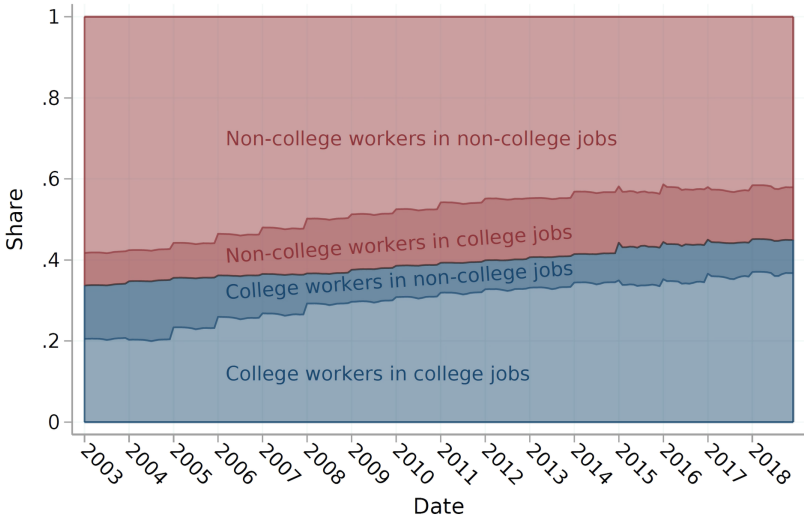
Figure 7 documents the allocation of workers to jobs using the baseline classification of occupations. The share of workers in college jobs grows, both among college and non-college workers. As shown in section 5.5, this is driven mostly by increased premiums changing the classifications from non-college to college over time. The overall share of mismatched workers, however, is stable slightly above 20% of workers throughout the period. The share of mismatch is generally decreasing among college workers, and growing for non-college workers.

Table 4 documents some stylized facts about workers by education and job classifi-

¹¹While classifying occupations with a college share below 10% and above 90% using the college share.

cation. Workers in college jobs have higher wages, are on average older, and have a lot more experience than workers in non-college jobs. These patterns suggest that workers tend to move towards college jobs after spending some time in non-college jobs, consistent with search and information frictions. The table also documents that non-college workers in college jobs have higher ability test scores and socioeconomic backgrounds than non-college workers in non-college jobs, suggesting that high ability can substitute for college workers in some of the occupations classified as college jobs.

Figure 7: Matched and Mismatched Workers by Education



Note: This figure shows the monthly allocation of college and non-college workers into college and non-college occupations, measured in hours.

Table 4: Descriptive Statistics by Education and occupation classification

Education:	Non-college		College	
Job:	Non-college	College	Non-college	College
<i>Panel 1: Average across jobs</i>				
Wage	206.84	272.67	232.49	295.74
Log wage	5.24	5.52	5.35	5.61
Age	39.33	45.42	37.11	42.79
Woman	0.48	0.36	0.55	0.57
In education	0.13	0.05	0.22	0.08
Immigration background	0.11	0.06	0.19	0.10
Tenure	4.54	5.75	3.38	5.12
Experience	9.95	13.63	9.29	12.73
Hours (yearly)	1234.76	1554.89	1124.61	1480.03
Father's income rank (51-53)	45.59	51.49	56.85	58.90
Ability test score	4.48	5.09	5.98	6.29
<i>Panel 2: Observations</i>				
Observations	21,473,810	4,507,825	4,520,359	11,386,580
Full time equivalent (fte)	13,528,102	3,5761,11	2,593,703	8,598,208

Note: This table describes workers by education and by job-type (college or non-college). Panel 1 shows averages across jobs, meaning yearly combinations of employees, employers, and occupations. Fte is total number of hours within a year divided by 1,960. Note that fathers' income rank and ability test scores are observed for a subset of 30% and 74% of observations, respectively.

5.4 Decomposition of Mismatch

This section documents the decomposition of mismatch into a (i) a structural mismatch that captures misalignment between underlying demand and supply factors (ii) a geographical component, and (iii) a residual within-location assignment component.

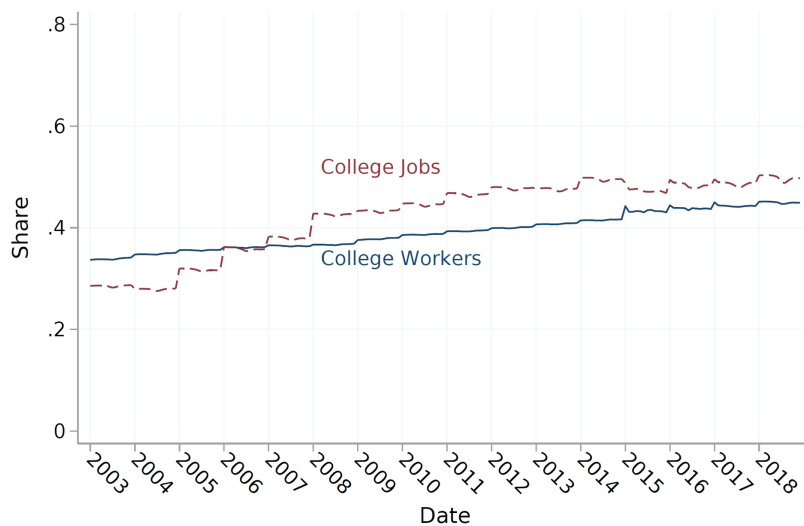
5.4.1 Structural Mismatch

To recover the component of mismatch driven by underlying demand and supply, imagine that we can freely reallocate workers between occupations. As an example, consider January 2006 in figure 7. A feasible reduction in mismatch within this month is to swap every mismatched college with a mismatched non-college worker. Because there are approximately the same number of mismatched college and non-college workers, the lowest attainable level of educational mismatch within this month is approximately zero. Compare this to January 2003, where there are fewer mismatched non-college workers

than college workers, meaning that the lowest level of mismatch is greater than zero.

As the example above makes clear, structural mismatch is apparent in Figure 7 as the absolute value of the difference between mismatched college and non-college workers. Similarly, the lower bound of mismatch is equal to the difference between the share of college workers and college jobs, documented in Figure 8. This figure shows the share of college jobs growing more than the share of college workers, in particular in the first half of the period. Assignment mismatch is the residual mismatch caused by the realized assignment of workers to jobs.

Figure 8: Share of College Workers and Jobs (monthly)



Note: This figure documents the monthly share of college workers and college jobs, both measured in hours. Within year changes in the share of college jobs are driven by changes in hours worked in occupations. The yearly jumps in the share of college jobs are driven by changes the binary classification of occupations.

5.4.2 Geographical Mismatch

The decomposition above is based on the thought experiment of moving workers freely between jobs. To recover the geographical component of mismatch, this section considers the lower bound of mismatch by restricting the reassignment of workers within commuting zones.

The geographical component of mismatch depends on the correlation between the share of college workers and jobs across commuting zones. Figure 9 documents the joint

distribution of college workers and jobs. Panel (a) documents a high correlation in the pooled data set. Panel (b) documents that most commuting zones had a higher share of college workers than jobs in 2003, a relationship that is inverted in 2018. Both panels suggest that geographical mismatch might not be a prominent source of mismatch.

Figure 9: Share of College Workers and Jobs Across Commuting Zones



Note: These figures compare the share of college jobs to the share of college workers, documenting that commuting zones with a large share of college workers have correspondingly high shares of college jobs. The scatter plot in Panel (a) plots every commuting zone-year combination, showing regions with more workers as less transparent, while Panel (b) compares 2003 to 2018. The local average lines weights by hours worked in both panels.

5.4.3 Results

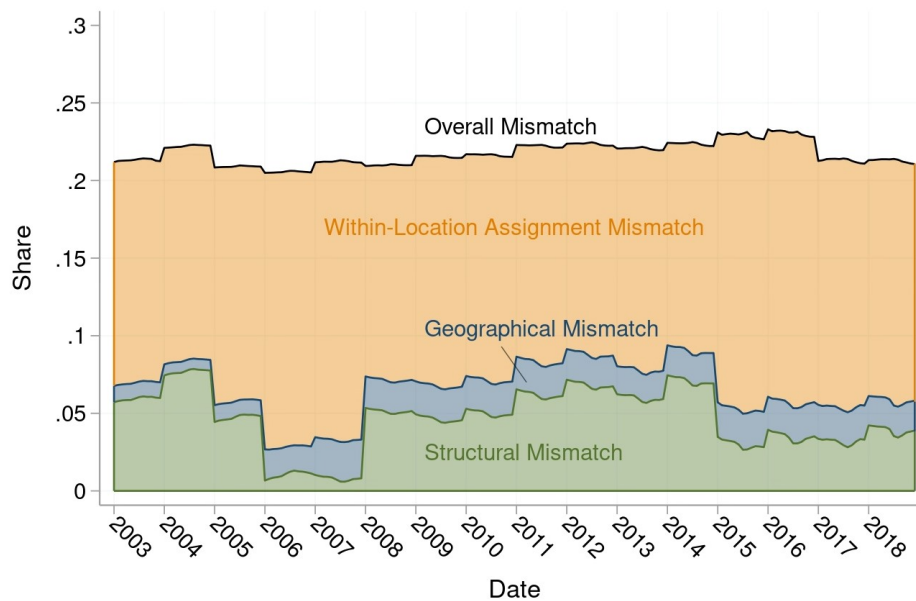
Figure 10 decomposes overall mismatch into (i) a structural mismatch that captures misalignment between underlying demand and supply factors (ii) a geographical component, and (iii) a residual within-location assignment component.

Overall mismatch is stable, affecting around 20% of workers. The structural component is somewhat stable with a slight dip around 2006-2008. Comparing this to Figure 7 reveals that the structural mismatch is driven by a higher share of college workers before

2006, while this relationship is reversed for the remainder of the period. Within-location assignment mismatch accounts for 60% or more of mismatch throughout the period.

Year-to-year changes are affected by variation in estimated premiums around the threshold and are visible as discontinuities at the beginning of most years. Regardless, of these instabilities, within-location assignment mismatch is accountable for a majority of mismatch throughout the period.

Figure 10: Mismatch Decomposed



Note: This figure documents overall mismatch, and the decomposition into three distinct components. Structural mismatch captures misalignment between underlying demand and supply factors, geographical mismatch captures the misalignment between the location of workers and jobs across commuting zones, and within-location assignment is the residual mismatch attributed to the assignment of workers to jobs within commuting zones. The discontinuities around the beginning of most years are driven by small changes in estimated premiums around the threshold used to classify occupations.

5.5 What is Driving the Growth in the Share of College Jobs?

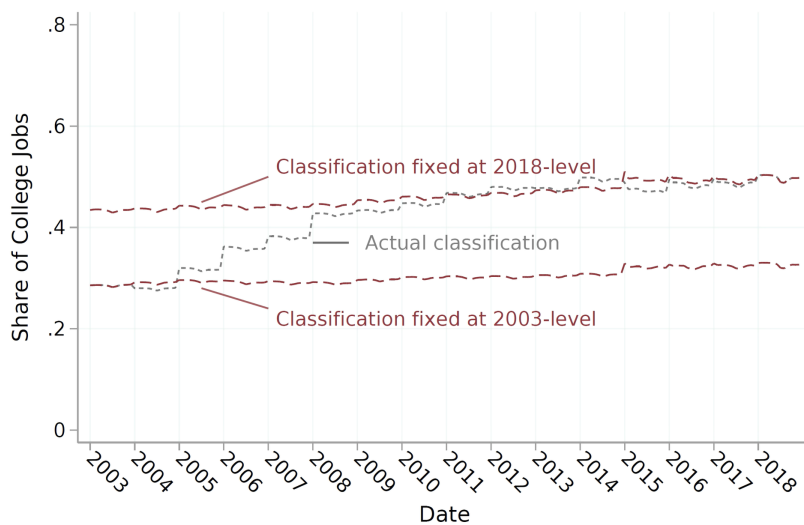
This section shows that the increase in the share of college jobs observed in Section 5.2 is driven by an upward trend in estimated premiums, consistent with a growing number of jobs requiring college education.

The observed increase in the share of college jobs can in theory come from many changes in the labor market. First, growth in college jobs can be driven by employment moving from non-college to college occupations, or alternatively by occupations changing

classification over time. Second, changes in classifications can be driven by trends in estimated premiums or college worker shares ¹². Third, changes in classification can be an artifact of small changes around the arbitrary thresholds used in the classification procedure, or systematic shifts in premiums over time.

Figure 11 documents that the time trend in the share of college jobs is almost flat when classification is not allowed to change over time, meaning that changes in classification drive the increasing share of college jobs.

Figure 11: College Job Share with Fixed and Variable Classification



Note: This figure compares the estimated time trend in the share of college jobs to alternative classifications that fix occupations to the 2003 and 2018 classification. “Actual classification” allows occupation to change between college and non-college across years by changes in estimated premiums or the share of college workers, while the red lines use the estimated premiums and shares in 2003 (2008) across all years.

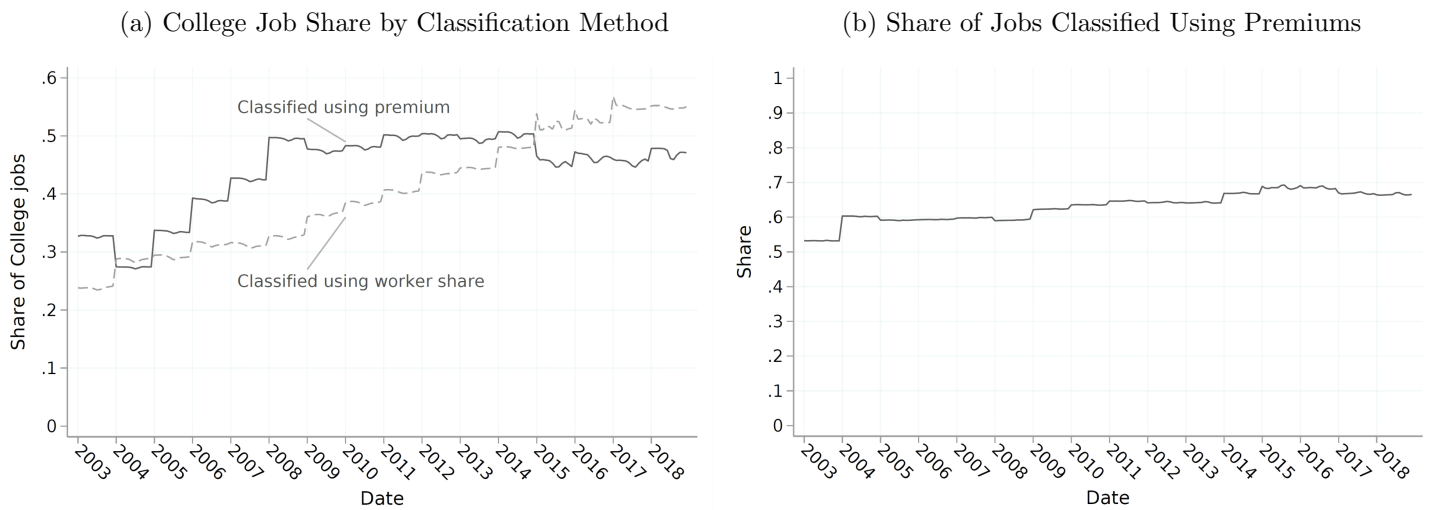
Figure 12 shows that both classification methods¹³ leads to an increase in the share of college jobs. Among occupations classified using premiums, this is happening mostly before 2008, while the college job share is growing steadily throughout the period among occupations with a college share below 10% or above 90%. The share of college jobs among the first group shows a sizable decline between December 2014 and January 2015, possibly indicating inconsistency between the two registry data sources (see Section 4). Around 50 – 60% of jobs are classified using estimated premiums, growing somewhat

¹²Because occupations with more than 90 or less than 10% college workers do not use estimated premiums for classification.

¹³College premium above 9.2% or share of college workers above (below) 90%. (10%).

throughout the period, as seen in Panel (b).

Figure 12: College Job Share by Classification Method



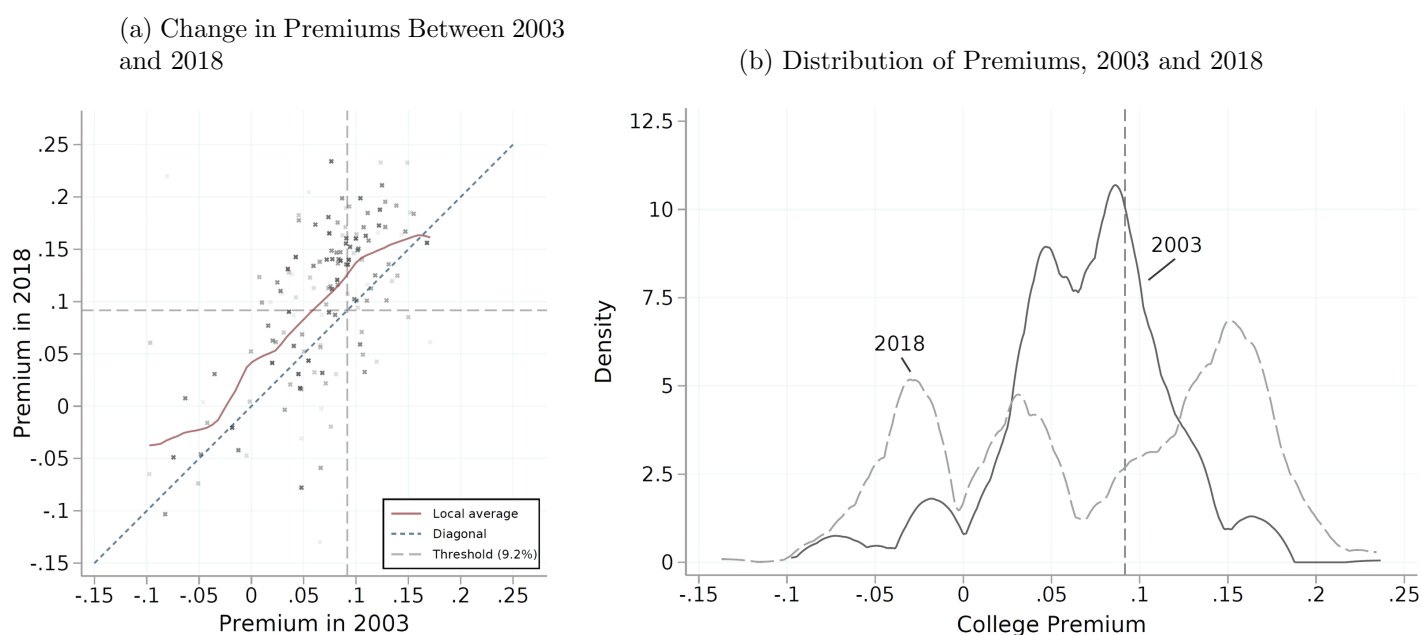
Note: This figure documents the relative contribution of the two methods used to classify occupations. Jobs with a college worker share below (above) 10% (90%) are classified as non-college (college) without inspecting the estimated premiums. The remaining jobs are classified using estimated premiums. Panel (a) shows that both classification methods contribute to a rise in the share of college jobs, with a slightly higher contribution from using the college job share. Panel (b) shows a relatively stable contribution in the share of jobs classified using premiums.

Figure 13 Panel (a) documents how estimated premiums change between the first and last year of the period. On average, premiums tend to increase over time, and this pattern holds regardless of estimated premiums in 2003. The occupations located in the upper left square of Panel (a) change classification from non-college in 2003 to college in 2018.

Panel (b) shows how the overall distribution of premiums shifts over time. The distribution in 2003 is close to unimodal, indicating many occupations in the middle of the skill distribution between 2 and 15 percent in 2003. This picture is severely altered in 2018, with one peak for jobs with negative premiums, and a large mass of jobs with estimated premiums above 15 percent.

The polarization of premiums visible in Figure 13, Panel (b), is analogous to the polarisation of the US labor market documented in Autor and Dorn (2013). Autor and Dorn shows that occupations around the top and bottom of the skill distribution experienced growth in employment and wages relative to jobs around the middle of the skill distribution. The premium estimates here suggest that there is also a tendency for polarization in

Figure 13: Estimated Premiums in 2003 and 2018



Note: This figure compares estimated premiums in 2003 and 2018. Panel (a) compares estimated premiums within occupations, while panel (b) shows the overall distributions of premiums in 2003 and 2018.

skill requirements. While many occupations had a premium of 5 to 10% in 2003, the 2018 labor market is split between jobs giving small college premiums and occupations with large premiums. The significance of this distributional shift, however, is not assessable due to the lack of standard errors of estimated premiums.

Table 5 takes a closer look at the occupations that are driving most of the change in the share of college jobs. Almost 20% of the increase is driven by computing professionals and public administration professionals. While it might make intuitive sense to think about some of these occupations as requiring a college education, the prevalence of non-college workers in these occupations suggests that at least some of the tasks performed by workers classified using these occupations may be performed by workers without college education (potentially workers with particular skills acquired by on-the-job learning or elsewhere).

Lastly, Figure 14 documents why there is a growth of college jobs among jobs classified using the college worker share. Panel (a) shows that there has been a steady decline in occupations with less than 10% college workers, and a steady increase in occupations with

Table 5: Occupations Accountable for Increase in College Job Share

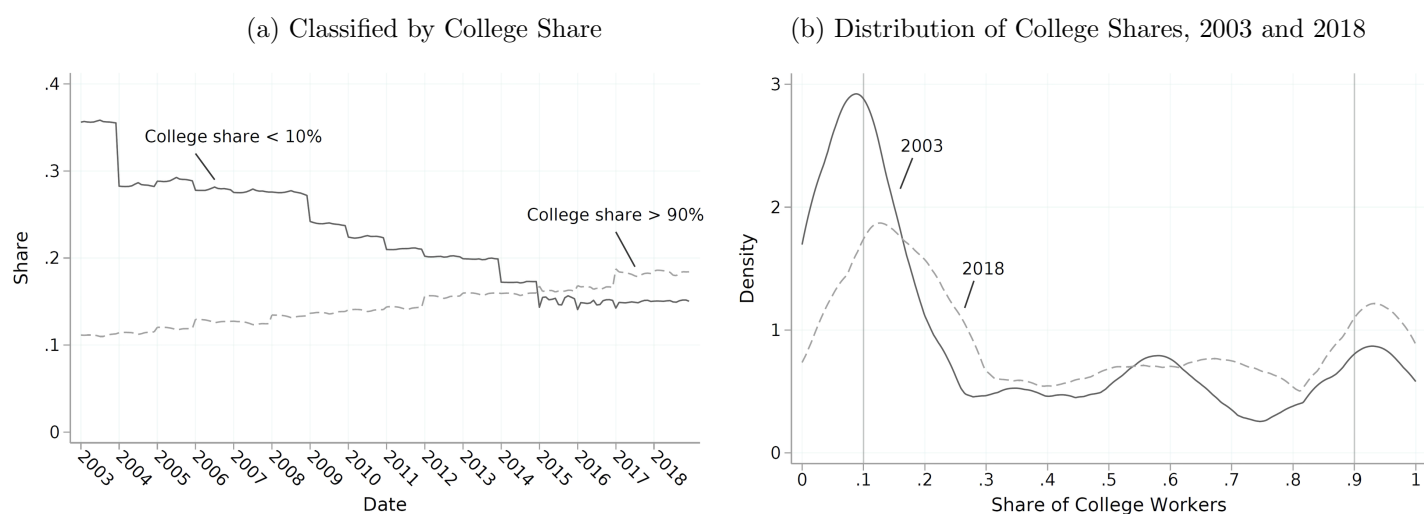
Occupation	Share of college workers		Estimated premium		Share of increase in demand
	2003	2018	2003	2018	
Computing professionals	60.2%	70.2%	7.4%	16.5%	10.7%
Public administration professionals	58.6%	71.7%	8.2%	12.1%	8.4%
Civil Engineers (construction)	76.5%	64.7%	9.1%	16.1%	4.8%
Physical and engineering science technicians	40.2%	52.9%	7.7%	14.1%	4.7%
Social workers and child welfare officers	86.6%	92.6%	7.1%	10.4%	4.7%
Auditors and accountants	52.3%	65.4%	4.2%	14.3%	4.3%
Customer service in banking	28.0%	59.3%	3.5%	13.1%	3.7%
Computer engineers and technicians	55.6%	57.3%	8.5%	13.9%	3.7%
Economics and business administration	63.2%	74.5%	7.4%	18.1%	3.0%
Civil engineers (natural resources)	80.8%	79.3%	4.5%	17.8%	2.7%

Note: This table documents the share of college workers and estimated in premiums in occupations that are contributing most to the increase in the college job-share between 2003 and 2018. The rightmost column is calculated by comparing the share of college jobs with and without fixing the occupation at the 2003 classification (college or non-college). The 10 occupations listed here account for around 50% of the observed increase in the share of college jobs.

a college worker share above 10%. This trend is partially mechanical: As the share of college workers increases in the overall economy, the expected number of occupations with less than 10% college workers decreases, and the number of occupations with more than 90% college workers increases, all else constant. Panel (b) shows that there were many jobs in occupations with almost no college workers in 2003, and that these occupations increase their college share over time. This reduction is the main driver behind the increase in college jobs among occupations with a college share outside the interval 10 – 90%.

Overall, around half of the increase in college job share can be attributed to a general increase in estimated college premiums, while the other half of the increase is due to a reduction in the number of occupations with less than 10% college workers. In Appendix C, I show that the large increase in the share of college jobs is not due to the mechanical effect of increasing the share of college workers pointed out above. This is because occupations with less than 10 or more than 90% college workers have estimated premiums that lead to the same classification as using the college share.

Figure 14: Change in College Share



Note: This figure documents the share of jobs with less than (more than) 10% (90%) college workers. Panel (a) shows that there is a decline (increase) in the share of jobs in occupations with less than 10% (more than 90%) college workers. Panel (b) compares the distribution of college shares across occupations in 2003 and 2018. Both panels calculate shares and weights with hours worked.

6 Conclusion

In this paper, I provide evidence of the evolution of mismatch in the college market between 2003 and 2018. I show that about 20% of Norwegian workers are mismatched, with similar numbers for college and non-college workers.

Despite large shifts in the share of college workers, from around 30 to around 40%, I find little contribution from misalignment between overall supply and demand for college workers. This result is driven by large positive shifts in estimated college premiums across a wide set of occupations, suggesting a rising requirement for college skills within occupations. Additionally, I find that college workers and college jobs are largely concentrated in the same areas, and almost no contribution from geographical misalignment between college workers and jobs for mismatch. The within-location assignment component can account for more than two-thirds of overall mismatch, and its contribution has remained relatively unchanged over time. This evidence points to the vital role of reassignment policies that improve matching in the college market.

References

- Abowd, John M., Francis Kramarz, and David N. Margolis**, “High Wage Workers and High Wage Firms,” *Econometrica*, 1999, *67* (2), 251–333.
- Autor, David, Claudia Goldin, and Lawrence F. Katz**, “Extending the Race between Education and Technology,” *AEA Papers and Proceedings*, May 2020, *110*, 347–51.
- Autor, David H. and David Dorn**, “The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market,” *American Economic Review*, August 2013, *103* (5), 1553–97.
- Bhuller, Manudeep Singh**, “Inndeling av Norge i Arbeidsmarkedsregioner,” *Statistics Norway*, 2009.
- Bonhomme, Stéphane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler**, “How Much Should We Trust Estimates of Firm Effects and Worker Sorting?,” *Journal of Labor Economics*, 2023, *41* (2), 291–322.
- Correia, Sergio**, “Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator,” Technical Report 2016. Working Paper.
- Duncan, Greg J. and Saul D. Hoffman**, “The incidence and wage effects of overeducation,” *Economics of Education Review*, 1981, *1* (1), 75–86.
- Fredriksson, Peter, Lena Hensvik, and Oskar Nordström Skans**, “Mismatch of Talent: Evidence on Match Quality, Entry Wages, and Job Mobility,” *American Economic Review*, November 2018, *108* (11), 3303–38.
- Gottschalk, Peter and Michael Hansen**, “Is the Proportion of College Workers in Noncollege Jobs Increasing?,” *Journal of Labor Economics*, 2003, *21* (2), 449–471.
- Grazier, Suzanne, Nigel C. O’Leary, and Peter J. Sloane**, “Graduate Employment in the UK: An Application of the Gottschalk-Hansen Model,” IZA Discussion Papers 3618 July 2008.

- Hvinden, Even C. and Einar W. Nordbø**, “The fall in oil prices and the labourmarket,” Technical Report 7, Norges Bank (central bank of Norway) 2016.
- ILO**, “International Standard Classification of Occupations (ISCO-08), Conceptual Framework (Annex 1),” <https://www.ilo.org/public/english/bureau/stat/isco/docs/annex1.pdf> 2006.
- Kiker, B. F., Maria C. Santos, and M. Mendes de Oliveira**, “Overeducation and undereducation: Evidence for Portugal,” *Economics of Education Review*, 1997, 16 (2), 111–125.
- Lise, Jeremy and Fabien Postel-Vinay**, “Multidimensional Skills, Sorting, and Human Capital Accumulation,” *American Economic Review*, August 2020, 110 (8), 2328–76.
- Liu, Kai, Kjell G. Salvanes, and Erik Ø. Sørensen**, “Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession,” *European Economic Review*, 2016, 84, 3 – 17. European Labor Market Issues.
- Marinescu, Ioana and Roland Rathelot**, “Mismatch Unemployment and the Geography of Job Search,” *American Economic Journal: Macroeconomics*, July 2018, 10 (3), 42–70.
- Markussen, S. and K. Røed**, “Ulikhet og Sosial Mobilitet (Inequality and Social Mobility),” in I. Frønes and L. Kjølørød, eds., *Det norske samfunn*, Gyldendal Akademiske forlag, 2016, chapter 17.
- Norwegian Ministry of Finance**, “Official Governmental document ”Meld. St. 1 (2015-2016)”,” 2015.
- OECD**, *Education at a glance: OECD indicators 1996*, OECD, 1996.
- , *Education at a Glance 2021* 2021.

O’Leary, Nigel and Peter Sloane, “Too many graduates? An application of the Gottschalk–Hansen model to young British graduates between 2001–2010,” *Oxford Economic Papers*, 2016, 68 (4), 945–967.

Verdugo, Richard R. and Naomi T. Verdugo, “The Impact of Surplus Schooling on Earnings: Some Additional Findings,” *The Journal of Human Resources*, 1989, 24 (4), 629–643.

Appendices

A Robustness to Alternative Classifications

A.1 Alternative Methods of Classifying Occupations

In this section I investigate how well the employed classification corresponds to alternative ways of distinguishing between college and non-college occupations.

ILO classification: ILO classifies managers, professionals and technicians as college occupations ILO (2006)¹⁴. The remaining occupations are classified as requiring high school or less. ILO does not classify occupations in the Armed Forces as requiring college education or not. Table 6, Panel (a), documents high correspondence between ILO’s classification and the one used in this paper. The exception is some occupations classified as requiring college education by ILO but not in the baseline specification used in this paper.

Share of college workers: A method used in the literature is to classify occupations that employ more college workers than non-college workers as college occupations (Kiker et al., 1997; Verdugo and Verdugo, 1989). As mentioned in Section 3.2, this approach is sensitive to the overall share of college workers in the economy.

To see this, consider a stylized example of an economy with two occupations, A and B , 100 non-college workers and 5 college workers. Suppose that 6 non-college workers and all 5 college workers are employed in occupation A , and that the remaining non-college workers work in occupation B . Although college (non-college) workers are strongly selected into (out of) occupation A , this occupation is classified as a non-college occupation.

In the lens of the model outlined in Section 2, the modal worker in occupation j is realized by the following equation:

$$\frac{L_1^j}{L_0^j} = \frac{p_1}{p_0} \frac{\left(\frac{\exp(w^j + \delta^j + \gamma_1^j)}{\sum_{k=1}^J \exp(w^k + \delta^k + \gamma_1^k)} \right)}{\left(\frac{\exp(w^j + \gamma_0^j)}{\sum_{k=1}^J \exp(w^k + \gamma_0^k)} \right)} > 1 \quad (10)$$

¹⁴See 4 for more details

Where the modal worker is not only a function of the distribution of preferences and wages, but also the overall ratio of college to non-college workers. A simple correction for this is to divide the within-occupation ratio of workers with the overall ratio of workers:

$$\frac{L_1^j/L_0^j}{p_1/p_0} = \frac{L_1^j/p_1}{L_0^j/p_0} > 1 \quad (11)$$

This can be interpreted as the actual share of college workers relative to the share under random assignment. Alternatively, the expression above can be interpreted as how strongly college workers are selected into occupation j relative to non-college workers.

Table 6, Panel (b) and (c), compares these two alternative methods for classifying occupations with the baseline method, measured as pooled hours worked. These classification methods are well aligned with the main classifications, using the same labels for 86% and 90% of hours worked in the economy.

Table 6: Comparison of Baseline Classification to Alternative Classifications
(a)

		ILO's Classification			
		Non-college	College	Not classified	
Baseline	Non-college	48.4%	8.0%	0.6%	57.0%
	College	0.1%	42.0%	0.0%	43.0%
		49.4%	50.1%	0.6%	

(b)

		Modal Worker		
		Non-college	College	
Baseline	Non-college	52.5%	4.53%	57.0%
	College	9.1%	33.9%	43.0%
		61.6%	38.4%	

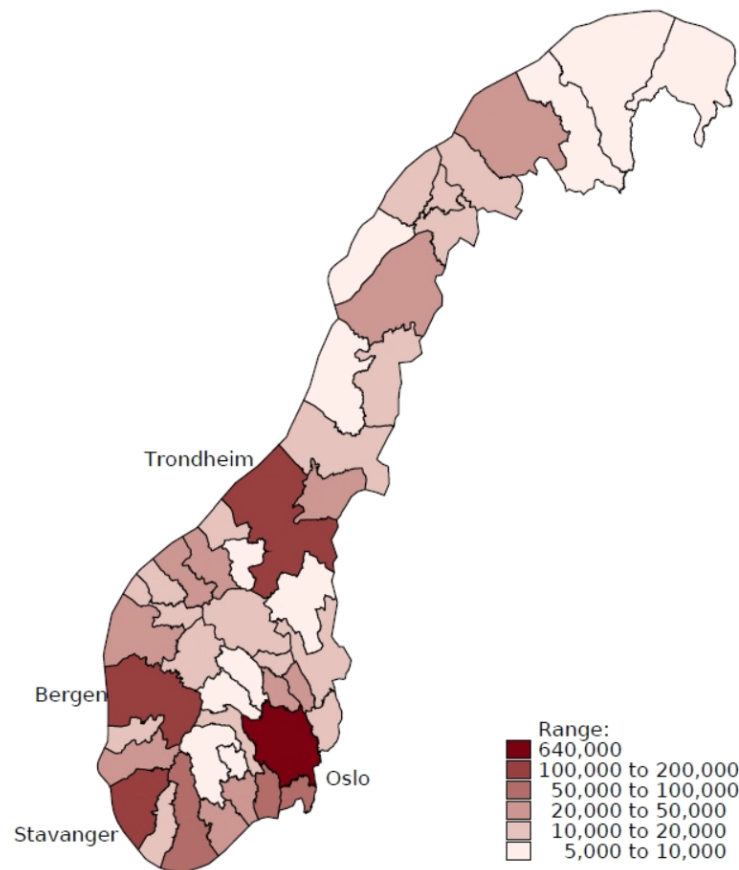
(c)

		Adjusted Share		
		Non-college	College	
Baseline	Non-college	50.7%	6.3%	57.0%
	College	4.1%	38.9%	43.0%
		54.9%	45.1%	

Note: This table compares the baseline classification of occupations with three alternative approaches to classifying occupations. Panel (a) uses ILO's classification of occupations into college and non-college using the first digit. Panel (b) classifies an occupation as college if the share of college workers is larger than 50%. Panel (c) uses a similar approach to Panel (b), but compares the share of college workers in an occupation to the overall share of college workers in the labor market. College worker shares and relative cell-sizes are computed using hours worked.

A.2 Commuting Zones

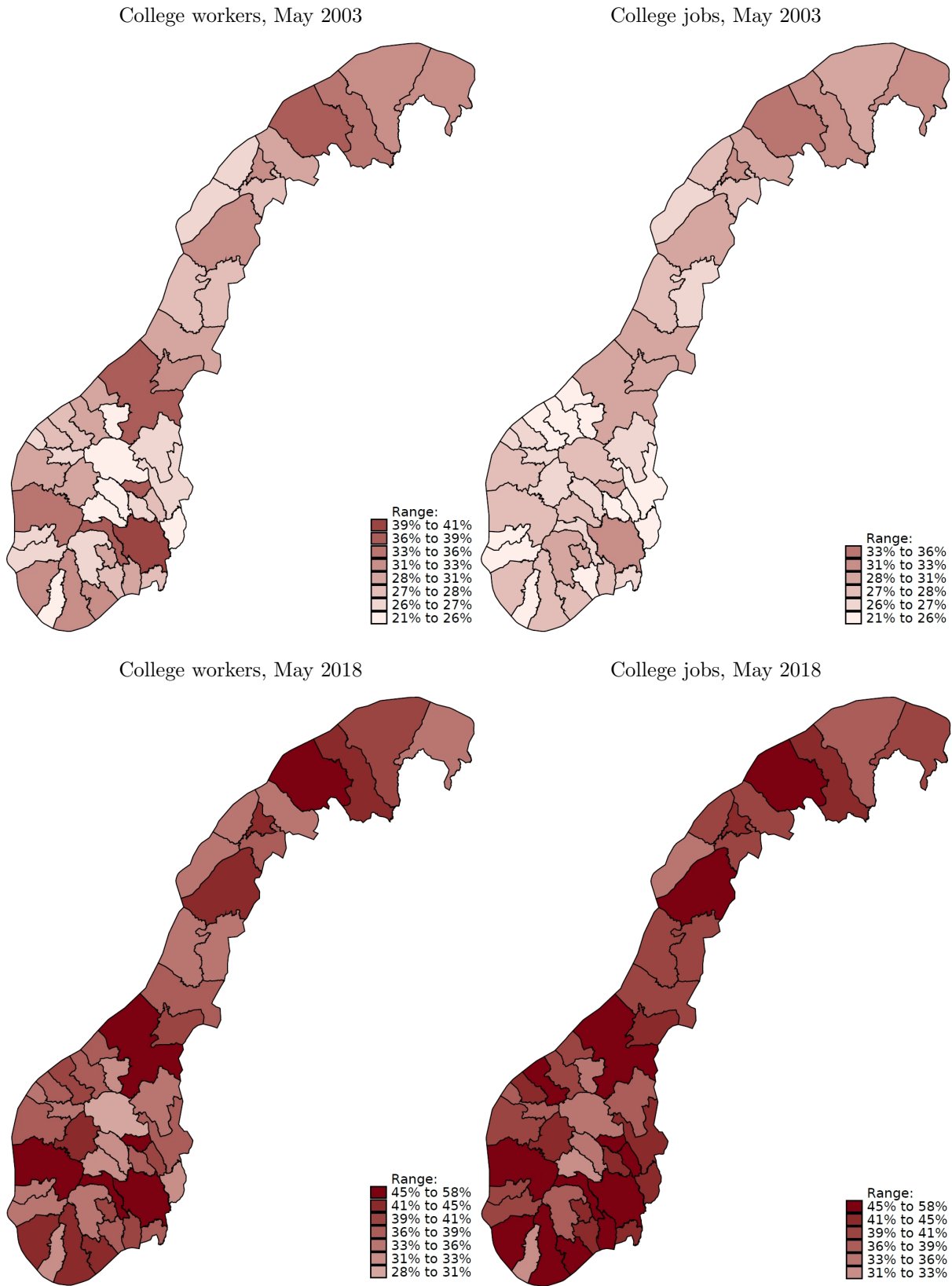
Figure 15: Commuting Zones and Number of Workers



Note: The map displays the number of workers within each commute zones in May 2010. The total number of workers in this month is 1.9 million, meaning that 34% of workers lived in the largest region (Oslo). The three commuting zones with the largest number of workers after Oslo, (Bergen, Stavanger and Trondheim) consisted of 6 – 7% of workers each.

B College Workers and Jobs Across Commuting Zones

Figure 16: College Worker and Job Share Across Commuting Zones



Note: This figure displays the share of college workers and jobs in 2003 and 2018, measured in hours worked.

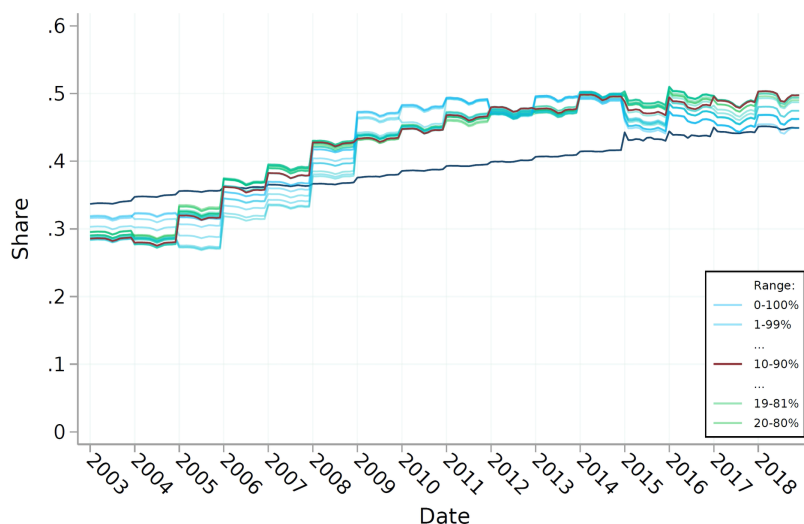
C Occupations with High and Low College Shares

The baseline method classifies occupations using estimated college shares, except for occupations with less than 10 or more than 90 percent college workers. This section documents the robustness of the main results with respect to changing how these occupations are classified.

Figure 17 shows how the share of college jobs varies with the choice of which occupations to classify using the share and which occupations are classified using estimated premiums. The baseline classifies occupations with a share of college workers in the range 10 to 90 percent, indicated with the red line in Figure 17. The remaining lines show the share of college jobs corresponding to varying this range.

The share of college jobs increases more than the increased share of college workers regardless of the threshold. However, the increased share of college jobs falls from 20 to 13 percentage points when classifying all occupations using estimated premiums. This is possibly due to the large amount of noise in estimated premiums for occupations where almost all workers are either college or non-college workers.

Figure 17: Sensitivity of College Job Share to Classification Method

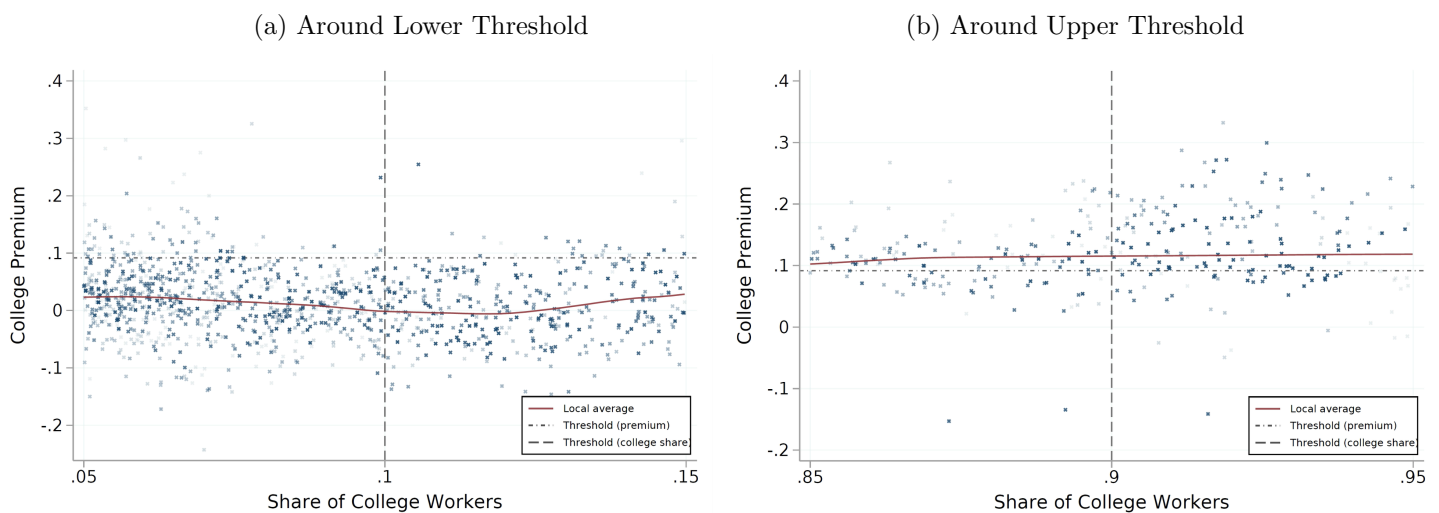


Note: This table documents sensitivity of the share of college jobs to classification method. As a baseline, occupation-years with more than 90% (less than 10%) college workers are classified as college (non-college) occupations. This figure shows how changing this threshold from the baseline case to alternative thresholds affects the calculated share of college workers.

Figure 18 documents estimated premiums around the thresholds of 10 and 90 percent college workers. On average, estimated premiums in occupations with around 10% (90%) college workers are below (above) the premium threshold 9.2%, meaning that the classification induced by using college worker shares corresponds to the classification induced by estimated premiums. Note that estimated premiums tend to get noisy as the college job share decreases (increases) towards zero (one).

Note that every occupation with more than 10 and less than 90% college workers has been merged with other occupations to make sure there are at least 50 college and non-college workers every year. This is not done for occupations with less than 10 or more than 90% college workers.

Figure 18: Estimated College Premiums Around Share Thresholds



Note: This figure documents estimated premiums for occupations with close to 10 and 90% college workers. Occupations with college shares below (above) 10% (90%) are classified using the share, while the remaining occupations are classified using estimated premiums.