Talent allocation, aggregate productivity, and income inequality: Evidence from Finland*

Juhana Siljander†

This paper is evolving. For the latest version, please click here.

November 18, 2021

Abstract

This paper studies the impact of occupational talent allocation on aggregate productivity and income inequality. I use Finnish administrative microdata to estimate a Roy model of occupational choice with unobservable skill and preference heterogeneity. I show that workers' sorting behavior changes have not driven aggregate productivity growth in the recent past. Holding skill distribution fixed, potential future gains for aggregate productivity by improving sorting are also limited. However, different time trends on occupational sorting patterns explain up to 40% of relative wage growth in certain occupations during 1995 - 2005. In particular, differential occupational sorting between genders explains 3.5 percentage points, or 16%, of the gender earnings gap. When accompanied by changes in the skill distribution, workers' occupational sorting behavior also matters for aggregate output. Removing gender differences in skills would lead to a 28% higher GDP effect than complete gender equalization with identical sorting patterns. I augment this analysis by leveraging the staggered implementation of the Finnish Comprehensive School Reform to discipline another counterfactual exercise. I use the model to decompose the reform effect into its skill and sorting components. I show that the reform's differential impact on women increased aggregate productivity by one percent, half of which was due to the sorting channel.

JEL codes: E24, J16, J11

Keywords: Talent Allocation, Factor Misallocation, Occupational Sorting, Income Inequality, Income Distribution, Gender Earnings Gap

*Special thanks to my thesis advisors Lars Peter Hansen, Stephane Bonhomme, Veronica Guerrieri, Thibaut Lamadon and Joseph Vavra for their guidance and support. I also appreciate helpful feedback from Erik Hurst, Magne Mogstad, Derek Neal, Tuomas Pekkarinen, Timm Prein, Antti Ripatti, Matti Sarvimäki, Alex Torgovitsky, Oskari Vähämaa, and participants of the Capital Theory and Applied Microeconomics seminars, and Economic Dynamics and Financial Markets working group at the University of Chicago, and the Macroeconomics seminar at Helsinki Graduate School of Economics. I acknowledge external funding from the Yrjö Jahnsson Foundation, and Statistics Finland and Aalto University for support in data access. Any errors that remain are my sole responsibility.

†Author's contact information: siljander@uchicago.edu

1 Introduction

During the last few decades, disentangling worker and firm heterogeneity for explaining observed variation in wages has been a central labor economics research problem. Many papers find that worker skill heterogeneity is especially important. (Abowd, Kramarz, and Margolis (1999); Bagger and Lentz (2019)) With heterogeneous skills, the allocation of workers across occupations could have aggregate effects. In this paper, I study how much talent (re)allocation contributes *quantitatively* to (i) aggregate productivity and (ii) income inequality. For each outcome, I separately ask how much this mattered in the recent past and how much can it plausibly matter in the future?

To motivate the question, I use administrative data from Finland to document significant differences in workers' occupational choices. For example, 23% of Finnish female workers worked in health and social care in 2005, whereas only 3% of men did so. I study, are skills that different, or are there other factors driving occupational sorting, and what is their significance for the economy at large.

I estimate a structural Roy (1951) model, where workers sort into occupations based on their occupational skills and preferences. These are assumed unobservable to the econometrician, as selection on unobservables is a potentially significant driver of occupational sorting. Indeed, I emphasize *unobservable* heterogeneity and *equilibrium effects* on wages. Concerning the latter, workers' sorting patterns determine the location of the labor supply curve in each occupational labor market. I show that shifts in these supply curves have a quantitatively substantial effect on equilibrium wages and occupational wage trends.

My analysis concentrates on talent allocation since the 1990s in the adult population of 30 - 55-year-olds. *Talent* thus describes the total worker aptitude encompassing all the educational and environmental factors affecting worker skills in adulthood. My data tracks occupations on the individual level, and I use job-movers to identify the model parameters (Abowd, Kramarz, and Margolis, 1999; Bonhomme, Lamadon, and Manresa, 2019). Labor demand for different occupations is modeled via a representative Cobb-Douglas producer, contributing to a labor market equilibrium, together with the Roy model of labor supply.

I use the model to disentangle shifts in the labor supply curves from other components affecting occupational wages. For example, increased occupational labor demand, factor-augmenting productivity growth, and reduced labor supply may all increase observed annual earnings. With a Cobb-Douglas producer, the relative labor demands for different occupations are identified from occupational wage shares. The Roy model, in turn, enables separating the non-wage component affecting labor supply from wage and productivity effects.

In my model, information about these labor supply effects is dispersed across a large number of sorting parameters for different heterogeneous agents. To aggregate this information, I consider a baseline counterfactual, where all workers sort into occupations based on skills. I then define an aggregate *preference wedge* in each occupation as the ratio of the counterfactual and actual equilibrium wages. An occupation with a low preference wedge has a relatively low equilibrium labor supply, and the increase in labor supply in the counterfactual economy drives down the occupational wage. Different preference wedges, therefore, describe occupational sorting patterns, conditional on wage. They measure and summarize the non-wage components of occupational selection, including true preferences, discrimination, and occupational entry barriers.[1] Time trends in the wedges show how occupational sorting patterns evolve. An increasing preference wedge is associated with suppressed wage growth due to time-varying occupational sorting behavior.[2]

My quantitative analysis consists of two parts. The first part discusses the effects of the sorting patterns, mainly in the past, considering the preference wedges and the above baseline counterfactual. In the second part, I allow for changes also in the skill distribution. I concentrate on gender differences and use the *Finnish comprehensive school reform* to provide exogenous variation in skills and sorting patterns to discipline my counterfactual analysis.

The first main finding of the paper is that holding adult employment and the distribution of skills constant, the effect of evolving talent allocation on aggregate productivity has been negligible in Finland during my study period of 1995 to 2005. When all workers sort into occupations purely on skills, aggregate output increases by approximately three percent annually. To the extent talent allocation matters for aggregate output, most potential gains seem to have been attained. Labor markets seem efficient in allocating workers into occupations based on their skills.

The second main contribution of the paper is to consider distributional outcomes and income inequality. I continue with the counterfactual, where everyone sorts into occupations based on their skills, and establish several findings. First, I document substantial relative differences in the counterfactual equilibrium wages, ranging from -57% to +20%, across different occupations in 2005. Some jobs have considerably lower labor supply due to a relatively significant preference wedge compared to others. These jobs are found dislikable or challenging to enter. Second, time trends in the preference wedges explain a substantial part of occupational wage growth, or lack thereof. Aggregate real earnings grew by 24% in Finland from 1995 to 2005. The time evolution of the preference wedges explains up to ± 10 percentage point differences in occupational wage trends. In other words, depending on the occupation, evolving sorting patterns increased or suppressed the occupational wage growth by up to 40%, relative to the average growth in the economy. Third, if all workers sorted into occupations purely on skills, Finland's gender earnings gap would have increased

[1] In the context of my model, "preferences" are defined broadly to include occupational discrimination or any other sorts of entry barriers affecting occupational choices.

[2] Throughout the paper, "occupational sorting" refers to sorting patterns, holding wages constant.

by 4.8 percentage points, or 22%, in 2005. Women, in particular, work in occupations with substantial wage premia due to the jobs' non-wage qualities.

In the second part of my quantitative analysis, I extend the analysis by also considering a counterfactual skill distribution, reflecting a potential long-run skill transformation. I study the relative importance of sorting patterns and skill growth for aggregate productivity. I concentrate on gender differences while allowing for unobservable within-gender heterogeneity. If women's skills matched men's, aggregate productivity would increase by 13.2%. However, the gender earnings gap would remain positive at 3.5% due to their differential occupational sorting patterns, which thus explain 16% of the gender earnings gap. Removing all gender heterogeneity would lead to a comparatively lower 10.2% increase in output, with lower earnings for both men and women. In these counterfactuals, the differential occupational sorting between men and women, conditional on wage, thus contributes to an additional 2.9 percentage point, or 28%, increase in aggregate productivity, disproportionally benefitting men. The analysis suggests that, when accompanied by substantial changes in the skill distribution, workers' sorting patterns may be essential drivers also for aggregate productivity.

Finally, I provide further evidence for the quantitative importance of this sorting channel by considering a specific natural experiment, the Finnish comprehensive school reform. The reform postponed the educational tracking of elementary school students into academic and vocational tracks by five years, from age 10 - 11 to age 15 - 16. The staggered implementation of the reform produces plausibly exogenous cross-sectional variation in workers' occupational choices and wages between genders. I shut down this variation to calibrate another counterfactual exercise on gender equality, thus using the reform to discipline the analysis.

Employing the staggered reform implementation, Pekkarinen (2008) documents a causal 4.1 percentage point reduction in the gender wage gap due to the reform in 2000. I replicate his analysis and find a 1.7 percentage point reduced-form reduction in the gender earnings gap in 2011 - 2017, with solid pre-trends. I further document a statistically significant 0.52 percentage point differential increase in women's probability of choosing health care occupations compared to men.

Given these substantial effects on wages and occupational sorting patterns, I use my structural model to decompose the reform impact on its *skill effect*, holding the selection of occupations constant, and the *sorting effect*, holding skills constant. I feed these partial equilibrium results into my general equilibrium model using the reform to provide causal interpretations for specific model parameters. Shutting down the reform effects, I find that its differential impact on women increased aggregate productivity by one percent. The skill and sorting components contributed 52% and 48% percent, respectively. The quantitative significance of the sorting channel in this natural experiment provides further confidence in the importance of occupational sorting when accompanied by significant changes in the skill distribution.

In conclusion, the paper shows that workers' occupational sorting patterns are fundamental drivers of wages and income inequality. While talent reallocation has not driven aggregate productivity growth in the past, workers' sorting behavior may substantially impact aggregate output if the skill distribution faces a material transformation, for instance, due to educational reform. In particular, talent reallocation may be important for aggregate productivity if gender differences diminish in the future.

The paper is organized as follows. Next I discuss my contributions to the literature. Section 2 lays out the roadmap for my analysis, discussing the analytical framework used to produce the counterfactual analyses. Section 3 develops the model. In Section 4, I discuss the identification and estimation of the model parameters. Section 5 shows the main counterfactual analyses, and in Section 6 I introduce the school reform with the accompanying analysis on gender differentials. Section 7 concludes.

Contributions to existing research. This paper builds on several existing strands of the literature. The efficient allocation of factor inputs is a classic economic research topic (Moscarini and Vella, 2008; Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009; De Vries, 2014; Kalemli-Ozcan and Sørensen, 2016; Dias, Marques, and Richmond, 2016).[3] Hsieh, Hurst, Jones, and Klenow (2019) document that improvement in talent allocation explains 20 to 40% of U.S. economic growth between 1960 and 2010. They concentrate on increased labor market participation of traditionally underrepresented groups, such as women and African-American men, mainly before the 1990s. Their central identification assumption is that differences in workers' innate talent between these groups have been constant since the 1960s, and the evolution in observed behavior can be attributed to reductions in labor market barriers.

I augment their work on several fronts. First, I consider income inequality in addition to aggregate productivity. Second, my analysis emphasizes unobservable heterogeneity within observable groups. I consider the effects of talent allocation since the 1990s in a relatively homogenous population with little racial disparities but significant unobservable differences. Third, my identification strategy allows for arbitrary time-varying unobservable heterogeneity in workers' occupation-specific skills and preferences. Workers' heterogeneity is taken at face value, and my counterfactual analysis uses a natural experiment, the Finnish comprehensive school reform, to discipline the quantitative exercise.

Another strand of the existing literature studies sorting heterogeneous workers into heterogeneous jobs (Shimer and Smith, 2000; Postel-Vinay and Robin, 2002; Lise and Robin, 2017; Hagedorn, Law, and Manovskii, 2017; Card, Cardoso, Heining, and Kline, 2018; Bonhomme, Lamadon, and Manresa, 2019). The classic approach is a Roy (1951) model, where heterogeneous workers sort into different industries based on their skills (Heckman and Sedlacek, 1985). However, the frictionless setup of this approach has faced challenges

[3] For an overview, see Restuccia and Rogerson (2017).

in matching micro-level patterns on labor mobility and wages.

In contrast, Postel-Vinay and Robin (2002) introduce rich heterogeneity in both workers and firms to a search model. They can explain interesting micro behavior such as wage cuts in job-to-job transitions. In a canonical search model (Pissarides, 1979; Diamond, 1982a,b; Burdett and Mortensen, 1998), a worker is matched with a firm, and Nash bargaining divides the output between them. This approach de-emphasizes equilibrium forces affecting wages in a traditional neoclassical supply-demand framework and concentrates on explaining the micro evidence (Mortensen, 2011).

I consider a discrete choice model of occupational choice with no labor market frictions, but I introduce rich, multidimensional heterogeneity in workers' skills. Similar to Lindenlaub (2017) and Lise and Postel-Vinay (2020), I depart from the assumption of a unidimensional skill common in macro models (Acemoglu and Autor, 2011) but still emphasize the equilibrium effects on wages and wage inequality. Expanding the macroeconomic forces in labor models is an active area of current research (Moscarini and Postel-Vinay, 2017).

In my model, workers sort into different occupations based on their heterogeneous occupational skills and (revealed) preferences, broadly defined to include labor market discrimination and occupational entry barriers. The preferences can be interpreted as compensating differentials for entering jobs, which are dislikable or difficult to enter (Lavetti and Schmutte, 2016; Sorkin, 2018; Lamadon, Mogstad, and Setzler, 2019).

Conditional on skills, these preferences determine the locations of the labor supply curves in each occupational labor market. Reduced labor supply in an occupation increases the equilibrium wage and affects income inequality between occupations. I show that these equilibrium effects are potentially important drivers of income inequality and, under certain circumstances, aggregate productivity when accounting for the general equilibrium externalities.

While I consider a producer employing workers in *different* occupations, my model does not have firm heterogeneity. The relative significance of worker and firm effects for the observed variation in wages constitutes a significant research stream. Several papers find worker heterogeneity especially important (Abowd, Kramarz, and Margolis, 1999; Bagger and Lentz, 2019).

Taber and Vejlin (2020) study the relative importance of pre-market skills, search frictions, occupational preferences, and human capital for explaining wage dispersion. Using Danish data, they show that shutting down job preferences between job types (establishments in the data) average earnings would rise by 0.20 log points. The misallocation literature often finds substantial misallocation at the establishment level; see Restuccia and Rogerson (2017). I consider occupational choices on a higher level of aggregation and find substantially smaller earnings impact but substantial effects for income inequality.

In the final part of my paper, I use a natural experiment, the Finnish comprehensive school reform, to discipline a quantitative exercise on the relative significance of skills and sorting effects for aggregate productivity. My analysis augments earlier research on the reform (Pekkarinen, 2008; Pekkarinen, Uusitalo, and Kerr, 2009; Pekkala Kerr, Pekkarinen, and Uusitalo, 2013). This analysis relates to an extensive literature on educational reforms and career choices.

Bell, Chetty, Jaravel, Petkova, and Van Reenen (2019) document the importance of "exposure effects" in career selection and show that women and disadvantaged youth are relatively unlikely to become inventors, potentially reflecting a suboptimal talent allocation. Abebe, Caria, and Ortiz-Ospina (2021) do a field experiment in Ethiopia to study how a monetary prize affects talent allocation. While I rely on an existing reform, I document the reform's causal impact on occupational choices. Like Khanna (2015), who considers returns to schooling for a major educational reform in India, I extend the reduced-form analysis by a general equilibrium and use my model to decompose the equilibrium effect on its skill and sorting components. Part of my analysis also considers the gender earnings gap, which has spanned a vast literature on its own (Goldin, 2014; Polachek, Tatsiramos, and Zimmermann, 2015). Again, I emphasize the equilibrium effects on wages.

2 Motivation and roadmap for the analysis

The results in this paper consist of three different parts. I first study the impact of talent allocation on aggregate productivity, then occupational wages, and finally, the gender earnings gap. In the final part, I also introduce the Finnish comprehensive school reform to discipline my quantitative analysis. This section motivates these questions in more detail and builds a roadmap for tackling these problems.

2.1 Analytical framework

My analysis builds on an estimated version of a structural Roy (1951) model. In section 3, I describe the whole dynamic setup, but to fix ideas, consider a static model of occupational choice, where a worker i maximizes her log-income conditional on preferences:

$$o^* \in \arg\max_o \{w_o + \alpha_{io} + \nu_{io}\}, \tag{2.1}$$

where her log-income consists of the occupational equilibrium log-wage w_o and her talent/skill, α_{io}, in that occupation. Additionally, the worker cares about her occupational preferences ν_{io}. Eventually, the model is estimated by using workers' wages and job mobility in adulthood. A worker's *talent*, therefore, describes her skills after education and takes into account all the possible environmental factors affecting her occupational aptitude.

To simplify the problem, I assume the worker heterogeneity is captured by a small number of latent types indexed by $j \in \{1, \dots, J\}$, and $\alpha_{io} = \alpha^{g(i)}_{j(i)o}$ and $\nu_{io} = \nu^{g(i)}_{j(i)o} + \zeta_{io}$, where ζ_{io} is a multivariate extreme-value distributed idiosyncratic taste shock, and $g(i)$ denotes the gender of worker i. This contributes to a standard discrete choice model.

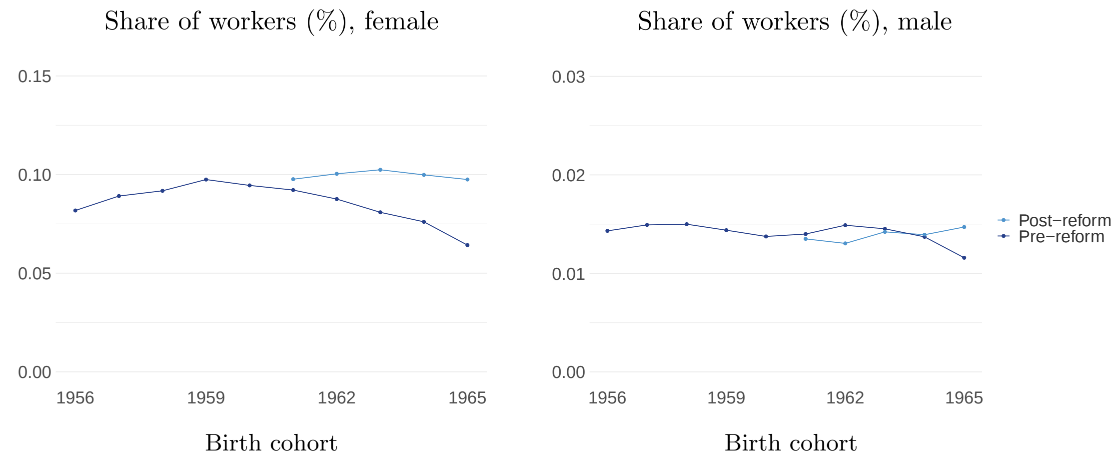


Figure 1: Occupational selection into health care jobs by gender

Notes: The picture shows the share of workers occupied in health care jobs for females and males, respectively, in 2011 − 2017. Beginning with the 1961 birth cohort, the shares are divided between areas affected and not affected by the school reform. The dark blue graphs include workers, who were unaffected by the reform, by birth cohort, whereas the light blue graphs include individuals, who studied under the post-reform school system. No causality is claimed and the issue of causal inference is revisited in section 6.

In each period, workers' only decision concerns their occupation. In the chosen occupation, they inelastically supply $\alpha^{g(i)}_{j(i)o}$ efficiency units of labor to a representative Cobb-Douglas producer, which aggregates the labor inputs to a single consumption good. Occupational wages w_o adjust to clear labor markets.

I use this framework to analyze the relative importance of the skill and sorting components, modeled via α^g_{jo} and ν^g_{jo}, on aggregate productivity and income inequality. Economic shocks may plausibly affect both. This raises the question of their quantitative importance. To consider a specific example, we may motivate the discussion with data from the Finnish comprehensive school reform, which I later in section 6 also use to discipline a particular counterfactual exercise.

Using the reform's staggered implementation, Pekkarinen (2008) documents that the reform reduced the gender wage gap by 4.1 percentage points in 2000. I show in section 6 that the reform also had a differential impact on workers' occupational choices between genders. Most notably, women's probability of choosing a high-education health care job relative to men's increased by 0.52 percentage points. Figure 1 documents this in the aggregate data, and in section 6, I discuss causal evidence leveraging an event-study setup. Disentangling these sorting effects from skill growth and decomposing their roles for aggregate productivity and income inequality are central research questions of this paper.

Indeed, workers have different occupation-specific skills $\alpha^{g(i)}_{j(i)o}$ and my main research question concerns their allocation to these occupations o. Employers observe the skills, which affect earnings and occupational sorting directly. Additionally, workers have other factors impacting their occupational choices. These are modeled by the sorting parameters ν_{jo}^g, which are my main objects of interest. With a given skill distribution, these parameters determine the allocation of talent relative to a frictionless baseline, where workers sort into occupations based on productivity.

My analysis faces two main challenges. First, I need to identify shocks on sorting (ν_{jo}^g) from shocks on skills (α_{jo}^g). Second, the information captured in the distribution of sorting parameters $\{\nu_{jo}^g\}_{j,o,g}$ is highly dispersed and complicated to analyze.

2.2 Identification

Relying on Bonhomme et al. (2019), I use both job-movers' and job-stayers' data on earnings and occupational choices to identify the model parameters. A fundamental identification assumption, detailed in Section 4, is that movers from occupation o to o' are different from movers from occupation o' to o.

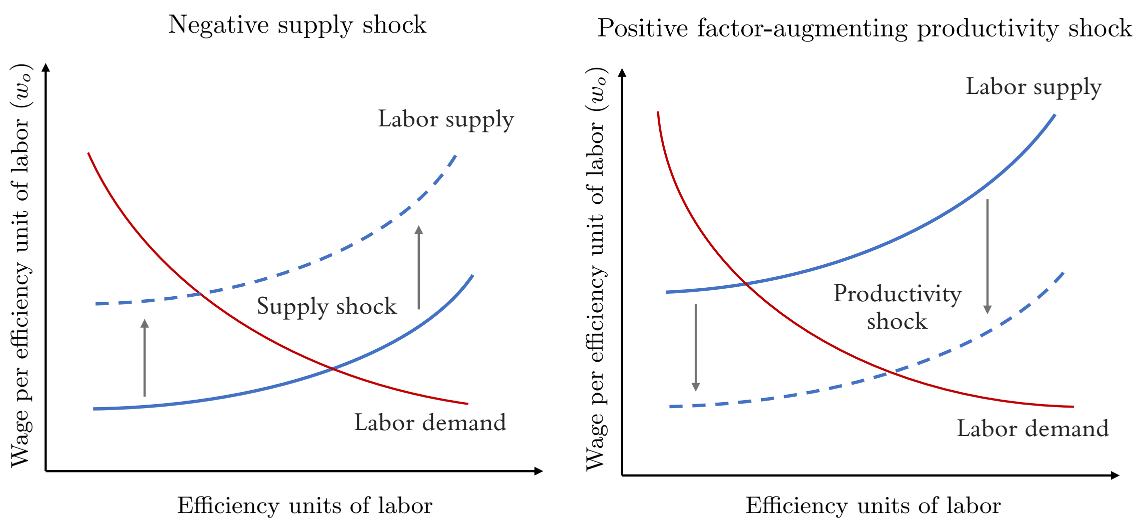


Figure 2: Partial equilibrium reduction in labor supply (ν_{jo}), and increase in skills (α_{jo})

Notes: The picture shows two types of labor market shocks, which both increase observed annual earnings. On the left, I consider a reduction in the value of the non-wage component ν_{jo} of sorting for one types of agents in one occupation. The partial equilibrium effect in that occupation is a reduction in labor supply, which drives up the equilibrium wage w_o. With unit elastic labor demand, the wage bill remains unaltered (in partial equilibrium) and annual per worker earnings increase. On the right, I consider a positive productivity shock increasing the value of α_{jo} for one type of workers in one occupation. The shock increases the efficiency units of labor offered to that occupation. The equilibrium wage w_o is reduced, which incentivizes the unaffected, other types of workers to change occupations. In partial equilibrium, annual average earnings increase for workers, who choose to stay in this occupation, since wage bill remains unaltered due to the unit elastic labor demand. Data on earnings and job mobility is used to distinguish these effects from each other.

To illustrate the identification challenge, note that holding labor demand constant, both factor-augmenting productivity growth (increase in α_{jo}^g) and reduced labor supply (decrease in ν_{jo}^g) may increase observed annual earnings.[4] Figure 2 provides a graphical illustration of this identification challenge for disentangling labor supply and productivity shocks. The figure depicts the partial equilibrium effects of a negative labor supply shock (left) and positive factor-augmenting productivity shock (right) in one occupation for a particular worker type.

A negative supply shock by a worker type shifts the labor supply curve inward as the affected workers leave the occupation. This reduction in labor supply increases equilibrium average earnings, assuming constant total wage bill due to a unit elastic labor demand. A positive factor-augmenting productivity shock, in turn, increases the supply of efficiency units of labor and suppresses the equilibrium wage. The lower salary incentivizes unaffected workers to leave the occupation. With a constant total wage bill, the average earnings increase for the workers staying in the field.

While both types of shocks lead to increased average earnings, the job mobility patterns associated with these shocks are entirely different. With a negative supply shock, wages increase for everyone staying in the job as the affected workers leave the occupation. With a productivity shock, wages rise for the workers with increased productivity but decrease for all other workers, who are incentivized to leave the profession. Leveraging these differential mobility and earning patterns allows for disentangling the sorting parameters (ν_{jo}^g) from wage effects ($w_o + \alpha_{jo}^g$).

2.3 Counterfactual analyses

The information about the labor supply shocks is highly dispersed across all the sorting parameters $\{\nu_{jo}^g\}_{j,o,g}$. To summarize this information, I consider a counterfactual, where these occupational preference parameters in equation (2.1) are set to zero, $\nu_{io} = 0$, for everyone. This counterfactual corresponds to a frictionless baseline, where all workers sort into occupations based on their skills. I solve the model under both parameterizations and define a *preference wedge* as the ratio of the counterfactual and actual equilibrium wages as

$$\tau_o = \frac{W_o^{counter}}{W_o} - 1.$$

These wedges now aggregate and summarize the potentially very complicated sorting behaviors captured in the dispersed set of sorting parameters $\{\nu_{jo}^g\}_{j,o,g}$. In a static counterfactual, these preference wedges represent the aggregate revealed preferences for each occupation, conditional on wage. These encompass the true occupational preferences but also occupational discrimination and entry barriers. Time trends of these wedges describe

[4] With a Cobb-Douglas producer differential labor demand shocks across the different occupations can be identified from occupational wage shares.

how changes in the sorting patterns affect occupational wage growth.[5] The preference wedges are my main object of interest for considering the quantitative role of occupational sorting patterns for occupational wage trends in Section 5.

2.4 The gender earnings gap

The final part of the paper discusses gender equality. I consider a counterfactual, where I remove all the gender differences in the skill and preference parameters and report the results. I show that differential sorting, conditional on wage, between the genders is essential for aggregate productivity, given such an extensive transformation in the skill distribution. However, removing all gender heterogeneity is perhaps unrealistic and, as such, an uninteresting counterfactual. Therefore, I use the Finnish comprehensive school reform to produce an alternative calibration for the counterfactual analysis. The reform's staggered implementation offers a causal interpretation for specific model parameters. Estimating the reform effects within the model thus allows for an alternative and arguably more realistic approach for disciplining the calibration of the counterfactual exercise on gender differences. Indeed, leveraging the exogenous variation in wages and occupational sorting provided by the reform, I use my model to decompose its productivity effect on the skill and sorting components.

3 Model

This section presents the quantitative model. I consider a representative final goods producer, which uses constant returns to scale Cobb-Douglas production technology to combine labor in different occupations to a single consumption good, which is the numéraire in the economy. The firms is owned by the workers, who choose the occupation they work in using a multinomial discrete choice Roy (1951) model. Thus, each period, each worker has one unit of indivisible time, which is used for production following the worker's occupational choice. Occupational wages adjust endogenously to clear markets. Workers are heterogeneous and belong to one of a finite number of types. Types differ in their productivity and occupational preferences.

In principle, the model is dynamic and agents are forward-looking, but given the assumptions specified in the following subsections, it turns out that the model may be solved period-by-period. Therefore, for all practical purposes, the model is a static one. The dynamic set-up is due to two reasons: (i) identification and estimation of model parameters relies on job movers, which is a meaningless concept in a static model, and (ii) estimation of the reform effect by a proper comparison requires the economy to be populated by both

[5] Note that occupational wage growth cannot be directly used to measure sorting effects, since it faces the identification problem discussed in Section 2.2: wage growth alone does not allow for disentangling productivity growth from sorting effects.

pre-reform and post-reform workers; this never holds in a model steady-state, since in the end all workers have been treated by the reform.

3.1 Set-up

Time is discrete and horizon is infinite. I consider an overlapping generations model, where each period $t = 0, 1, 2, \dots$, a new worker cohort of mass N_t is born. Workers live for T periods. The economy has a single representative firm, which employs workers in O distinct occupations, indexed by o. Workers belong to one of J types, indexed by j. The types are unobservable to the econometrician, but observable to the agents in the model. For each worker, we also observe birth cohort, region of birth and gender, which are indexed, respectively, by b, r and g.

3.1.1 Workers

Agents consume a consumption good C. An agent i of type $j(i)$ with observable characteristics $b(i), r(i)$ and $g(i)$ has period utility function

$$u_{it}(C, o) = \log(C) + \nu_{j(i)ot}^{b(i)r(i)g(i)} + \zeta_{iot},$$

where ζ_{iot} are multivariate extreme-value distributed preference shocks, which are independent across agents and time. Workers have preferences ν_{jot}^{brg} for working in occupation o. These include barriers, such as discrimination, for working in a specific occupation, and may depend both on the unobservable type j and the observable characteristics b, r and g.

3.1.2 A representative firm

A representative firm maximizes profits by solving

$$\begin{aligned} \pi_t = \max_{l_o} \quad & Y_t - \sum_{o=1}^{O} W_{ot} l_o \\ & Y_t = e^{z_t} F(l_1, \dots, l_O), \end{aligned}$$

where W_{ot} is the equilibrium wage in occupation o for an efficiency-unit of labor, F is the production function using labor from each occupation, and z_t is a vector of TFP shocks:

$$z_{t+1} = \rho z_t + b_{t+1} + \xi_{t+1}, \tag{3.1}$$

for innovations ξ_t. It follows that labor demand for occupation o is given by

$$e^{z_t} \frac{\partial F}{\partial l_o} = W_{ot}. \tag{3.2}$$

I assume the firm has Cobb-Douglas technology $F(l_1, \dots, l_O) = \prod_{o=1}^{O} l_o^{\gamma_{ot}}$, with $\sum_{o=1}^{O} \gamma_{ot} = 1$ for every t.

3.1.3 Vacancies and job search

Each period the firm post vacancies for each occupation o at a wage W_{ot} for an efficiency unit of labor. Workers maximize their utility by freely choosing their occupation. Workers are heterogeneous in their occupation-specific skills. After choosing an occupation, a worker of type j with characteristics b, r and g inelastically supplies

$$L_{iot} = \exp\left[\alpha_{j(i)ot}^{b(i)r(i)g(i)} + \varepsilon_{iot}^{b(i)r(i)g(i)}\right]$$

efficiency units of labor, where $\varepsilon_{iot}^{brg} \sim N\big(0, \sigma_{j(i)o}^{brg}\big)$ are independent productivity shocks across agents and time. Thus, L_{iot} represents the productivity of agent i in occupation o at time t.

3.1.4 Worker problem

A worker of type j with characteristics b, r and g, born at year $t+1$, maximizes her expected lifetime utility

$$\mathbb{E}\sum_{\tau=1}^{T}\beta^{\tau}[\log(C_\tau) + \nu_{jo\tau}^{brg} + \zeta_{io\tau}]$$

subject to a budget constraint given by $C_\tau \leq W_{io,t+\tau} + \Pi_{t+\tau}$, where Π_t denotes the worker's share of total firm profits, which are evenly distributed across agents, and $W_{iot} = W_{ot}L_{iot}$ is the worker's salary.

Worker Bellman equation before output, but after observing the preference shock ζ_{iot} is given by

$$V_{it}^{brg}(z) = \max_{o' \in \{1,\ldots,O\}} \; v_{it}^{brg}(o' \mid z) + \zeta_{io't},$$

where

$$v_{it}^{brg}(o \mid z) = \mathbb{E}_{\varepsilon}[\log(W_{iot}) + \nu_{j(i)ot}^{brg}] + \beta\mathbb{E}_{\zeta',z'}[V_{i,t+1}^{brg}(z') \mid z].$$

3.1.5 Occupational sorting

The choice probabilities for a worker of type j with characteristics b, r and g, for choosing occupation o is now given by

$$\mathbb{P}_{jt}^{brg}(o_i = o \mid z, \mathbf{w}) = \frac{\exp[v_{jt}^{brg}(o \mid z)]}{\sum_{o'=1}^{O}\exp[v_{jt}^{brg}(o' \mid z)]},$$

where $\mathbf{w} = (w_1, \ldots, w_O)'$ is the vector of occupational log-wages $w_o = \log W_o$. This implies

$$\log\left[\frac{\mathbb{P}_{jt}^{brg}(o_i = o \mid z, \mathbf{w})}{\mathbb{P}_{jt}^{brg}(o_i = o' \mid z, \mathbf{w}}\right] = w_{ot} - w_{o't} + \alpha_{jot}^{brg} - \alpha_{jo't}^{brg} + \nu_{jot}^{brg} - \nu_{jo't}^{brg}. \tag{3.3}$$

I normalize $\nu_{j1t}^{brg} = 0$ for all j and b, r, g and t. Note that by (3.2), we have

$$\log W_{ot} - \log W_{o't} = \log\left[\frac{\gamma_{ot}}{l_{ot}}\right] - \log\left[\frac{\gamma_{o't}}{l_{o't}}\right]$$

It follows that, if the equilibrium labor supply L_{sot} does not depend on z_t, then the occupational choice probabilities $\mathbb{P}_{jt}^{brg}(o_i = o \mid z, \mathbf{w})$ do not depend on z_t, either, which then justifies the fact that L_{sot} does not depend on z_t. It is, therefore, a natural guess to be verified that the aggregate state does not affect the equilibrium labor supplies. In my empirical implementation, I classify occupations to a small number of clusters. Since each sector employs workers from all occupational clusters, sectoral shocks affect all the occupational wages the same.

Lemma 3.1. *The occupational choice probabilities are independent of the aggregate shocks z_t: $\mathbb{P}_{jt}^{brg}(o \mid z, \mathbf{w}) = \mathbb{P}_{jt}^{brg}(o \mid \mathbf{w})$, but moreover, there is no TFP effect through $\mathbf{w}$, either.*

For the proof, I refer to Appendix A. Finally, I also allow for time-varying $\gamma_{ot}, \alpha_{jot}^{brg}$ and ν_{jot}^{brg}, which affect the equilibrium, but these are assumed deterministic.

3.1.6 Labor supply

Let m_{jt}^{brg} denote the mass of workers of type j with characteristics b, r and g, at time t. Mass of workers in each occupation is now given by

$$E_{ot}(\mathbf{w}) = \sum_{j,b,r,g} m_{jt}^{brg} \mathbb{P}_{jt}^{brg}(o \mid \mathbf{w}).$$

In order to determine the supply of efficiency units of labor, we need to weight the occupational choice probabilities by the type-specific productivity parameters. Hence, the effective human-capital-weighted labor supply is

$$L_{ot}(\mathbf{w}) = \sum_{j,b,r,g} m_{jt}^{brg} \mathbb{P}_{jt}^{brg}(o \mid \mathbf{w}) \exp\left(\alpha_{jot}^{brg} - \frac{(\sigma_{jot}^{brg})^2}{2}\right).$$

The total employment and labor supply are, respectively, given by

$$E_t(\mathbf{w}) = \sum_{o=1}^{O} E_{ot}(\mathbf{w}) \quad \text{and} \quad L_t(\mathbf{w}) = \sum_{o=1}^{O} L_{ot}(\mathbf{w}).$$

3.2 Equilibrium definition

A competitive equilibrium is a set of occupational wages $\mathbf{w}_t = \{w_{ot}\}_{o=1}^{O}$ and decision rules $\{\mathbb{P}_{jt}^{brg}(o \mid \mathbf{w})\}_{j,o,b,r,g}$ and $\{l_{ot}(\mathbf{w})\}_o$ such that

1. the occupation demand functions l_{ot} solve the firm's problem (3.2).

2. the workers occupation choice decisions (3.3) are utility maximizing,

3. the final goods market clears: $Y_t = C_t = \sum_o W_{ot} L_{ot} + \Pi$, and

4. occupation-specific labor markets clear

$$L_{ot} = l_{ot} \quad \text{for all } o \text{ and } t.$$

4 Identification and estimation

This section discusses the identification and estimation of the labor supply model parameters. In particular, I concentrate on the identification of the unobservable skill and sorting parameters, α_{jot}^{brg} and ν_{jot}^{brg}, respectively. The estimation is done by a two-step maximum likelihood similarly to the discussion in Bonhomme et al. (2019).

4.1 Skill and preference distributions

Let $o_t(i)$ denote the occupation of a worker i in period t. Observed annual income ω_{it} is then given by

$$\log \omega_{it} = w_{o_t(i)t} + \alpha_{j(i)o_t(i)t} + \varepsilon_{it}, \tag{4.1}$$

where I have suppressed the dependence of α_{jot} on the observable characteristics b, r and g. The identification argument is developed for each set of observables separately.

This model closely follows that of Bonhomme et al. (2019), who discuss sorting of workers to firms with unrestricted mobility, whereas I consider occupational choice in a Roy model. My set-up follows almost exactly that of Grigsby (2019), with the main difference that I allow for time-varying skills α_{jot}. In the labor supply model, Grigsby (2019) also assumes that the sorting parameters do not have unobservable heterogeneity, but instead $\nu_{jot} = \nu_o$. This is the key difference that allows me to consider heterogeneous sorting and its effects on aggregate quantities.

Observe that the identification of α_{jot} separately from w_{ot} is generally not possible. This is due to the fact, that we have not specified the units of measurement for labor supply. The scale of w_{ot} depends on, whether we measure labor supply e.g. in years or hours. Building on Bonhomme et al. (2019) I establish the identification for $\bar{\alpha}_{jot} \equiv w_{ot} + \alpha_{jot}$, and then fix units by normalizing

$$\sum_{j,b,r,g} m_{jt}^{brg} \exp\left(\alpha_{jot}^{brg}\right) = 1, \tag{4.2}$$

for every o and t. Note that since this normalization is done period-by-period, the units of measurements are potentially changed each period. This prevents from considering dynamic counterfactuals for the skill parameters α_{jot}, and only static comparisons within the same normalization are possible. However, this places no restrictions on ν_{jot}, which are my main object of interest.

The identification of the unobservable skill parameters relies on occupation changers. Let $m_{it} := \mathbf{1}\{o_t(i) \neq o_{t-1}(i)\}$ be a mobility indicator describing, whether a worker i switched her occupation between periods $t-1$ and t. Furthermore, I denote the history of realization for any random variable X up to period t be $X^t = \{X_{i1}, \dots, X_{it}\}$. Finally, I denote by $p_{oo'}(j)$ the type distribution of agents with occupations o and o' during subsequent periods, and by $q_o^{brg}(j)$ the type distribution of agents in occupation o.

Definition 4.1. A connecting cycle of length R is a pair of sequences of occupations $(o_1, \dots, o_R)$ in period t, and $(\widetilde{o}_1, \dots, \widetilde{o}_R)$ in period $t+1$, with $o_{R+1} = o_1$, such that $p_{o_r,\widetilde{o}_r(j)} \neq 0$ and $p_{o_{r+1},\widetilde{o}_r}(j) \neq 0$ for all r in $\{1, \dots, R\}$ and j in $\{1, \dots, J\}$.

Following Bonhomme et al. (2019) and Grigsby (2019), I make the following identification assumptions.

Assumption 4.2 (Identification Assumptions)**.**

1. (Mobility determinants) The realization of mobility $m_{i,t+1}$ and the choice of occupation in period $t+1, o_{t+1}(i)$, are independent of the history of productivity shocks ε_i^t, conditional on the latent and observable worker characteristics $j(i), b(i), r(i), g(i)$, and their history of moves and occupation choices $m_i^t, o^t(i)$.

2. (Serial independence) The realization of period $t+1$'s productivity shock $\varepsilon_{i,t+1}$ for worker i is independent of the history of shocks ε_i^t and occupation choices $m_i^t, o^t(i)$, conditional on the worker's current occupation choice $o_{t+1}(i)$, the latent and observable worker characteristics $j(i), b(i), r(i), g(i)$ and worker mobility decision $m_{i,t+1}$.

3. (Connecting Cycles) For all observable worker characteristics b, r, g, and for any two occupations o and $o' \in \{1, \dots, O\}$, within the set of agents i satisfying $b(i) = b, r(i) = r$ and $g(i) = g$, there exists a connecting cycle $(o_1, \dots, o_R), (\widetilde{o}_1, \dots, \widetilde{o}_R)$ such that $o_1 = o$ and $o_r = o'$ for some r, and such that the scalars $a(1), \dots, a(J)$ are all distinct, where

$$a(j) = \frac{p_{o_1\widetilde{o}_1}(j)p_{o_2\widetilde{o}_2}(j) \cdots p_{o_R\widetilde{o}_R}(j)}{p_{o_2\widetilde{o}_1}(j)p_{o_3\widetilde{o}_2}(j) \cdots p_{o_1\widetilde{o}_R}(j)}.$$

   In addition, for all o, o', possibly equal, there exists a connecting cycle $(o'_1, \dots, o'_R)$, $(\widetilde{o}'_1, \dots, \widetilde{o}'_R)$ such that $o'_1 = o$ and $\widetilde{o}'_r = o'$ for some r.

4. (Full Rank) There exist finite sets of M values for ω_t and ω_{t+1} such that, for all $r \in \{1, \dots, R\}$, the matrices $A^{brg}(o_r, \widetilde{o}_r)$ and $A^{brg}(o_{r+1}, \widetilde{o}_r)$ have rank J, where $A^{brg}(o, o')$ has $(\widetilde{\omega}_1, \widetilde{\omega}_2)$ element

$$\mathbb{P}(\omega_{it} \leq \widetilde{\omega}_1, \omega_{i,t+1} \leq \widetilde{\omega}_2 \mid o_t(i) = o, o_{t+1}(i) = o', m_{i,t+1} = 1, b(i) = b, r(i) = r, g(i) = g)$$

These assumptions allow for estimating the unobservable skill and sorting parameters by maximizing the likelihood function. Note that, while in general parts 2 and 3 of the above assumption may be relaxed, in this form they necessarily hold under the assumption that the model discussed in Section 3 is correctly specified. That said, under these assumptions, the skill distribution is identified – and can be estimated – even for a misspecified equilibrium model.

Assumption 4.2 has four components. The first part requires that, conditional on workers' history of occupations and characteristics $j(i), b(i), r(i), g(i)$, their occupational choice $o_{i,t+1}$ is uncorrelated with their wage draws ε_i^t. In particular, this implies that idiosyncratic productivity shocks cannot affect occupational decisions. This rules out models, where idiosyncratic wage draws incentivize switching occupation. Correspondingly, I assume the occupational choice is done prior to observing the ε_{it}.

The second item requires serial independence from the productivity shocks $\varepsilon_{i,t+1}$ conditional on occupational choice $o_{t+1}(i)$. While this follows directly from the model structure discussed in section 3, this is generally violated by typical workers, at least outside of the gig economy. In my implementation of the method, I will assume each period to last for five years, which alleviate serial independence in idiosyncratic wages to some extent.

The third piece is a somewhat standard-looking assumption in the labor literature: any two occupations must belong to a connecting cycle for every worker type. It implies graph connectedness in the sense of Abowd, Kramarz, and Margolis (1999). Note, however, that it does not imply bilateral worker flows between every pair of occupations (o, o').

Finally, the last item is a standard rank condition, essentially requiring that the worker types are not drawn from the same distribution.

The following version of Theorem 1 from Bonhomme et al. (2019) now provides the identification of the unobservable heterogeneity in the model. Proofs are given in Appendix B.

Theorem 4.3. *(Identification) Let Assumption 4.2 hold. Suppose the individual wages w, occupations o, birth cohort b, region of birth r and gender g are observed during two periods. Then, up to labeling of the types j, the parameters $\bar{\alpha}_{jot}^{brg}, \sigma_{jo}^{brg}$ and ν_{jot}^{brg} are identified for all j and o. Moreover, for all pairs (o, o'), $o \neq o'$, for which there are occupation switchers from o to o', $p_{oo'}^{brg}(j)$ is identified for all j, for the same labeling and fixed b, r and g. Last, the type proportions $q_o^{brg}(j)$ are all identified, again for the same labeling and fixed b, r and g.*

In the above theorem I allow for time-varying skill and preference distributions $\bar{\alpha}_{jot}^{brg}$ and ν_{jot}^{brg}. It is well known that job movers can be used to identify individual and firm/occupation fixed effects as shown by Abowd et al. (1999).[6] Bonhomme et al. (2019) generalize this for worker-job interactions, such as in the above model. The key mobility assumption embedded

[6] Relying on similar assumptions than those in Assumption 4.2, Abowd, Kramarz, and Margolis (1999) prove the identification of

$$\log \omega_{it} = \alpha_i + \psi_{k_t(i)} + \varepsilon_{it},$$

where $\{\alpha_i\}_{i=1}^N$ and $\{\psi_k\}_{k=1}^K$ are individual and firm fixed effects, and $k_t(i)$ is the firm, where agent i is employed at time t. Using job stayers, in addition to job movers, provides identification also for time-varying firm fixed effects $\{\psi_{kt}\}_{k,t}$.

in part three of Assumption 4.2 is that workers moving from occupation o to o' are different from workers from moving from occupation o' to o.

Recall that the classic argument of Abowd et al. (1999) does not allow for time-varying parameters. This turns out to be a minor technicality; for instance, in the fixed effects model of Abowd et al. (1999), one may use *job-stayers* to identify time-varying firm fixed effects. Similarly, job stayers allow for time-varying estimation in the set-up of Bonhomme et al. (2019). Intuitively, if using job movers is enough to identify time-independent skill parameters α_{jo}, comparing the mean wages among job-stayers between the two periods used in the estimation enables identifying the differential time-components in α_{jot}.

Having time-varying parameters is important, since it allows for controlling occu-pation-specific shocks between the estimation periods. In my implementation of the estimation, in order to generate enough mobility for the estimation, I use 5-year intervals between the two periods required for the estimation. There is, however, no guarantees that the skill and sorting parameters remain constant during these five years. In fact, that is unlikely. Allowing for time-varying parameters enables to control for changes in the parameter values during the estimation interval.

Moreover, at this level of generality, there is no interaction, whatsoever, between different observable groups (b_1, r_1, g_1) and (b_2, r_2, g_2). No statement is made about the labeling or equivalence of types between these groups. Therefore, comparing individuals or their types between these groups is essentially impossible. In fact, the analysis could be done completely separately across different groups of observable characteristics. However, later I will parameterize the model further in order to make assumptions, which allow for across-group comparisons. For now, I rely on the following distributional assumptions.

Assumption 4.4 (Distributional Assumptions)**.**

1. The idiosyncratic productivity shocks ε_{it} are independent across agents and over time, and they are log-normally distributed with mean 0 and standard deviation σ_{jo}^{brg} for a worker i of type j, with occupation o at time t, and observable characteristics b, r and g.

2. The idiosyncratic taste shocks ζ_{iot} have multivariate extreme-value distribution and they are independent across agents and over time.

The first part of the above assumption assumes log-normality of wages within a type. Note that this does not imply that the wage distribution is log-normal, but only that the log-wage distribution is a mixture of normals. The second part of the above assumption implies serial independence of the taste shocks ζ_{iot}. This is a strong assumption in that it implies much more occupational mobility than observed in the data. The model is not built to explain labor mobility, but rather the dynamics of the model are only used to justify the estimation of the static counterfactuals, which require two-period data. Importantly, and contrary to Grigsby (2019), the above distributional assumptions do not imply (close to) random mobility, but the mobility patterns have substantial variation across agent types. This is due to the fact, that different types of agents have different sorting preferences captured in the differential values of ν_{jot}.

The parameter estimates are produced by maximizing the log-likelihood by using the expectation-maximization (EM) algorithm. This is done in two steps. First, I estimate the productivity parameters $\bar{\alpha}_{jot}^{brg}$ and σ_{jo}^{brg} for two subsequent periods, using two periods of data. I then use two-period rolling windows to estimate the effects for the whole study period of 2005 – 2009. Assuming worker types and wage realizations are independent across workers conditional on mobility indicators and occupational classes, the log-likelihood takes the form

$$\sum_{i=1}^{N} \sum_{b,r,g} \mathbf{1}\{b(i)=b, r(i)=r, g(i)=g\} \log \left[\sum_{j=1}^{J} p_{o_t(i)o_{t+1}(i)}^{brg}(j) \prod_{\tau=t}^{t+1} \phi(\omega_{i\tau}; \bar{\alpha}_{jo_\tau(i)\tau}^{brg}, \sigma_{jo_\tau(i)}^{brg}) \right].$$

In the second step, I estimate the sorting parameters ν_{jot}^{brg} along with the type distribution m_j^{brg}. Through the the lens of my model and using the above distributional assumptions, the probability for a type j agent with characteristics (b, r, g) to choose an occupation o is given by

$$\mathbb{P}_{jt}^{brg}(o) = \mathbb{P}_{jt}^{brg}\big(o; \{\bar{\alpha}_{jot}^{brg}, \nu_{jot}^{brg}\}_o\big) = \frac{\exp(\bar{\alpha}_{jot}^{brg} + \nu_{jot}^{brg})}{\sum_{o'=1}^{O} \exp(\bar{\alpha}_{jo't}^{brg} + \nu_{jo't}^{brg})}, \tag{4.3}$$

with normalization $\nu_{j1t}^{brg} = 0$ for every j, b, r, g and t. Denoting by $\widehat{\alpha}_{jot}^{brg}$ and $\widehat{\sigma}_{jo_\tau(i)}^{brg}$ the estimated values of $\bar{\alpha}_{jot}^{brg}$ and σ_{jo}^{brg}, respectively, the log-likelihood of occupational choices along with observes wages now takes the form

$$\sum_{i=1}^{N} \sum_{b,r,g} \mathbf{1}\{b(i)=b, r(i)=r, g(i)=g\}$$
$$\times \log \left[\sum_{j=1}^{J} m_j^{brg} \prod_{\tau=t}^{t+1} \phi(\omega_{i\tau}; \widehat{\alpha}_{jo_\tau(i)\tau}^{brg}, \widehat{\sigma}_{jo_\tau(i)}^{brg}) \mathbb{P}_{j\tau}^{brg}\big(o_\tau(i); \{\widehat{\alpha}_{jo\tau}^{brg}, \nu_{jo\tau}^{brg}\}_o\big) \right].$$

This is maximized over $\{m_j^{brg}, \nu_{jot}^{brg}\}_{j,t,o,b,r,g}$. Validation of the identification and estimation methodology using simulated data is discussed in Appendix B.[7]

[7] In this two-step process, the estimation of m_j^{brg} is produced at the same time as the estimation of ν_{jot}^{brg} and thus relies on the model of occupational choice being correctly specified. A more robust approach would be to introduce a third estimation step, where these weights are estimated using the cross-sectional data on wages and estimated values of $\widehat{\alpha}_{jot}^{brg}$ before estimating the sorting parameters. This approach is robust for mis-specifying the model of occupational choice. This will be included in the next major revision of the paper.

4.2 Data and Implementation

I use Finnish administrative data to estimate the model. Annual income is collected from tax filings and it is available at an annual frequency since 1993 for the whole Finnish population. I estimate the model on adult population with age between 30 and 55 using a 5-year interval between the estimation periods used to identify the skill and sorting distributions. In the latter estimation period the age of individuals in the sample is thus between 35 and 60. This age group is chosen to exclude most students and retirees, and it represents the core of the working age population.

The data on occupations is based on the International Standard Classification of Occupations (ISCO), and is available annually (end of year) only since 2004, and at 5-year intervals since 1995. In 1995, 2000 and 2004 – 2009 the classification uses ISCO-88, and from 2010 onwards ISCO-08.

The model of Section 3 includes two-dimensional heterogeneity: heterogeneous workers and heterogeneous occupations. Since I allow for largely arbitrary (although discrete) skill distribution, the identification argument requires substantial labor mobility within each type. This certainly fails at the individual level, and the individual heterogeneity needs to be discretized to form a small number of groups, which I call *types*. This is a true restriction on the data generating process (DGP).

In order to generate enough mobility, I also classify the occupations to a small number $O = 8$ clusters, by using the k-means algorithm. Bonhomme, Lamadon, and Manresa (2017, 2019) show that the k-means classification has good large sample properties in a wide class of models, including the dynamic discrete choice model. Alternatively, the clustering of occupations may be considered a restriction on the DGP; such a restriction is not needed, if we believe the k-means clustering to be a good approximation of the reality, relying on the large sample properties of the k-means clusters.

4.2.1 Occupational clustering

For implementing the k-means clustering, for each occupation in the ISCO classification, I determine workers' educational distribution. I denote by $h_{k,f}$ the share of workers employed at 5-digit occupation $k \in \{1, \ldots, K\}$ with highest degree attained at educational field $f \in \{1, \ldots, F\}$. The educational field classification is done at 1-digit level using the International Standard Classification of Education (ISCED-F 2011) standard. The k-means problem is given by

$$\min_{o(1),\ldots,o(K),\mathbf{H}_1,\ldots,\mathbf{H}_O} \sum_{k=1}^{K} \left[\sum_{f=1}^{F} [h_{k,f} - H_{O(k),f}]^2 \right]^{1/2},$$

which maps the K 5-digit occupations to $O = 8$ occupation groups. I produce the k-means grouping using the data from 2004 – 2009 and the 1960 – 1966 birth cohorts. These birth cohorts are chosen, since they represent the group of individuals most relevant for my reform

Table 1: Summary of occupational clusters by k-means

#	Broad category	Sample Occupations
1	Agriculture and forestry	Farmer, lumberjack, forest technician
2	Low-skill service	Cleaner, Personal aid, Cook, Waitress, Firefighter, Policeman, Hairdresser
3	Social skilled	Sales representative, merchant, primary school and kindergarten teacher
4	Tradespeople	Carpenter, Plumber, Electrician, Machinist
5	Manual work	Cargo driver, storage worker, maintenance worker, guard, baker
6	High-skill service	High school teacher, interpreter, priest, cultural manager, graphic designer
7	Clerical	Salesman, Secretary, Bookkeeper, Office clerk, Personal banker
8	Health and social care	Doctor, registered nurse, nurse assistant, social worker, physical therapist

analysis in Section 6. The middle aged birth cohorts are also likely to have finished their education.

Note that the k-means problem is a computationally demanding one and standard solution algorithms are not guaranteed to find the global minimum. However, the algorithms are known to be relatively well-behaved (Bonhomme et al., 2017). In Table 1 I report the most common occupations in each group. The clustering seems mostly intuitive, and based on these, I produce "broad category" labeling for each occupation-cluster.

4.3 Estimation results

The previous sections discussed the identification arguments and estimation of the model parameters. In my actual implementation, I will consider a number of different special cases of the general theory. Most notably, partially due to data limitations, I will concentrate on gender differences, and first consider the special case, where the unobservable heterogeneity is aligned only by gender: $\alpha^{brg}_{jot} = \alpha^{g}_{jot}, \nu^{brg}_{jot} = \nu^{g}_{jot}$ and $\sigma^{brg}_{jot} = \sigma^{g}_{jot}$ for all birth cohorts and regions.

Table 2 reports the parameter estimates for α^{g}_{jot} and σ^{g}_{jot} for $t = 2005$. Due to the normalization, the most rigorous comparison can be done within occupational clusters and between types. The skills have substantial variation. For instance, type 2 male workers succeed in health and social care jobs, but are at disadvantage against type 3 and 4 workers in most occupations. Type 4 workers, in particular, seem to be typical high-achievers, whereas type 1 and 5 workers tend to have substantial lower skills. That said, the comparative advantage of type 1 workers seem to be in clerical jobs.

Similar skill variation can be observed also for females. Type 10 individuals are good in clerical and other health and social care jobs, but other than that, they are generally low-achievers in other occupations. Type 6 workers are low-achievers across the occupational

Table 2: Type-occupation productivities $\exp(\widehat{\alpha}^g_{jo})$ and standard deviations $\widehat{\sigma}^g_{jo}$

		Worker type ($J = 5$), male					Worker type ($J = 5$), female				
#	Occupation category	1	2	3	4	5	6	7	8	9	10
1	Agriculture and forestry	0.41	1.23	1.80	0.11	0.78	0.66	1.09	1.68	0.18	0.26
2	Low-skill service	0.86	1.23	1.79	1.46	0.89	0.57	0.76	0.91	1.21	0.27
3	Social skilled	0.63	1.13	1.51	2.20	0.80	0.56	1.08	0.75	1.56	0.37
4	Tradespeople	0.60	1.09	1.44	2.04	0.84	0.60	1.08	0.81	1.60	0.45
5	Manual work	0.67	1.07	1.47	2.30	0.81	0.58	1.02	0.77	1.63	0.35
6	High-skill service	0.46	1.08	1.37	1.91	0.79	0.59	0.93	1.19	1.44	0.37
7	Clerical	1.00	1.02	1.42	2.18	0.74	0.54	0.89	0.71	1.66	1.19
8	Health and social care	0.63	2.38	1.14	0.89	0.71	0.51	0.89	0.71	0.49	1.53
Type proportions (m_j)		0.16	0.27	0.21	0.11	0.24	0.26	0.24	0.33	0.10	0.07

(a) This subtable documents the estimated skill parameters $\exp(\widehat{\alpha}^g_{jot})$ and within-gender type proportions m_{jt} for $t = 2005$ for both males and females. The parameters are scaled according to (4.2).

		Worker type ($J = 5$), male					Worker type ($J = 5$), female					
#	Occupation category	1	2	3	4	5	6	7	8	9	10	w_o
1	Agriculture and forestry	0.96	0.16	0.27	1.69	0.37	0.41	0.17	0.33	2.01	1.04	10.00
2	Low-skill service	1.14	0.10	0.17	0.01	0.22	0.35	0.09	0.10	0.20	1.12	10.18
3	Social skilled	1.11	0.14	0.14	0.21	0.18	0.44	0.16	0.15	0.22	1.50	10.30
4	Tradespeople	1.04	0.13	0.13	0.18	0.18	0.44	0.15	0.12	0.24	1.40	10.31
5	Manual work	1.11	0.13	0.15	0.23	0.18	0.41	0.14	0.14	0.30	1.54	10.32
6	High-skill service	1.02	0.13	0.12	0.20	0.24	0.48	0.19	0.09	0.13	1.23	10.35
7	Clerical	1.25	0.14	0.16	0.26	0.20	0.53	0.12	0.11	0.37	0.16	10.40
8	Health and social care	1.00	0.31	0.18	0.10	0.20	0.38	0.11	0.12	1.14	0.35	10.41

(b) This subtable documents the estimated standard deviation parameters $\widehat{\sigma}^g_{jot}$ using the data on years 2005 and 2009 in the estimation. Numbers are reported for males and females, separately. The normalization component $w_o = \bar{\alpha}^g_{jot} - \alpha^g_{jot}$ (cf. equation (4.2) and the preceding discussion) is provided for comparison.

distribution, whereas type 9 females seem very similar to type 4 males in that they are high-achievers in everything else except in agriculture and health care jobs. In particular, we see substantial skill variation across different occupations. Allocation of talent according to ones' comparative advantages, therefore, seems potentially important.

We also observe substantial variation in skill dispersion. Type 1 (male) and type 10 (female) individuals have in many respects similar skills, but additionally, their skill variation is order of magnitude larger than for other types of individuals. Within the other types, high-achievers in both genders seem to have relatively high skill variation, possibly due to relatively fat tail of the wage/skill distribution.

Finally, we may build similar estimation table also for the sorting parameters ν^{g}_{jot}. They are, however, only identified up to differences, which makes proper comparisons somewhat more difficult. In Section 5, I discuss these parameters through the lens of my model, giving them a behavioral interpretation. That enables a more interesting analysis.

4.4 Externally calibrated parameters

I summarize the model calibration in Table 4. The labor supply model is estimated using maximum likelihood. For the remaining parameters, I assume $J = 5$ worker types and classify the occupations to $O = 8$ clusters. The number of occupational clusters is limited by worker mobility. Increasing the number of clusters reduces between-cluster mobility leading to an identification failure. The number of worker types is similarly limited by within-type mobility, and computational costs.

Table 4: Calibration overview

Parameter	Name	Source/target
	Structural estimation	
Skill distribution	α^{brg}_{jot}	Likelihood estimation
Preference distribution	ν^{brg}_{jot}	Likelihood estimation
Stand. dev. of idiosyncratic wage draws	σ^{brg}_{jot}	Likelihood estimation
Share of workers of type j	m^{brg}_{jt}	Likelihood estimation
	External calibration	
Number of types	J	5
Number of occupations	O	8
Production share of occupation o	γ_{ot}	Share in wage bill
TFP series	z_t	Sum of wages

4.4.1 Productivity shocks

For estimating the aggregate shock process, I use sum of wages as a measure of GDP. This comes directly from the firm FOC with CRS production

$$Y_t = \sum_{o=1}^{O} W_{ot} l_{ot}.$$

I also avoid estimating the full shock process (3.1), but instead sum over o in (3.2) to calibrate the realized shocks z_t as

$$z_t = \log\left(\text{Sum of wages}_t\right) - \sum_{o=1}^{O} \gamma_{ot} \log(l_{ot}),$$

where production shares γ_{ot} are calibrated each period to match the wage share as

$$\gamma_{ot} = \frac{W_{ot} l_{ot}}{Y_t}.$$

The equilibrium labor demand l_{ot}, in turn, matches the supply

$$l_{ot} = L_{ot} = \sum_{j,b,r,g} m_{jt}^{brg} \mathbb{P}_{jt}^{brg}(o) \exp\left(\alpha_{jot}^{brg} - \frac{(\sigma_{jot}^{brg})^2}{2}\right),$$

and is obtained after estimating and solving the model. This explicitly relies on Lemma 3.1 stating that the model solution is independent of the aggregate shocks z_t.

Note that the productivity series z_t is matched exactly and provides no restrictions on (or a test of) the model. Moreover, due to the normalization (4.2) the z_t series itself is also meaningless, since the units of measurement change each period: the z_t series includes the relevant price index associated with the particular measurement used, and the ratio z_{t+1}/z_t includes an unidentified change of measure. Therefore, the series is useless for understanding aggregate shocks, which are irrelevant to my analysis, though, as long as Y_t is correctly matched. Nevertheless, we are able to separate changes in labor demand (i.e. γ_{ot}) from changes in labor supply.

4.5 Model fit

In Table 5 (a) I compare actual and modeled annual earnings for males and females as well as the gender earnings gap. The model is able to explain the observed data relatively well, albeit the male earnings are slightly underestimated and female earnings slightly overestimated. The earnings gap is to a small extent correspondingly underestimated. In panel (b) of the table, I consider the occupational choices of males and females, separately. The model explains the data relatively well.

Table 5: Model fit: annual earnings (euros), and occupational distribution (%)

	Male		Female		Gender gap	
Year	Model	Data	Model	Data	Model	Data
1995	25,146	25,241	19,451	19,353	5,888	6,267
2000	30,720	31,495	23,523	22,725	7,197	8,771
2005	35,953	36,706	27,921	27,156	8,031	9,550

(a) This table documents the true and modeled annual earnings for males and females as well as the annual gender earnings gap.

	Male		Female	
Occupation	Model	Data	Model	Data
Agriculture and forestry	9.5	5.8	3.6	3.0
Low-skill service	4.6	5.1	13.0	12.8
Social skilled	17.2	17.3	22.8	23.4
Tradespeople	29.8	30.8	3.2	3.3
Manual work	22.2	24.9	7.5	7.1
High-skill service	2.9	2.8	4.2	4.3
Clerical	10.6	10.0	22.6	23.0
Health and social care	3.1	3.3	23.1	23.2
Total	100.0	100.0	100.0	100.0

(b) This table documents the true and modeled occupational distribution for males and females for $t = 2005$.

5 Aggregate effects of labor market sorting

With the estimated model at hand, we may now run counterfactuals to consider the role of talent allocation for the aggregate economy. The sorting effects in the model are driven by the non-pecuniary benefit parameters ν_{jo}^{brg}. Without specifying a channel, and considering a corresponding quasi-experiment, it is generally impossible to determine to what extent these parameters are driven by *misallocation* (discrimination, entry barriers etc.) and true *inherent heterogeneity* in societal norms, environmental factors, agents' preferences etc. In this Section, I will consider a number of counterfactuals without specifying the channel. In the next section, I consider a specific quasi-experiment, *The Finnish Comprehensive School Reform* and concentrate on a specific channel, the educational system, and its effect on talent allocation in the context of cross-sectional variation created by the reform.

In this section, I continue to align unobservable heterogeneity only by gender, and ignore the observable heterogeneity in birth cohorts and regions.

5.1 Aggregate productivity

A particularly simple counterfactual is one, where all the agents sort purely on their wages, which describe their productivity. This corresponds to setting $\nu_{jot}^{g} = 0$ for everyone, and describes the counterfactual economy with no barriers, discrimination or even preferential differences in occupational selection. The results of these counterfactuals are reported in Table 7.

The somewhat striking result is that the gain for aggregate productivity is almost negligible, in the order of 2 to 3 % depending on year. Changing talent allocation does not seem to have high potential for increasing output. In particular, there does not seem to be much room for substantial misallocation, even though the concept of "misallocation" is not properly defined in the context of my frictionless model; whatever the reasons for the current talent allocation, no substantial gains for aggregate productivity are to be achieved by improving it (independent of what "improving" means).[8]

Even though improving talent allocation does not seem to provide substantial gains for aggregate productivity with *existing skills*, it does not mean that it could not be an important part of economic growth. For instance, each new birth cohort enters the economy with new set of skills. They potentially sort to occupations differently from previous birth cohorts, and these differences may be important also for aggregate output.

[8] Workers' welfare depends not only on their productivity, but also their preferences. Increasing aggregate productivity is, therefore, not necessarily welfare-improving. In fact, taking my model literally, and interpreting the ν_{jot} as preferences, ignoring them in occupational sorting would almost necessarily lead to decreased welfare. More generally, I am however agnostic about the true interpretation of these sorting parameters, and they may also be considered to include different sorts of occupational entry barriers, manifested through (revealed) preferences.

Table 7: Counterfactual analyses of sorting patterns

	Sorting on productivity			Constant sorting with age		
Year	Earnings (data)	Earnings (counterfactual)	Δ%	Earnings (data)	Earnings (counterfactual)	Δ%
1995	22,339	23,070	3.3			
2000	27,174	27,729	2.0	28,199	28,250	0.2
2005	31,970	32,785	2.6	32,769	32,866	0.3

Notes: The table documents the true earnings per capita, which is the model-implied measure of output per capita, as well as earnings for two counterfactual exercises. The first one considers a counterfactual, where everyone sorts purely on skills. The second one considers one, where each individual sort in 2000 and 2005 similarly as they sorted in 1995 and 2000, respectively. The percentages report the relative increase to the true GDP. The true earnings data in the latter counterfactual analysis is calculated on the population of individuals at age 35-60, and the mean wages include an experience premium.

Here, comparing the unobservable types across age groups, however, requires parameterizing age in the model, in order to guarantee that labeling of types remains unaltered between different age groups. This produces challenges with the model identification, which requires substantial occupational mobility within each group of observable characteristics. I circumvent this challenge by comparing the estimated economy to one where everyone sorts on skills, and study the time trends in aggregate productivity under both worlds. As discussed, the difference between sorting purely on skills versus the model-estimated sorting is minor, and this is a robust result across the whole study period. Skill improvement alone seems to be able to explain essentially all of the growth in aggregate productivity, which suggest that evolving occupational sorting with new birth cohorts is not driving economic growth.

However, improving talent allocation may be an important part of a worker's life cycle. Workers accumulate experience, and this may potentially allow them to better sort according to their comparative advantages. To study this further, I consider a counterfactual, where workers occupational sorting is not updated with age. As discussed in section 4.1, I use two-period rolling windows to estimate the model parameters $\alpha^g_{jot}, \alpha^g_{jo,t+1}, \nu^g_{jo,t}$ and $\nu^g_{jo,t+1}$. Note that within each two-period estimation window, the type labeling remains the same. This allows for counterfactual of setting $\nu^g_{jo,t+1} := \nu^g_{jo,t}$, that is keeping the sorting preferences constant. The results are again reported in Table 7.

The effect of differential sorting seems negligible and, if anything, workers care less about wages, when they grow older. Most of the gains in aggregate productivity are due to improved skills. Note that in this analysis, we consider the same set of individuals at ages 30 – 55 and 35 – 60, in a general equilibrium model. This is not a true macro counterfactual, since here I am not including a new cohort of 30-34 year olds in the analysis, nor am I removing the retirees (who are still few at the age of 60). Instead, I am only looking at the effect of an aging population. That said, it seems unlikely that the new cohorts or the new retirees would have substantial general equilibrium externalities to the core of the working age population; at least not to the extent that the life cycle effect would be substantially misestimated.

5.2 Distributional effects of talent allocation

Next I turn to consider the distributional effects. I again compare the estimated model economy to a counterfactual economy, where everyone sorts on skills, i.e. $\nu^g_{jot} = 0$ for all j, o, t and g. I denote the occupational wages in the estimated model economy by W_{ot} and in the counterfactual economy by W^{skills}_{ot} and define the occupational *preference wedges* as

$$\tau_{ot} = \frac{W^{skills}_{ot}}{W_{ot}} - 1.$$

These wedges describe the effect of non-differential sorting patterns on equilibrium wages. Wedge of $\tau_{ot} = 1$ implies a 100 % increase in the equilibrium wage, if everyone sorts purely on skills. Occupations with particularly low τ_{ot} are, thus, considered generally dislikable, and are compensated with comparatively high wages, and vice versa for occupations with high values of τ_{ot}. I present these wedges in Figure 3 for year $t = 2005$.

Unsurprisingly, agriculture and forestry jobs have comparatively limited labor supply, which drives wages up. While sorting on productivity would provide only modest gains for aggregate productivity, it would decrease wages in these jobs by 57.1%. Similarly, high-skill service occupations would see a 49.4% decrease in wages, associated with corresponding increase in occupational labor supply. Notably, health and social care jobs are also considered "dislikable" with a -22.5% preference wedge. On the other hand, jobs requiring good social skills are found attractive, and the equilibrium wage is relatively lower, given the relatively higher labor supply in these jobs: sorting on skills would increase wages by 26.6% in 2005. This may be explained by primary school teachers' long summer vacations, and sales representatives' fringe benefits.[9]

Figure 4 documents time trends for certain occupational preference wedges. The wedge in Agriculture and forestry decreased from -47% in 1995 to -57% in 2005, possibly indicating that younger birth cohorts consider such jobs increasingly dislikable or at least do not see their potential to provide long-term income during the future decades. Interestingly, the wedge on health care jobs increased from -31.6% in 1995 to -22.5% in 2005 and the wedge in social skilled jobs from 15.8% in 1995 to 26.6% in 2005, respectively.

Other notable trends are the relative increase of wages in tradespeople occupations with decreasing preference wedges from 33.6% in 1995 to 19.7% in 2005. Altogether, these mean that economic growth has substantially different effects on wages across the occupational

[9]For instance, access to a company car may provide substantial non-taxable benefits.

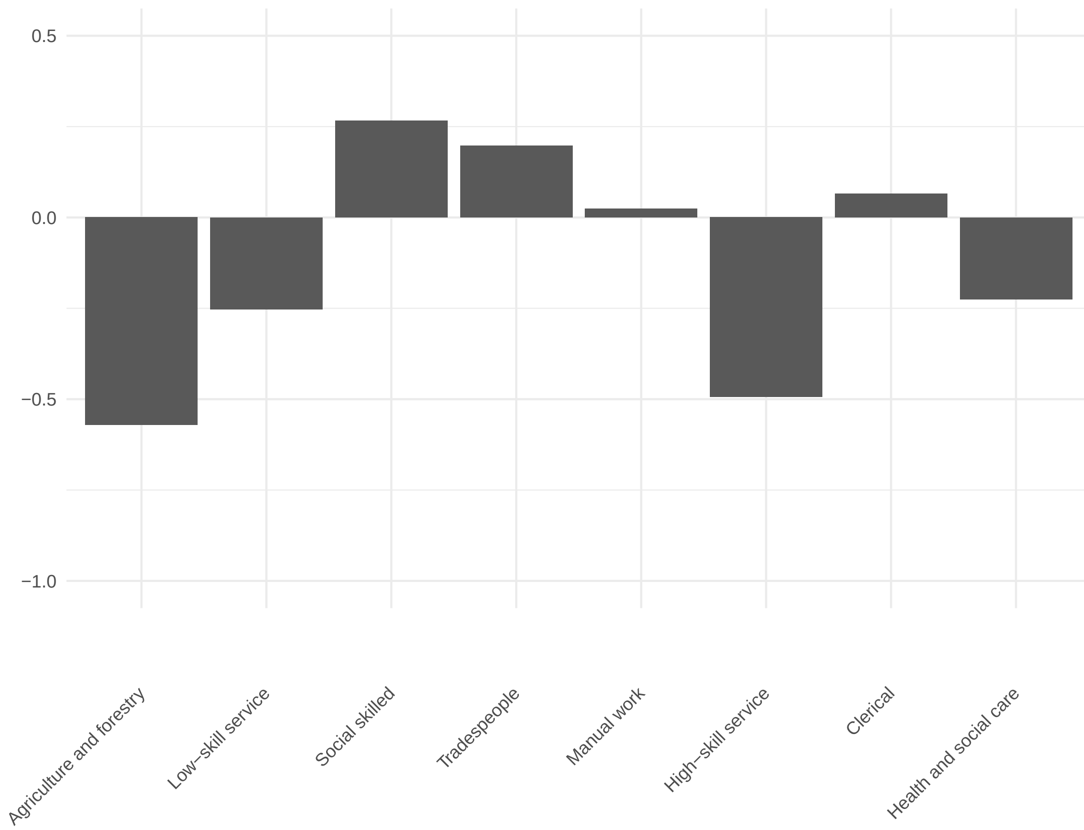


Figure 3: Occupational preference wedges

Notes: The figure reports the counterfactual effect on occupational wages, if workers sorted to occupations purely on skills in 2005. For instance, the equilibrium wage in Agriculture and forestry would decrease by 57.1%

distribution. In my study population of $30-55$ year old workers, the average wages increased 43.1% from 22,339 euros in 1995 to 31,970 euros in 2005, which contributes to 24.4% increase in real wages during these 10 years. However, in health and social care as well as social skilled occupations with increasing preference wedges, the increase in real wages was almost 10 percentage points lower.

These results suggest that occupational sorting patterns have substantial effects on the income distribution. Obviously, the varying time trends of occupational equilibrium wages differentially affect workers with different skills, working on the affected occupations. Much of these trends occur along unobservable lines differentially affecting the different unobservable worker types. We may, however, also consider the effects between observable worker groups; for instance, between females and males.

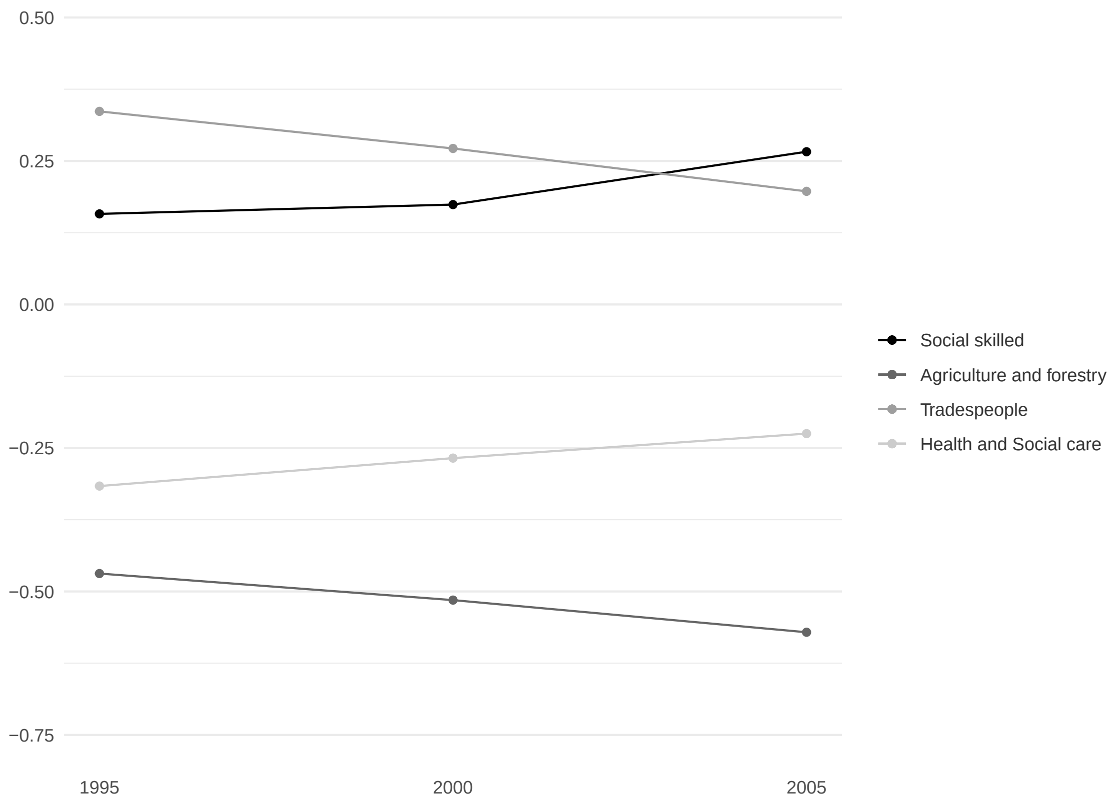


Figure 4: Occupational preference wedge trends

Notes: The figure reports the counterfactual effect on occupational wages, if workers sorted to occupations purely on skills in 1995, 2000, and 2005, for the four shown occupations.

5.3 Gender earnings gap

First, I note that if all workers sorted purely on skills in 2005, the gender earnings gap would increase by 4.8 percentage points, which corresponds to 21.7% in the modeled earnings gap. However, a potentially more interesting counterfactual is to consider removal of gender differences, by considering females acquiring the *current* skills and preferences of males, instead of comparing both to a hypothetical counterfactual, where everyone sorts on their skills. To consider such a counteractual, I now parameterize the skill parameters as

$$\alpha_{jot}^{brg} = \alpha_{jot} + \beta_{jt}\delta_{ot}D_o^g, \tag{5.1}$$

where D_o^g is an indicator, whether a worker is female and works in occupation o. I use normalization $\delta_{1t} = 1$. The value of α_{jot} describe the skill parameters for males. The wage regression (4.1) then takes the form

$$\log \omega_{it} = w_{o_t(i)t} + \alpha_{j(i)o_t(i)t} + \beta_{j(i)t}\delta_{o_t(i)t}D_{o_t(i)}^{g(i)} + \varepsilon_{it}.$$

This parameterization now allows for comparing males and females, by linking the type labelings between genders. For instance, in Table 2 the male type 1 and female type 10 seem to be generally similar in that they both are mostly low-achievers with very high standard deviations in their wage draws. Similarly male type 4 and female type 9 have similar skill patterns across occupations. The above parameterization allows for a more rigorous matching of these types between genders.

Table 8: Counterfactual analyses of gender differences

			Counterfactuals					
			Females sorting as males		Females with male skills		Gender equalization	
	Model	Data	Earnings	Δ%	Earnings	Δ%	Earnings	Δ%
Male earnings	36,017	36,706	35,539	-1.3	36,824	2.2	35,248	-2.1
Female earnings	27,856	27,156	27,454	-1.4	35,526	27.6	35,248	26.5
% of male earnings	77.3	77.7	77.3		96.5		100	
Aggregate earnings	31,970	31,970	31,529	-1.4	36,180	13.2	35,248	10.3

Notes: The table documents the modeled and true per capita earnings for females, males and in aggregate for 2005, as well as three different counterfactuals: (i) females sorting to occupations exactly as males, (ii) females having exactly same skills as males, (iii) both of (i) and (ii), i.e. removing all gender differences, whereby females adopt the skills and preferences of males. The model is calibrated to individuals with age 35-60 in 2005, with classified occupation and positive income both in 2005 and 2009.

Females may have gender specific implementation δ_{ot} of occupational skills, and different female types may have different sensitivities β_{jt} for these occupational gender effects, but the inherent (possibly biological) potential of type 9 females are similar to those of type 4 males in Table 2. These type labelings are now matched by considering the above skill parameterization. Note that by Theorem 4.3 the correct labeling of sorting parameters ν_{jot} also follows from this parameterization.

As a first check it is instructive to look at how well this parameterization matches the data. Obviously, the parameterized model does not exactly match the values obtained from the non-parametric model described in Table 5, but the differences seem minor. As expected, the differential sorting patterns between females and males, given existing skills, do not explain large differences in earnings or aggregate productivity. However, if the skill distribution of females converged to that of males, we would see a large reduction in gender earnings gap, but the gap would still not completely disappear, but females would still had average earnings 3.5% below that of males. This is due to differential sorting patterns of females and males.

Interestingly, however, these differences improve total output, and if females behaved exactly as males, both female and male earnings would decrease relative to this counterfactual, even if gender wage gap obviously disappeared (cf. Table 8). The differential sorting preferences of females relative to males have positive general equilibrium externalities, which increase aggregate productivity: with the existing skills these externalities benefit both females and males similarly, but if the female skill distribution were to converge to that of males', these externalities would mostly benefit males. This suggest that complete gender equality in measured wages might not be necessarily Pareto optimal.[10]

Finally, based on this analysis gender skill equalization would lead to 13.2% increase in aggregate productivity. However, this would be reduced to 10.3% if females also adopted male preferences in terms of occupational sorting. In other words, approximately 22% of the productivity effect can be contributed to the differential sorting patterns of females relative to those of males. The next section will consider this observation further in the context of a specific quasi-experiment, *The Finnish Comprehensive School Reform*, and asks whether occupational sorting could be important for aggregate effects in the context of a policy reform.

6 The Finnish Comprehensive School Reform

The Finnish comprehensive school reform was implemented during 1972 – 1977 with the specific goal to improve educational equity and access to higher education for underrepresented groups. The reform postponed educational tracking by 5 years: in the pre-reform system, students had to choose at the age of 10 – 11, whether they entered a vocational track or continued their academic studies in a general secondary school, leading to high school and potentially to college. After the reform, everyone continued within the same *comprehensive school* until the age of 15 – 16 before making this choice (cf. Figure 5). In both pre- and post-reform systems the students had (mostly compulsory) formal education until the age of 15-16.

The reform implementation was staggered, as Finland was divided into six regions (cf. Figure 6), which implemented the reform in different years during 1972 – 1977. The first affected birth cohorts were born in 1961 – 1966 depending on the region, where they lived during the implementation years. Leveraging this staggered reform implementation in a difference-in-differences set-up, Pekkarinen (2008) shows that the reform had a 4.1 percentage point negative causal effect on the gender wage gap in 2000. In this section, I extend his analysis by checking pre-trends, considering the reform's effect on occupational selection, and by introducing the estimated partial equilibrium effect into my general equilibrium model, which allows for analysis of the reform's equilibrium effects. I consider data covering

[10] A complete welfare discussion would require identifying whether the occupational preference wedges represent true differences in preferences, or possibly discrimination or other sorts of occupational entry barriers/frictions. My analysis cannot make this distinction, but based on this analysis, it should, however, be noted that the differential occupational sorting of males and females may have positive earnings effects for both, even if gender earnings gap persists.

the whole Finnish population from years 2011 – 2017.

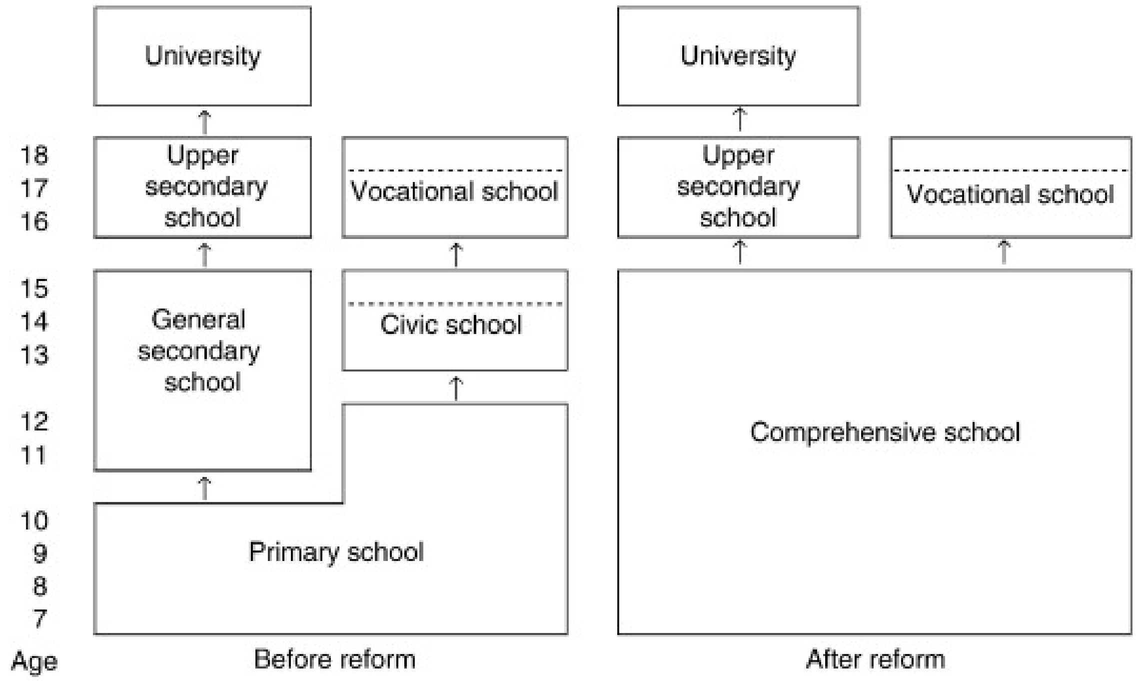


Figure 5: Pre and post-reform systems. Source Pekkarinen (2008).

6.1 Reduced form evidence: occupational selection and gender earnings gap

In Figure 1, I documented the occupational selection patterns into health care jobs by gender for different birth cohorts and depending on, whether students were affected by the school reform or not. We observed that the in the affected regions, unlike men, women were more likely to select into health care jobs than in the unaffected regions. I will now further motivate this analysis by using the staggered implementation of the reform to provide causal evidence to this effect. I begin by discussing the occupational clustering used for this analysis covering years 2011 – 2017.

6.1.1 Occupational selection

The occupational classification changed in 2010 in the Statistics Finland data, and to consider occupational selection after 2010 , I reclassify the occupations into nine new clusters. The new classification allows for more nuanced analysis of the occupations than the clustering using the occupational information prior 2010 in Section 4.2.1.

For producing the occupational classification, I first consider the share of college educated workers in each 5-digit occupation, and divide these into three educational terciles. Then within each tercile, the occupations are further divided into three clusters using k-

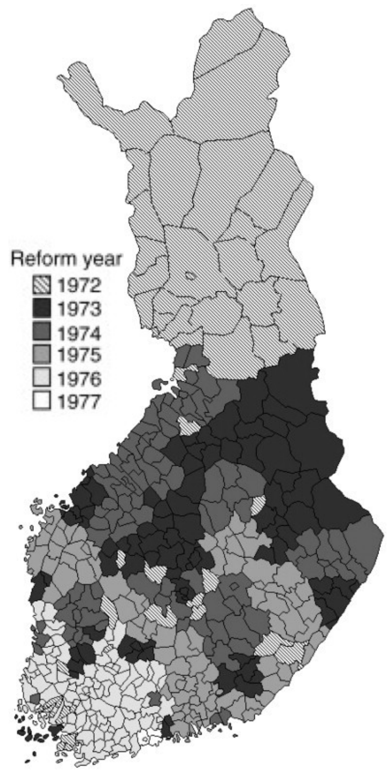


Figure 6: Finnish comprehensive school reform implementation by year and municipality. Source Pekkarinen (2008).

Table 9: Occupational grouping

#	Educational tercile	Broad category	Sample Occupations
1	Low share of college educated workers	Low-skill service	Cleaner, Cook, Waitress, Bartender, Firefighter, Policeman, Hairdresser, Cosmetologist
2		Tradespeople	Carpenter, Electrician, Plumber, Auto mechanic, Machinist
3		Routine	Salespeople, Taxi and bus drivers, Mailman, Nanny, Nurse assistant
4	Medium share of college educated workers	Manual labor	Farmer, Lumberjack, Gardener, Physical therapist, Caregiver, Guards, Police lieutenant
5		Clerical	Sales supervisor, Secretary, Bookkeeper, Office clerk, Personal banker
6		Technicians	Mechanical/electrical/construction technicians, Telecom specialist, Construction supervisor
7	Large share of college educated workers	Medical	Registered nurse, Medical doctor, Dentist, Pharmacist, Social worker
8		Teachers	Primary, secondary, vocational, polytechnic, kindergarten and special education teachers
9		Engineers	Mechanical/electrical/production engineer, Software developer, Architect

means as discussed in section 4.2.1. Table 9 documents a sample of common occupations in each cluster, and category labelings, which I have produced based on this information included in the table.

I study a difference-in-differences set-up with a triple difference specification comparing males and females, and consider the regression coefficients δ_l from the triple-difference event study regressions

$$\begin{aligned} Y_{it} = & Region_i + Birth\,Cohort_i + Gender_i + \\ & + Region_i \times Birth\,Cohort_i + Region_i \times Gender_i \\ & + Birth\,Cohort_i \times Gender_i + \sum_{l=-5}^{4} \delta_l D_{il} + \varepsilon_{it}, \end{aligned} \tag{6.1}$$

where $Region_i, Birth\,Cohort_i$ and $Gender_i$ denote region, birth cohort and gender fixed effects, and where $D_{il} = 1$ if individual i is female and belongs to a birth cohort, which is $-l$ years away from being treated: in a given region, $l = -1$ includes the individuals in the last non-treated birth cohort, whereas $l = 0$ corresponds to the first treated birth cohort etc. I normalize $\delta_{-1} = 0$, and note that we may test parallel pre-trends by testing $\delta_l = 0$ for $l < -1$. For technical details on identifying the treatment effects I refer to Appendix C.[11]

Table 10 reports the triple difference coefficients for the school reform's differential effect on women occupational choices relative to men. We observe the reform had a substantial and statistically significant 0.52 percentage point positive increase in women's relative probability of choosing health care occupation compared to men. Note that, since the effects in Table 10 are measured in percentage points, the sum of the measured effects must be zero: increase in the probability of choosing one occupation, must be compensated by a corresponding decrease in the probability of choosing another occupation.

Moreover, Pekkarinen (2008) documents that the school reform had a 4.1 percentage point causal reduction in the gender wage gap in 2000. I replicate his analysis, and find 1.7 percentage point reduced-form reduction in the gender earnings gap in 2011 – 2017. In Figure 7 I depict the event-study plot for these analysis on earnings gap (left) and the reform's differential effect on choosing health care jobs (right) between men and women.

6.2 Worker types and estimation

I now turn to consider the structural model. I estimate the reduced form (partial equilibrium) effect of the reform and input that into the general equilibrium framework. The staggered implementation of the reform allows for estimating the reform's effect on skills using difference-in-differences, and conditional on skill, we may further estimate the effect on the sorting parameters ν_{jot}. Since all the agents (treated and non-treated) work in the same labor market and face the same prices and wages, the reduced form effects on skills and occupational sorting are not biased by general equilibrium forces.

The regression specification differs from that of the previous subsection in that now I introduce type-occupation controls by considering the log-earnings regression

$$\log \omega_{iot}^{brg} = \delta_{1o}^{b} + \delta_{2o}^{r} + \delta_{3o}^{g} + \delta_{4o}^{br} + \delta_{5o}^{bg} + \delta_{6o}^{rg} + \beta_o D_{io} + \alpha_{jo} + \varepsilon_{it}, \tag{6.2}$$

[11] While this project is first to consider the macroeconomic effects of the Finnish comprehensive school reform, existing work on the reform has suggested a decrease in gender wage gap (Pekkarinen, 2008) and reduction in intergenerational income elasticity (Pekkarinen, Uusitalo, and Kerr, 2009).

Table 10: Triple difference effect on occupational selection (percentage points)

Occupation category	Estimate	Standard error
Low-skill service	0.07	0.32
Tradespeople	-0.24	0.041
Social Routine	0.12	0.51
Manual labor	-0.07	0.26
Clerical	-0.34	0.27
Technicians	-0.29*	0.14
Medical	0.52***	0.11
Teachers	-0.05	0.36
Engineers	0.27	0.17

Notes: The table documents the estimates and standard errors for the triple-difference regression parameter studying the reform's differential impact on women's relative to men's occupational choices.

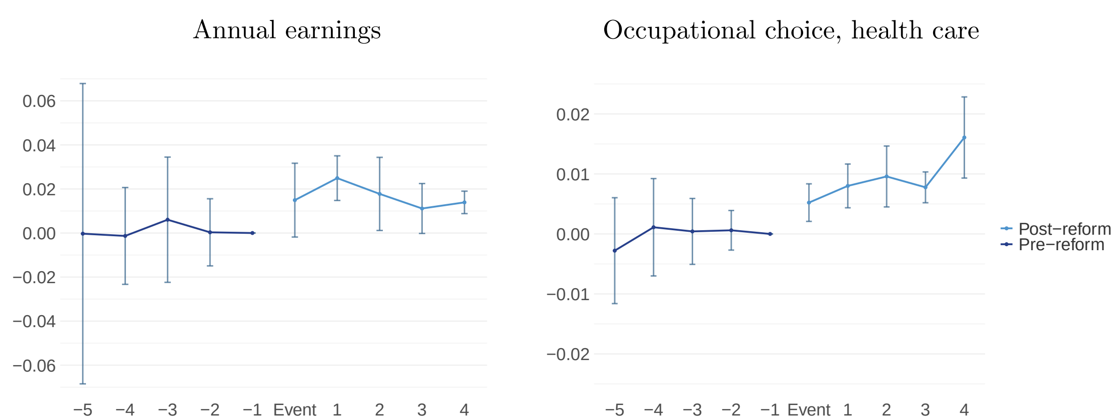


Figure 7: Event-study plot on the reform's differential impact on females, earnings (left) and occupational selection into health care jobs (right).

Notes: The figure plots the coefficients δ_l – describing the reform's effect on the negative of gender gap – of the triple-difference regression (6.1) using annual ($\sinh^{-1}$) income (left) and occupational choice to health care occupations (right) as dependent variable. The dark blue graph on the left hand side correspond to pre-treatment birth cohorts, whereas the light blue on the right hand side describe post-treatment effects. Under parallel pre-trends the pre-treatment effect should be zero. The error bars depict 95% confidence intervals.

where D_{io} is an indicator for whether the individual is female, and was treated by the reform, and β_o is the causal parameter of interest. Under a suitable parallel trends assumption, this identifies the differential average treatment effect on treated between women and men. For theoretical details of the analysis, I refer to Appendix C. Note that this specification is a special case of that discussed in Sections 3 and 4, as we have merely introduced the parameterization

$$\alpha_{jot}^{brg} = \delta_{1o}^{b} + \delta_{2o}^{r} + \delta_{3o}^{g} + \delta_{4o}^{br} + \delta_{5o}^{bg} + \delta_{6o}^{rg} + \beta_o \bar{D}_o^{brg} + \alpha_{jo}, \tag{6.3}$$

where the treatment status $D_{io} = \bar{D}_o^{b(i)r(i)g(i)}$ is fully determined by the worker's observable characteristics $b(i), r(i)$ and $g(i)$.

The first five columns of Table 11 now report my estimation for the occupation-skill matrix $\{\exp(\widehat{\alpha}_{jo})\}$. Again, we observe substantial heterogeneity in worker skills. Type 1 and type 4 workers constitute to high-achievers in high-education and low-education jobs, respectively. Type 5 workers are low-achievers across the whole occupational distribution, and type 3 workers have comparative advantage in low-education jobs, but with lower ability than type 4 workers. Type 2 workers have a strong comparative advantage in engineering, while in other occupations they are low-achievers, although with very high within-occupation wage dispersion.

Indeed, in Table 12 I document the standard deviations σ_{jo} for the idiosyncratic wage draws ε_i. The within-occupation wage dispersion is an order of magnitude higher for type 2 workers than for the other types.

The sixth column of Table 11 shows the reform's effect on gender wage gap (β_o), when ignoring type-specific heterogeneity ($J = 1$), whereas the seventh column shows the effect with five heterogeneous types ($J = 5$). The differences in the point estimates between these two cases suggest shifts in occupational sorting between the pre- and post-reform workers.

Estimating the reform's effect on the sorting parameters ν_{jot} is somewhat more complicated, since they are not produced by a simple regression, but instead estimated by the MLE. By Theorem 4.3, the type labelings of ν_{jot} match with those of α_{jot}. Having parameterized the skill parameters with (6.3), the type labelings for ν_{jot}^{brg} thus necessarily match across different sets of observable characteristics b, r and g. This means that we may directly estimate the reform effect on the sorting parameters ν_{jot}^{brg} non-parametrically, under a suitable parallel trends assumption.

I employ the potential outcome framework of Rubin (1974), and define the potential outcomes $\nu_{it}(r) := \nu_{j(i)o_t(i)t}^{b(i)r(i)g(i)}(r), r = 1, \dots, R$ and $\nu_{it}(0) := \nu_{j(i)o_t(i)t}^{b(i)r(i)g(i)}(0)$ for a worker i having been treated or not, respectively, depending on their region of schooling r. Note that the actual treatment status is fully determined by the triplet $b(i), r(i), g(i)$. The true value of the sorting parameter for a worker i is then given by

$$\nu_{it} = \nu_{j(i)o_t(i)t}^{b(i)r(i)g(i)} = \nu_{it}(0) + \sum_{r=1}^{R} [\nu_{it}(r) - \nu_{it}(0)] \cdot D_i^r,$$

Table 11: Estimated type-occupation productivities and reform triple-difference effect

		Worker type ($J = 5$)					Reform wage effect (%)	
#	Occupation category	1	2	3	4	5	$J = 1$	$J = 5$
1	Low-skill service	0.83	0.42	1.06	1.53	0.64	0.93	1.38
2	Tradespeople	0.81	0.41	1.07	1.50	0.70	-1.21	2.47
3	Routine	0.83	0.35	1.05	1.55	0.67	-0.09	0.56
4	Manual labor	0.83	0.30	1.07	1.55	0.63	4.70	3.11
5	Clerical	0.80	0.55	1.03	1.57	0.66	-2.23	0.31
6	Technicians	1.85	0.54	0.75	1.09	0.54	1.67	2.79
7	Medical	1.88	0.66	0.73	1.04	0.56	2.66	0.54
8	Teachers	1.83	0.68	0.74	1.12	0.50	-1.98	2.37
9	Engineers	1.79	1.03	0.72	1.07	0.56	-0.63	-1.55
Type proportions (m_j)		0.21	0.06	0.38	0.19	0.16		

Notes: The first five columns of the table reports the occupational skills parameters $\widehat{\alpha}_{jo}$ for each of the five types. Due to the implicit normalization this corresponds to male skills in the 1960 birth cohort in the northern most part of Finland. The sixth and seventh columns report the reform's triple-difference effect on occupational gender wage gaps assuming homogeneity ($J = 1$) and with five heterogeneous types ($J = 5$) following the specification (6.2), respectively. The differences in these regression results represent the impact of selection on unobservables.

where D_i^r is an indicator whether the worker was treated, and went to school in region r.

This approach now allows for directly producing the non-parametric average treatment effect on treated (ATT) for these sorting parameters. In the triple-difference set-up, my object of interest is the differential effect on females defined by

$$\begin{aligned} ATT^j_{\nu,diff}(r,t) =& \mathbb{E}[\nu_t(r) - \nu_t(0) \mid Type = j, Gender = Female, D^r = 1] \\ &- \mathbb{E}[\nu_t(r) - \nu_t(0) \mid Type = j, Gender = Male, D^r = 1], \end{aligned}$$

I identify this treatment effect under a standard parallel trends assumption, and estimate it non-parametrically using the estimated values $\widehat{\nu}^{brg}_{jot}$. In my counterfactual analysis, I then replace the realized values of ν_{it} with the estimated potential outcomes $\nu_{it}(0)$ produced by this analysis. The counterfactual analysis, and decomposing the effect of the reform on its skill and sorting components then relies on separately shutting down the skill effect encompassed in the parameters β_o in the wage regression, and the non-parametrically estimated ATT on ν_{jot}. For details, I refer to Appendix C.

Table 12: Estimated type-occupation wage dispersion (σ_{jo}) and occupational wages

		Worker type ($J = 5$)					
#	Occupation category	1	2	3	4	5	$\log W_o$
1	Low-skill service	0.11	1.09	0.14	0.15	0.27	10.52
2	Tradespeople	0.14	1.27	0.16	0.19	0.35	10.57
3	Routine	0.12	1.26	0.15	0.19	0.34	10.56
4	Manual labor	0.13	1.46	0.16	0.21	0.36	10.57
5	Clerical	0.13	1.60	0.15	0.22	0.40	10.62
6	Technicians	0.28	1.31	0.15	0.19	0.27	11.00
7	Medical	0.28	0.99	0.15	0.21	0.26	10.96
8	Teachers	0.31	1.37	0.17	0.22	0.35	10.96
9	Engineers	0.30	1.42	0.15	0.21	0.37	11.04

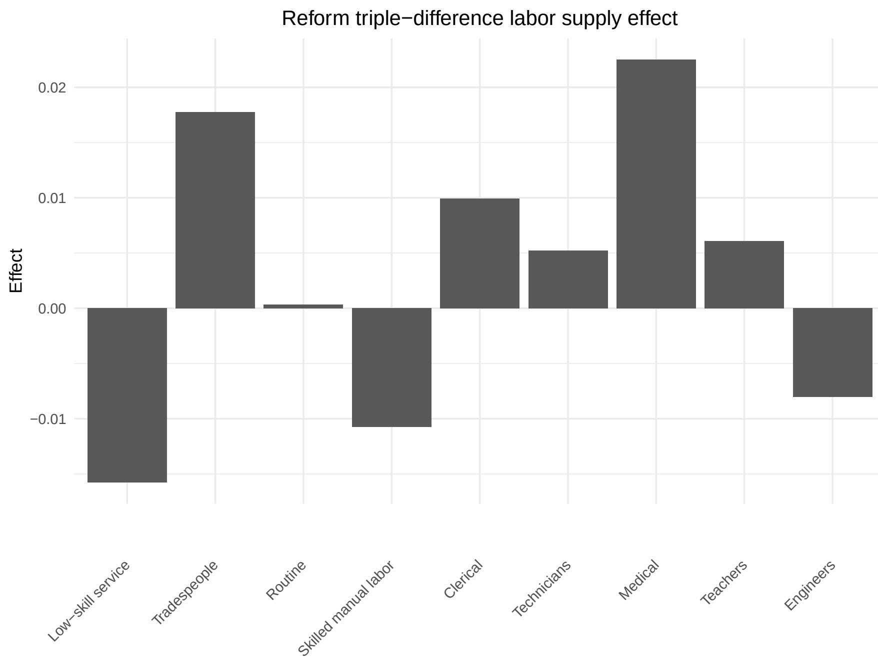
Notes: The table reports the estimate standard deviations $\widehat{\sigma}_{jo}$ for the productivity shocks in the wage regression (6.2). In this analysis, the standard deviations are assumed to be constant across observables.

6.3 Aggregate productivity and labor supply

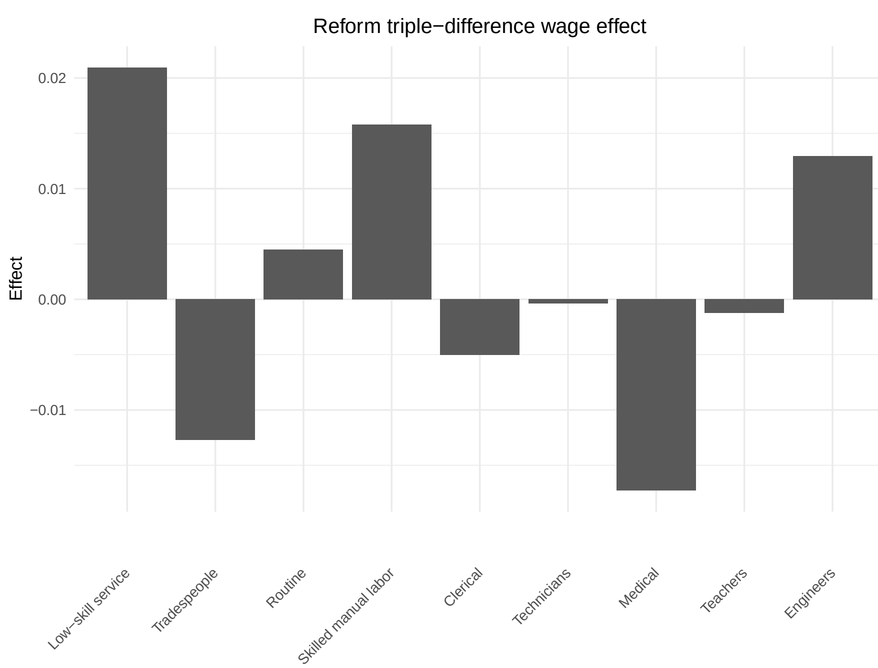
The main counterfactual I study is a shock to the sorting parameters ν_{jot}. As documented in Section 6.1, in the cross section of workers, occupational sorting has systematic variation based on the school system the workers studied in. I use this variation to quantify a the shock on the sorting parameters ν_{jot}, as described above. I shut down the effect of the reform on these parameters and resolve the equilibrium. This allows for comparing the aggregate productivity across the two economies with pre- and post-reform sorting parameters, respectively.

In panels (a) and (b) of Figure 8, respectively, I show the general equilibrium effects on occupational labor supply and wages, due to the school reform's differential impact on female and male occupational sorting. The wage effects have a positive bias relative to the labor supply effects due to a total 1.00 % increase in aggregate productivity (no-reform vs. all females treated). Given the relatively small changes in occupational labor supply, the productivity effect is quantitatively substantial. I further decompose the effect to its skill and sorting components, by separately shutting down the skill and sorting effects. The

Figure 8: Occupational selection, medical and clerical jobs by gender



(a) General equilibrium effect on occupational labor supply due to the school reform's differential impact on female vs. male occupational sorting.



(b) General equilibrium effect on occupational wages due to the school reform's differential impact on female vs. male occupational sorting.

skill effect explains 52% of the increase in aggregate productivity, whereas the sorting effect explains 48%. This suggests that talent allocation may have quantitatively substantial aggregate effects in the context of policy reforms, both for income inequality and aggregate productivity.

7 Conclusion

This paper studies the question of talent allocation by using the exogenous variation in skills and sorting produced by the Finnish comprehensive school reform to discipline the quantitative analysis. I build a structural model of the labor market, estimate it using Finnish administrative data, and also consider number of counterfactuals. In the context of policy reforms affecting the skill distribution, talent reallocation seems an important additional force for potentially improving aggregate productivity. That said, the reallocation seems fast enough that ex post we observe little potential for raising aggregate output by improving talent allocation. Nevertheless, continuing talent reallocation still seems important for trends in occupational equilibrium wages. Occupational wages have consistent time trends explained by evolving non-wage components of occupational sorting.

References

Girum Abebe, A Stefano Caria, and Esteban Ortiz-Ospina. The selection of talent: Experimental and structural evidence from ethiopia. *American Economic Review*, 111(6): 1757–1806, 2021.

John M Abowd, Francis Kramarz, and David N Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999.

Daron Acemoglu and David Autor. Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier, 2011.

Jesper Bagger and Rasmus Lentz. An empirical model of wage dispersion with sorting. *The Review of Economic Studies*, 86(1):153–190, 2019.

Alex Bell, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen. Who becomes an inventor in america? the importance of exposure to innovation. *The Quarterly Journal of Economics*, 134(2):647–713, 2019.

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. Discretizing unobserved heterogeneity. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2019-16), 2017.

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. A distributional framework for matched employer employee data. *Econometrica*, 87(3):699–739, 2019.

Kenneth Burdett and Dale T Mortensen. Wage differentials, employer size, and unemployment. *International Economic Review*, pages 257–273, 1998.

David Card, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1):S13–S70, 2018.

Gaaitzen J De Vries. Productivity in a distorted market: The case of b razil's retail sector. *Review of Income and Wealth*, 60(3):499–524, 2014.

Peter A Diamond. Aggregate demand management in search equilibrium. *Journal of political Economy*, 90(5):881–894, 1982a.

Peter A Diamond. Wage determination and efficiency in search equilibrium. *The Review of Economic Studies*, 49(2):217–227, 1982b.

Daniel A Dias, Carlos Robalo Marques, and Christine Richmond. Misallocation and productivity in the lead up to the eurozone crisis. *Journal of Macroeconomics*, 49:46–70, 2016.

Claudia Goldin. A grand gender convergence: Its last chapter. *American Economic Review*, 104(4):1091–1119, 2014.

John Grigsby. Skill heterogeneity and aggregate labor market dynamics. Technical report, Working Paper, University of Chicago 2019., 2019.

Marcus Hagedorn, Tzuo Hann Law, and Iourii Manovskii. Identifying equilibrium models of labor market sorting. *Econometrica*, 85(1):29–65, 2017.

James J Heckman and Guilherme Sedlacek. Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market. *Journal of political Economy*, 93(6):1077–1125, 1985.

Chang-Tai Hsieh and Peter J Klenow. Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448, 2009.

Chang-Tai Hsieh, Erik Hurst, Charles I Jones, and Peter J Klenow. The allocation of talent and us economic growth. *Econometrica*, 87(5):1439–1474, 2019.

Sebnem Kalemli-Ozcan and Bent E Sørensen. *5. Misallocation, Property Rights, and Access to Finance: Evidence from within and across Africa.* University of Chicago Press, 2016.

Gaurav Khanna. Large-scale education reform in general equilibrium: Regression discontinuity evidence from india. *unpublished paper, University of Michigan*, 2015.

Thibaut Lamadon, Magne Mogstad, and Bradley Setzler. Imperfect competition, compensating differentials and rent sharing in the us labor market. Technical report, National Bureau of Economic Research, 2019.

Kurt Lavetti and Ian M Schmutte. Estimating compensating wage differentials with endogenous job mobility. 2016.

Ilse Lindenlaub. Sorting multidimensional types: Theory and application. *The Review of Economic Studies*, 84(2):718–789, 2017.

Jeremy Lise and Fabien Postel-Vinay. Multidimensional skills, sorting, and human capital accumulation. *American Economic Review*, 110(8):2328–76, 2020.

Jeremy Lise and Jean-Marc Robin. The macrodynamics of sorting between workers and firms. *American Economic Review*, 107(4):1104–35, 2017.

Dale T Mortensen. Markets with search friction and the dmp model. *American Economic Review*, 101(4):1073–91, 2011.

Giuseppe Moscarini and Fabien Postel-Vinay. The job ladder: Inflation vs. reallocation. *Draft, Yale University*, 2017.

Giuseppe Moscarini and Francis G Vella. Occupational mobility and the business cycle. Technical report, National Bureau of Economic Research, 2008.

Sari Pekkala Kerr, Tuomas Pekkarinen, and Roope Uusitalo. School tracking and development of cognitive skills. *Journal of Labor Economics*, 31(3):577–602, 2013.

Tuomas Pekkarinen. Gender differences in educational attainment: Evidence on the role of tracking from a finnish quasi-experiment. *Scandinavian Journal of Economics*, 110(4): 807–825, 2008.

Tuomas Pekkarinen, Roope Uusitalo, and Sari Kerr. School tracking and intergenerational income mobility: Evidence from the finnish comprehensive school reform. *Journal of Public Economics*, 93(7-8):965–973, 2009.

Christopher A Pissarides. Job matchings with state employment agencies and random search. *The Economic Journal*, 89(356):818–833, 1979.

Solomon W Polachek, Konstantinos Tatsiramos, and Klaus F Zimmermann. *Gender Convergence in the Labor Market*. Emerald Group Publishing, 2015.

Fabien Postel-Vinay and Jean-Marc Robin. Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica*, 70(6):2295–2350, 2002.

Diego Restuccia and Richard Rogerson. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*, 11(4):707–720, 2008.

Diego Restuccia and Richard Rogerson. The causes and costs of misallocation. *Journal of Economic Perspectives*, 31(3):151–74, 2017.

Andrew Donald Roy. Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2):135–146, 1951.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Robert Shimer and Lones Smith. Assortative matching and search. *Econometrica*, 68(2): 343–369, 2000.

Isaac Sorkin. Ranking firms using revealed preference. *The quarterly journal of economics*, 133(3):1331–1393, 2018.

Christopher Taber and Rune Vejlin. Estimation of a roy/search/compensating differential model of the labor market. *Econometrica*, 88(3):1031–1069, 2020.

A Model Appendix

In this section I present the Model solution. The discussion here culminates to the proof of Lemma 3.1.

Proof of Lemma 3.1. I guess and verify that the value functions for a worker i with characteristics b, r and g take the forms

$$V_{it}^{brg}(z) = a_{j(i)t}^{brg} z + b_{it}^{brg}. \tag{A.1}$$

and

$$v_{it}^{brg}(o \mid z) = c_{j(i)t}^{brg} z + d_{j(i)ot}^{brg}. \tag{A.2}$$

I will also guess and verify that the choice probabilities $\mathbb{P}_{jt}^{brg}(o \mid z)$ are independent of z. Note first that

$$w_{ot} = \log(W_{ot}) = z_t + \log\left[\frac{\gamma_o F}{l_{ot}}\right]$$

and

$$w_{ot} - w_{o't} = \log\left[\frac{\gamma_o}{l_{ot}}\right] - \log\left[\frac{\gamma_{o'}}{l_{o't}}\right].$$

If the choice probabilities are independent of z, then by (3.3) labor supply l_{ot} is independent of z for every o and t. This is consistent with the guess of the choice probabilities being independent of z. Moreover, the choice probabilities are now given by

$$\begin{aligned}\mathbb{P}^{brg}_{jt}(o_i = o \mid z) &= \frac{\exp[v^{brg}_{jt}(o \mid z)]}{\sum_{o'=1}^{O} \exp[v^{brg}_{jt}(o' \mid z)]} \\ &= \frac{\exp\left(d^{brg}_{jot}\right)}{\sum_{o'=1}^{O} \exp\left(d^{brg}_{jo't}\right)}.\end{aligned}$$

It follows that

$$\begin{aligned}\mathbb{E}_{\zeta',z'}[V^{brg}_{i,t+1}(z') \mid z] &= \mathbb{E}\left[\sum_{o'} \mathbb{P}^{brg}_{j,t+1}(o') v^{brg}_{i,t+1}(o' \mid z') \mid z\right] \\ &= \sum_{o'} \mathbb{P}^{brg}_{j,t+1}(o')[c^{brg}_{j(i),t+1}\rho z + d^{brg}_{j(i)o',t+1}] \\ &= c^{brg}_{j(i),t+1}\rho z + \sum_{o'} \mathbb{P}^{brg}_{j,t+1}(o') d^{brg}_{j(i)o',t+1}\end{aligned}$$

which gives

$$v^{brg}_{it}(o \mid z) = z + \log\left[\frac{\gamma_o F}{l_{ot}}\right] + \alpha^{brg}_{j(i)ot} + \nu^{brg}_{j(i)ot} + \beta\big[c^{brg}_{j(i),t+1}\rho z + \sum_{o'} \mathbb{P}^{brg}_{j,t+1}(o') d^{brg}_{j(i)o',t+1}\big],$$

We obtain

$$c^{brg}_{jt} = 1 + \beta\rho c^{brg}_{j,t+1}, \quad c^{brg}_{j,b+T} = 0$$

and

$$\begin{aligned}d^{brg}_{jot} &= \log\left[\frac{\gamma_o F}{l_{ot}}\right] + \alpha^{brg}_{jot} + \nu^{brg}_{jot} + \beta \sum_{o'} \mathbb{P}^{brg}_{j,t+1}(o') d^{brg}_{j(i)o',t+1}, \\ d^{brg}_{jo,b+T} &= 0, \quad \text{for all } o,\end{aligned}$$

which is a well-posed system of difference equations, given

$$\frac{F}{l_{ot}} = \frac{\prod_{o=1}^{O} l_{ot}^{\gamma_o}}{l_{ot}} \quad \text{and} \quad l_{ot} = \sum_{j,b,r,g} m^{brg}_{jt} \mathbb{P}^{brg}_{jt}(o).$$

Assuming the solution for this system, $V^{brg}_{it}(z)$ is characterized by

$$a^{brg}_{jt} = c^{brg}_{jt} \quad \text{and} \quad b^{brg}_{it} = d^{brg}_{j(i)o^*_t(i)t} + \zeta_{io^*_t(i)t},$$

where

$$o^*_t(i) = \arg\max_{o}\, d^{brg}_{j(i)ot} + \zeta_{iot}.$$

These choices for $a^{brg}_{jt}, b^{brg}_{it}, c^{brg}_{jt}$ and d^{brg}_{jt} solve the problem with the value functions described in (A.1) and (A.2). This finishes the proof of Lemma 3.1 as well as characterizes the model solution.

□

(a) α_{jo1} and α_{jo2} (b) ν_{jo1} and ν_{jo2} (c) σ_{jo1} and σ_{jo2}

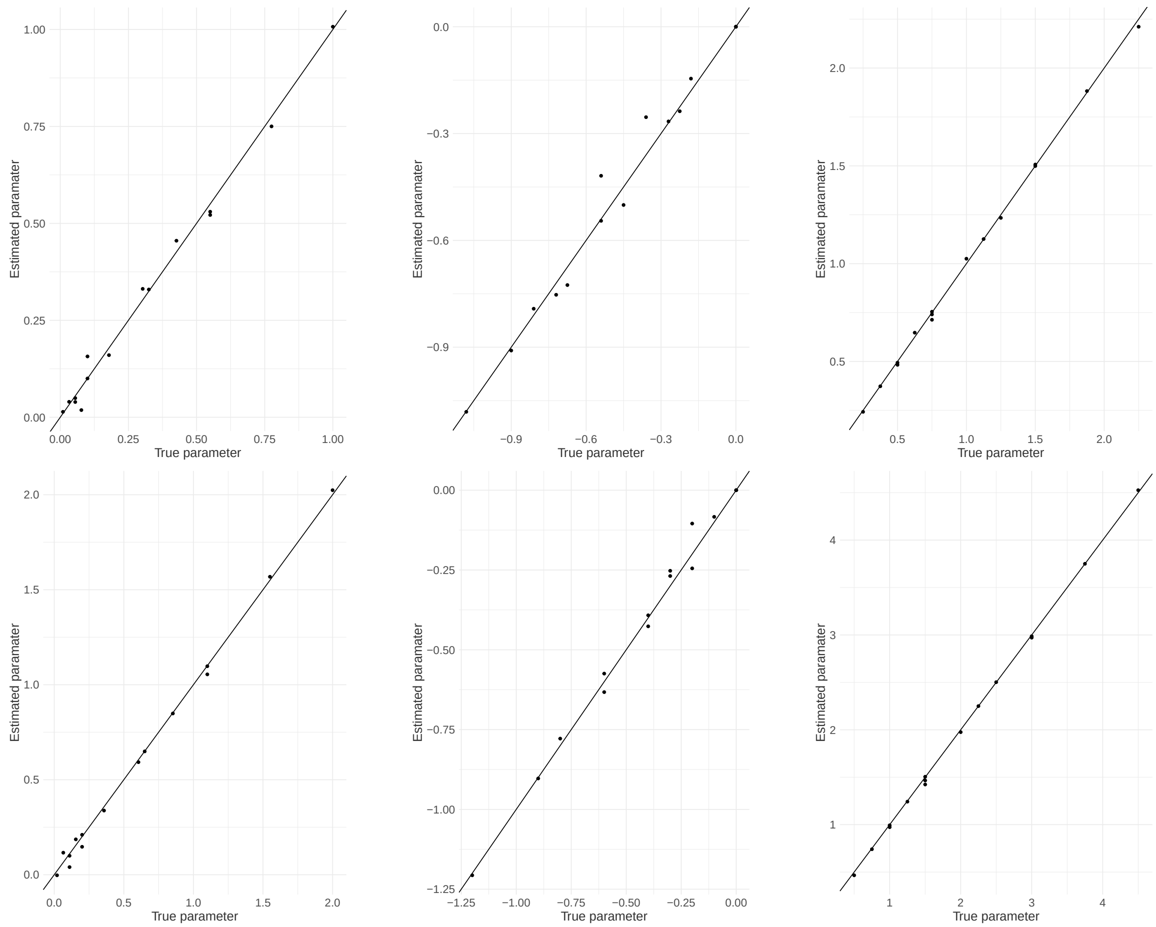


Figure 10: True vs. estimated parameter values in a simulated data

Notes: The figures depict the simulated values (x-axis) of α_{jo1}, ν_{jo1} and σ_{jo1} (top row), and α_{jo2}, ν_{jo2} and σ_{jo2} (bottom row) against the estimated values $\widehat{\alpha}_{jo1}, \widehat{\alpha}_{jo2}, \widehat{\sigma}_{jo1}, \widehat{\sigma}_{jo2}, \widehat{\nu}_{jo1}, \widehat{\nu}_{jo2}$ (y-axis) using the Expectation Maximization algorithm, with $J = 3$ types and $O = 5$ occupations.

B Identification argument and validation

The proof of Theorem 4.3 follows from Bonhomme et al. (2019, Theorem 1). To provide further confidence on the point identification of the model skill and preference parameters, I study the estimation of the model parameters from simulated data. The idea in this exercise relies on the fact consistency of an estimator implies point identification.

While short of suggesting consistency, I provide evidence towards this direction, by showing that the expectation maximization algorithm used to estimate the model parameters correctly finds the true parameters in a simulated data set. For consistency, one would need to do a similar exercise for a larger selection of parameter values in a Monte Carlo simulation, and show that the mean square error reduces with the sample size N at rate

N^{-1}.

In Figure 10 I show the parameters $\alpha_{jo1}, \alpha_{jo2}, \sigma_{jo1}, \sigma_{jo2}, \nu_{jo1}$ and ν_{jo2} estimated from a simulated data, against the true parameters used for the model simulation. The synthetic data is simulated using the model as the data generating process.

C The Finnish comprehensive school reform: heterogeneous treatment effects

C.1 Treatment effect

I begin by defining the treatment effect of interest for my analysis. Let Y_{it} denote the outcome (wage, occupational sorting) of a worker i at time t. I consider the treatment of having studied under the new comprehensive school, and denote by D_i^r the treatment indicator of being treated at group/region $r = 1, \ldots, R$, where the regions are enumerate from the first treated birth cohort to the last. The potential outcomes are denoted by $Y_{it}(r), r = 1, \ldots, R$ and $Y_{it}(0)$ for treated and non-treated individuals, respectively. Then

$$Y_{it} = Y_{it}(0) + \sum_{r=1}^{R}[Y_{it}(r) - Y_{it}(0)] \cdot D_i^r,$$

In my structural model, I assume the effect to be identical within a type. Correspondingly, I define the average treatment effect on the treated within a type j as

$$ATT^j(r,t) = \mathbb{E}[Y_t(r) - Y_t(0) \mid Type = j, D^r = 1].$$

I concentrate on gender differences in a triple-difference set-up. My main object of interest is

$$\begin{aligned} ATT^j_{Y,diff}(r,t) =& \mathbb{E}[Y_t(r) - Y_t(0) \mid Type = j, Gender = Female, D^r = 1] \\ &- \mathbb{E}[Y_t(r) - Y_t(0) \mid Type = j, Gender = Male, D^r = 1], \end{aligned}$$

where the subscript $Y \in \{\text{skill, sorting}\}$ on the left hand side specifies the type of outcome variable considered.

By the identification argument in Section 4, identification and consistent estimation of $\mathbb{E}[Y_i \mid Type, Gender, D]$ is possible. However, fully controlling for the unobservable types for the average treatment effect ATT_{diff} is not possible, since we cannot generally guarantee the same labeling of types across different groups of observable characteristics. For this reason, I assume the following parametric form (6.3) for the skill parameters.

Without unobservable heterogeneity (i.e. $J = 1$) the above parameterization represents standard triple-difference set-up with occupation specific effects. The triple-difference estimation does not require assumptions about the data generating process, but only a suitable parallel trends assumption. With unobservable heterogeneity, the standard triple-difference argument is not possible, since we need to make sure that type labelings across observable groups match. The above parameterization provides this common type labeling at the cost of assuming that type-occupation interactions take the multiplicative form. The parallel trends assumption used to identify the reform effect is as follows.

Assumption C.1 (Parallel trends)**.** Let G be an indicator, whether one has been treated by the reform. The parallel trends assumption states that

$$\begin{aligned}
&\mathbb{E}[Y_t^b(0) - Y_t^{b-1}(0) \mid Type, Occupation, Gender = Female, D^r = 1] \\
&- \mathbb{E}[Y_t^b(0) - Y_t^{b-1}(0) \mid Type, Occupation, Gender = Female, G = 0, D^r = 0] \\
&= \mathbb{E}[Y_t^b(0) - Y_t^{b-1}(0) \mid Type, Occupation, Gender = Male, D^r = 1] \\
&\quad - \mathbb{E}[Y_t^b(0) - Y_t^{b-1}(0) \mid Type, Occupation, Gender = Male, G = 0, D^r = 0].
\end{aligned}$$

We have the following theorem for identifying the treatment effects.

Theorem C.2. *Suppose the Assumption C.1 and the assumptions of Theorem 4.3 hold as well as that the data generating process for skills satisfies* (6.3)*. Then*

$$\begin{aligned}
ATT^{j,c}_{skills,diff}(r,t) &= [\alpha^{br,1}_{jot} - \alpha^{b-1,r,1}_{jot}] - [\alpha^{b,c,1}_{jot} - \alpha^{b-1,c,1}_{jot}] \\
&\quad - [\alpha^{br,2}_{jot} - \alpha^{b-1,r,2}_{jot}] - [\alpha^{b,c,2}_{jot} - \alpha^{b-1,c,2}_{jot}], \\
ATT^{j,c}_{sorting,diff}(r,t) &= [\nu^{br,1}_{jot} - \nu^{b-1,r,1}_{jot}] - [\nu^{b,c,1}_{jot} - \nu^{b-1,c,1}_{jot}] \\
&\quad - [\nu^{br,2}_{jot} - \nu^{b-1,r,2}_{jot}] - [\nu^{b,c,2}_{jot} - \nu^{b-1,c,2}_{jot}],
\end{aligned}$$

where $c = r+1, \ldots, R$ is one of the control regions.

Proof. To be added. □

As discussed, the parameterization (6.3) ensures that types are comparable across different sets of observable characteristics, when estimating the skill effects. By Theorem 4.3 the labeling of types for the sorting parameters ν^{brg}_{jot} agree with those of the skill parameters α^{brg}_{jot} and, hence, the type-specific differences across different regions and birth cohorts is meaningful also for these sorting parameters.

In my equilibrium analysis, I assume constant effects across regions obtained by averaging the above estimates as

$$\widehat{ATT}^{j}_{Y,diff}(r,t) = \frac{1}{(R-1)(R-r)} \sum_{r=1}^{R-1} \sum_{c=r+1}^{D} \widehat{ATT}^{j,c}_{Y,diff}(r,t),$$

where $\widehat{ATT}^{j,c}_{Y,diff}(r,t)$ is the sample analogue of the formulas in Theorem C.2.