ISSN 1833-4474

College Majors and Skill Mismatch in Labour Markets

Michelle Rendall, Satoshi Tanaka and Yi Zhang

School of Economics
The University of Queensland,
St Lucia, QLD 4072, Australia

Version: July, 2024

Keywords: Skill Mismatch, College Major, Occupation, Roy Model, Administrative Tax Record

JEL Classification: E24, J24, J31

College Majors and Skill Mismatch in Labour Markets*

Michelle Rendall† Satoshi Tanaka‡ Yi Zhang§

July 31, 2024

Abstract

This paper studies the extent of skill mismatch across college major-occupation combinations. In the framework of this paper, skill mismatch is measured through individual wage losses in partial equilibrium and through output losses in general equilibrium. The model relies on the estimation of major-occupation returns based on the Roy model. Using Australian administrative tax panel data containing employment history and university degree information, we find sizeable wage losses, up to 28 percent, but smaller general equilibrium output losses, up to 10 percent – at the upper bound – from workers being allocated to occupations not well linked to their majors. Our results show that STEM, Commerce, and Social Science and Arts majors are the main drivers of aggregate mismatch, but a worker's occupational mismatch declines over the life cycle, and, most importantly, in general equilibrium, disappears almost entirely by age 35.

Keywords: Skill Mismatch, College Major, Occupation, Roy Model, Administrative Tax Record

JEL Classification: E24, J24, J31.

*We thank Bruce Chapman, Fatih Guvenen, Chris Edmond, Michael Keane, Todd Scholleman, and the participants at Society for the Advancement of Economic Theory Conference 2022, 2nd Australian Workshop on Public Finance, VAMS Workshop, Econometric Society Australian Meeting 2023, Labour Econometrics Workshop, Human Capital Conference at LAEF, OZMac Conference 2023, CIGS End of Year Conference 2023, GRID Conference 2024, University of Melbourne, Peking University, Hitotsubashi University, Australian National University, Keio University, University of Tokyo, Waseda University, Osaka University, and University of Queensland. Also a special thank you to Andrew Carter, Son Nguyen, and Justin Holland at the Australian Taxation Office for their support in working with the data. We gratefully acknowledge financial support from the Australian Research Council (ARC Grant DP200100105). The usual disclaimers apply.

†Department of Economics, Monash University; CEPR; Email: michelle.rendall@monash.edu

‡School of Economics, University of Queensland; Email: s.tanaka@uq.edu.au

§School of Economics, University of Queensland; Email: yi.zhang1@uq.edu.au

1 Introduction

One of the most important roles of a tertiary education system is to supply individuals with skills, which are demanded in the economy. In almost all countries, colleges and universities offer various fields of study that students can freely choose between. Perspective students tend to pick courses based on preferences and expectations of future labor demand. However, changes in the macroeconomic environment and technological progress across industries make it hard to predict what kind of skills will be in shortage and, thus, demanded in the economy upon graduation and beyond. Economists and policymakers are, therefore, concerned with the concept of talent misallocation or skill mismatch, a situation where workers with certain types of skills are working in occupations that do not necessarily require their skill set. This mismatch can lead to inefficiencies in individuals coupling with the labor market, and, consequently, exacerbating income inequality and generating aggregate production inefficiencies or welfare losses.

The goal of this paper is to develop an economic framework to consistently measure mismatch between the skills learned and the jobs held by workers not only at the individual level but in general equilibrium or aggregate. Using the Roy model (Roy, 1951), we first estimate returns to college major by occupation. To account for possible selection into occupations, we control for the selection probability of choosing a certain occupation at any point in time, allowing workers to select into occupations based on their unobserved pecuniary and non-pecuniary benefits. Using estimated returns, we calculate penalties for not working in the occupation that gives the highest payoff to a given major for all major-occupation combinations. These estimated penalties allow us to construct our mismatch measure at the individual level, which measures individual wage losses for not working in one's top occupation—the occupation with the largest return to a given college major. This measure captures individual mismatch given fixed returns to skill.

We next extend the framework to the general equilibrium by introducing a CES production function that nests workers with various college majors as inputs to produce occupational output. Using the estimated college returns from the Roy model we use time and regional variation to estimate the CES production function parameters. The estimated production function parameters are then used to solve the optimal problem of how workers with different college majors should be allocated to occupations. Similar to the literature on factor misallocations (Hsieh and Klenow, 2009), we obtain the optimal allocations from the condition that workers with the same college major have the same marginal product of labor across occupations. A general equilibrium mismatch measure is then constructed as the production loss of the current allocation of workers relative to the optimal allocation.

We illustrate the framework through a study of the aggregate and individual evolution of mismatch in Australia. Data providing information on both the skills of individuals and detailed work histories is rare. The Australian ALife data (Carter et al., 2021), which consists of a panel of a 10 percent population sample of tax-paying individuals in Australia and education outcomes for terminal degrees, provides us with a unique dataset to study the evolution of skill match in the economy. We analyze the evolution of mismatch in Australia both at the aggregate level and over workers' life cycles.

Our results show that there are large penalties of major-occupation mismatch: penalties range from 3.6 percent of wage loss for Food, Hospitality and Personal Services majors to 23.7 percent of wage loss for Creative Arts majors. Addressing selection bias increases the overall mismatch, in particular for STEM, Management and Commerce, and Society and Culture majors. Correcting the occupational selection bias enlarges the penalty by the range from 3.7 to 23.1 percentage points for different STEM majors and 28.8 percentage points for Management and Commerce majors. Overall, at the individual level, average wage penalties are 28.3 percent.

This figure, however, falls considerably when prices adjust due to a reallocation of factors. That is, based on our benchmark specification of the general equilibrium, removing all mismatches within majors by reallocating talent across occupations suggests an average output loss of 10.0 percent during the sample period of 2003-2019. This loss represents the upper bound of the cost of mismatch when all mismatch is due to frictions rather than to individual tastes. The loss is primarily driven by Management and Commerce (3.5 percent), STEM-related majors (2.5 percent), and Society, Culture, and Arts majors (2.4 percent). Our results also indicate that mismatch falls over the life cycle and, on average, is greater for males than females due to selection into different majors. Most skill mismatch disappears as agents age when returns to skills are endogenous.

The remainder of the paper is organized as follows. Section 2 reviews the existing literature. Section 3 outlines the framework used to study mismatch. Section 4 describes the specific ALife dataset. Section 5 discusses the estimation. Section 6 discusses the results. Finally, Section 7 concludes.

2 Literature Review

The study of skill mismatch has reemerged in recent years with the availability of large-scale data and the regularly updated database of occupational characteristics, O*NET. One strand of the literature proposes a measure of skill mismatch using the information on workers' skills and occupational requirements.[1] Two main contributions in this literature are Guvenen, Kuruscu, Tanaka, and Wiczer (2020) and Lise and Postel-Vinay (2020), who estimate models with multidimensional skills, sorting, and human capital accumulation, using the National Longitudinal Survey of Youth 1979 and O*NET.[2] This approach of combining worker and occupation information is conceptually simple, but requires detailed data on both sides and relies heavily on text information for mapping skills and occupations.[3] In this paper, we propose a new framework to measure skill mismatch. As in the misallocation literature that measures factor misallocations through TFP dispersions (Hsieh and Klenow, 2009), we measure the degree of occupation-major mismatch from occupational wage differentials for workers with the same college majors.[4] In this spirit, we also extend our measure to a general equilibrium framework.

This paper also relates to an active literature on the skill content of occupations, which uses data on occupation characteristics (e.g., from O*NET) and explores how they are related to the wages in those occupations (Autor, Levy, and Murnane, 2003; Ingram and Neumann, 2006; Poletaev and Robinson, 2008; Gathmann and Schönberg, 2010; Bacolod and Blum, 2010; Rendall, 2010; Yamaguchi, 2012; Autor and Handel, 2013; Deming, 2017; Deming and Noray, 2020; Atalay, Phongthiengtham, Sotelo, and Tannenbaum, 2020; Edmond and Mongey, 2021). This connects to research on estimating the Roy (1951) model, where controlling selectivity is one of the key issues.[5] Dahl (2002) and Lee (1983) develop a control function approach that deals with selection in high-dimensional settings. More recently, Eckardt (2019) extends this framework to estimate returns to training by occupation and analyzes the mismatch between training and occupation in the German apprenticeship context. This paper builds on these studies by estimating returns to college majors by occupation, constructing a mismatch measure between college majors and occupations, and studying the general equilibrium implications of skill mismatch.

Finally, our paper also relates to the literature on estimating returns to college majors. Altonji, Blom, and Meghir (2012) and Altonji, Arcidiacono, and Maurel (2016) survey the recent contributions in this literature, including Arcidiacono (2004), Altonji, Kahn, and Speer (2014),

[1] Rathelot et al. (2017) provide an excellent review of this literature, including the earlier works by Pellizzari and Fichen (2013) and McGowan and Andrews (2015) that contributed to the development of the OECD's measure of skill mismatch.

[2] Fredriksson, Hensvik, and Skans (2018) take a different approach and defines mismatch as the gap between the skills of an entrant and his incumbent peers in the same occupation and establishment. Groes, Kircher, and Manovskii (2015) study how the probability of occupational switching depends on mismatch, defined as the gap between a worker's wage from the average of his coworkers.

[3] In a similar approach, Şahin et al. (2014) and Patterson et al. (2016) quantify mismatch of unemployment and resulting productivity loss using information on the unemployed's previous jobs and information on vacancies.

[4] Another closely related paper is Hsieh, Hurst, Jones, and Klenow (2019) who estimate worker misallocations across race and gender from their labor market outcomes.

[5] See French, Taber, et al. (2011) for a comprehensive review of the estimation of the Roy model.

Altonji, Kahn, and Speer (2016) and Kirkeboen, Leuven, and Mogstad (2016). The main focus of those papers is to estimate the effects and determinants of college major choice, controlling for selection. The focus of our paper, on the other hand, is to estimate the returns to college majors by occupation and to construct a measure of the mismatch between majors and occupations. Lemieux (2014) and Kinsler and Pavan (2015) are a few exceptions in this literature, which estimate returns to college majors separately for those with and without a career related to their major. Liu, Salvanes, and Sørensen (2016) is another exception, which estimates returns to college majors by industry for Norway and studies major-industry mismatch, similar in spirit to our paper. Our paper differs from Liu, Salvanes, and Sørensen (2016) in three distinct ways. First, their measure of mismatch is an index based on the ranking of wage payments across industries for a given major, whereas our measure is directly tied to wage losses or production losses. Second, their mismatch measure does not consider the selection problem, and, therefore, is potentially biased, while our mismatch measure addresses this issue under the framework of the Roy model. Third, we provide a general equilibrium framework for evaluating the mismatch of major-occupation combinations, which allows us to consider the general equilibrium effects of worker reallocation.

3 Skill Mismatch Framework

The economy consists of K intermediate firms, which are each defined by a unique occupation. Intermediate firms operate a CES production function and hire workers with different skills (from different college majors). Occupations produce a differentiated good that is aggregated with a CES production function to a final good. There is no capital in the economy. Workers choose the occupation that maximizes their utility based on the Roy model. Workers differ in their acquired skills - college majors. We take these skills as given and fixed. Lastly, we consider two cases, (i) perfect and (ii) imperfect mobility of labor across occupations.

3.1 Model Components

Individual Wages Suppose that, after graduating from college, a worker enters the labor market and provides college-major-specific human capital $\tilde{h}^j_{i,t} = h^j a_i$, where h^j is a college-major-specific skill for which we assume $h^j = 1$ if the worker studied major j at college, and a_i is general labor efficiency units.[6] Let $\mathcal{J} = (1, \ldots, j, \ldots, J)$ represent the set of college majors. In any period t the individual chooses one occupation from the set $\mathcal{K} = (1, \ldots, k, \ldots, K)$. Income for

[6]We also consider a robustness exercise where the efficiency unit, a_i depends on college-major a^j_i.

an individual i with college major j, who works in occupation k is,

$$w_{i,j,k} = p_j^k \tilde{h}_i = p_j^k h^j a_i \tag{1}$$

where p_j^k is the price for major-j specific human capital in occupation k which equals the marginal product of labor. Taking the logarithm on both sides of equation (1), we have an empirical specification to estimate returns to college majors:

$$\ln w_{i,j,k,t} = \sum_{j'}^{J} \sum_{k'}^{K} \underbrace{\beta_{j'}^{k'}}_{\ln p_{j'}^{k'}} D_{j',k'} + \ln(a_{i,t}) + \epsilon_{i,j,k,t} \tag{2}$$

where $w_{i,j,k,t}$ is annual earnings, $\beta_{j'}^{k'}$ is the rate of return to major-j-specific human capital in occupation k (or logarithm of the price ($p_{j'}^{k'}$) of major j' in occupation k'), $D_{j',k'}$ is a dummy variable that takes 1 if $k' = k$ and $j' = j$, and $\epsilon_{i,j,k,t}$ is the error term.

Firms To determine the marginal product of labor, assume the economy's production function is a nested CES of major-occupation pairs.[7] We start with the intermediate production functions—the occupation-specific production functions. For each occupation k, we assume a CES production function that takes workers' skills, which are summarized by his/her major j, as inputs,

$$\begin{aligned} Y_t^k &= F^k(H_{1,t}^k, ..., H_{j,t}^k, ..., H_{J,t}^k) \\ &= A_{k,t}\left(\alpha_{1,t}^k \left(H_{1,t}^k\right)^{\sigma_k} + ... + \alpha_{j,t}^k \left(H_{j,t}^k\right)^{\sigma_k} + ... + \alpha_{J,t}^k \left(H_{J,t}^k\right)^{\sigma_k}\right)^{\frac{1}{\sigma_k}}, \end{aligned} \tag{3}$$

where $H_{j,t}^k = \sum_{i \in I^{j,k}} \tilde{h}_{i,t}$ is the supply of labor efficiency units of individuals with college major j to occupation k in year t, with $I^{j,k}$ the set of indexes for the individuals with major j in occupation k. Note that majors summarize different bundles of skills, where $\left(\alpha_1^k, \alpha_2^k, ..., \alpha_J^k\right)$ are the weights for all majors and σ^k is the substitution parameter across majors, all of which are occupation-specific.

The competitive final good producer then aggregates occupational outputs Y^k with a CES

[7] We allow for two versions: (1) without uneducated workers and (2) including uneducated workers. First, we focus only on the production function of college graduates. In this case, we could assume a final goods sector aggregates university and non-university labor. This will not affect the results of the analysis as long as university and non-university labor is separable and aggregated in the final CES goods sector only. Second, we also include uneducated workers as an additional input in the CES occupational production function. That is, we assume within occupations educated and uneducated workers have the same elasticity of substitution, but different productivity weights. Thus, uneducated workers can simply be considered another "major" choice. We discuss some of the issues with this approach in the results section.

production function,

$$
\begin{aligned}
Y_t &= F\left(Y_t^1, \ldots, Y_t^k, \ldots, Y_t^K\right) \\
&= A_t \left(\alpha_{1,t}\left(Y_t^1\right)^{\sigma} + \ldots + \alpha_{k,t}\left(Y_t^k\right)^{\sigma} + \ldots + \alpha_{K,t}\left(Y_t^K\right)^{\sigma}\right)^{\frac{1}{\sigma}},
\end{aligned} \tag{4}
$$

where $\left(\alpha_1, \alpha_2, ..., \alpha_J\right)$ are the weights for all occupations and σ is the substitution parameter across occupations.

The marginal product of labor in occupation k for college major j is,

$$
p_j^k = \frac{\partial Y_t}{\partial Y_t^k}\frac{\partial Y_t^k}{\partial H_{j,t}^k} = \alpha_{k,t} A_t Y_t^{1-\sigma}\left(\sum_{j=1}^{J} \alpha_{j,t}^k \left(H_{j,t}^k\right)^{\sigma_k}\right)^{\frac{\sigma}{\sigma_k}-1} \alpha_{j,t}^k \left(H_{j,t}^k\right)^{\sigma_k - 1}, \tag{5}
$$

Perfect Mobility and No Individual Tastes We first discuss an economy with perfect mobility across occupations and no tastes of individual workers for occupations. In this case, workers move across occupations to maximize their wages, which equalizes prices for major-specific human capital. Therefore, conditional on college majors, workers' earnings only differ in their efficiency units of labor which are common across occupations.

Individual i, who graduates in major j, chooses occupation k in calendar year t to maximize utility, $U_{i,j,t}$, or equivalently maximizes income,

$$
U_{i,j,t} = \max\left\{w_{i,j,1,t}, \ldots, w_{i,j,k,t}, \ldots, w_{i,j,K,t}\right\} \tag{6}
$$

Thus, the labor supply to an occupation is infinitely elastic. Workers will relocate to the occupation with the highest marginal product until marginal products conditional on college major equalize across occupations,

$$
\frac{\frac{\partial Y_t}{\partial Y_t^{k'}}\frac{\partial Y_t^{k'}}{\partial H_{j,t}^{k'}}}{\frac{\partial Y_t}{\partial Y_t^{k}}\frac{\partial Y_t^{k}}{\partial H_{j,t}^{k}}} = \frac{\alpha_{k',t}\left(\sum_{j=1}^{J} \alpha_{j,t}^{k'}\left(H_{j,t}^{k'}\right)^{\sigma_{k'}}\right)^{\frac{\sigma}{\sigma_{k'}}-1} \alpha_{j,t}^{k'}\left(H_{j,t}^{k'}\right)^{\sigma_{k'}-1}}{\alpha_{k,t}\left(\sum_{j=1}^{J} \alpha_{j,t}^{k}\left(H_{j,t}^{k}\right)^{\sigma_{k}}\right)^{\frac{\sigma}{\sigma_{k}}-1} \alpha_{j,t}^{k}\left(H_{j,t}^{k}\right)^{\sigma_{k}-1}} = 1. \tag{7}
$$

Imperfect Labor Mobility Relaxing the assumption on perfect mobility we now assume that workers differ in their tastes for different occupations. The utility, $U_{i,(k|j),t}$ of choosing occupation k given major j is now determined by earnings $w_{i,j,k,t}$ and non-pecuniary utility $\eta_{i,(k|j),t}$ received from unobserved occupational characteristics (i.e., working conditions, taste for an occupation, etc.),

$$U_{i,(k|j),c,r,t} = \underbrace{w_{i,j,k,t}}_{\text{Earnings}} + \underbrace{\eta_{i,(k|j),t}}_{\text{Non-pecuniary Utility}} \qquad \text{for each } j \in \{1,...,J\}. \tag{8}$$

A larger $\eta_{i,(k|j),t}$ means that a worker has a particular taste for occupation k. Alternatively, $\eta_{i,(k|j),t}$ could also represent labor market frictions faced by the individual or alternatively a combination of the two – tastes and frictions. Irrespective of the origin of η, workers pick the occupation where their utility $U_{i,j,t}$ is maximized,

$$U_{i,j,t} = \max\left\{U_{i,(1|j),t}, \ldots, U_{i,(k|j),t}, \ldots, U_{i,(K|j),t}\right\}. \tag{9}$$

The key difference now is that the marginal products of labor, conditional on a college major, do not necessarily equalize any longer, as only the *marginal* worker will be indifferent between different occupation choices,

$$\frac{\frac{\partial Y_t}{\partial Y_t^{k'}} \frac{\partial Y_t^{k'}}{\partial H_{j,t}^{k'}}}{\frac{\partial Y_t}{\partial Y_t^{k}} \frac{\partial Y_t^{k}}{\partial H_{j,t}^{k}}} = \frac{\alpha_{k',t}\left(\sum_{j=1}^{J} \alpha_{j,t}^{k'}\left(H_{j,t}^{k'}\right)^{\sigma_{k'}}\right)^{\frac{\sigma}{\sigma_{k'}}-1} \alpha_{j,t}^{k'}\left(H_{j,t}^{k'}\right)^{\sigma_{k'}-1}}{\alpha_{k,t}\left(\sum_{j=1}^{J} \alpha_{j,t}^{k}\left(H_{j,t}^{k}\right)^{\sigma_{k}}\right)^{\frac{\sigma}{\sigma_{k}}-1} \alpha_{j,t}^{k}\left(H_{j,t}^{k}\right)^{\sigma_{k}-1}} = \frac{p_j^{k'}}{p_j^{k}}. \tag{10}$$

3.2 Cost of Skill Mismatch

Define skill mismatch both at the individual level and the aggregate economy. We measure individual mismatch as wage gains by reallocating workers across occupations while keeping returns fixed. Second, we measure aggregate mismatch as output gains in a general equilibrium set-up where prices adjust when workers relocate across occupations.

3.2.1 Mismatch due to Frictions

Individual Mismatch If skills acquired in college major j earn the highest wage reward in occupation $k^*(j)$, we refer to this as the *top*-occupation for major j. The wage return to the combination of college major j and occupation k is captured by p_j^k. Thus, $p^* \equiv p_j^{k^*(j)}$ is the return to a top-match, while all remaining p_j^ks are wage returns to other matches. Suppressing time subscripts, we define average individual *within* major-occupation mismatch for major j as,

$$m_j = \frac{\sum_{i \in I^j} p^* \tilde{h}_i - \sum_{i \in I^j} p_j^k \tilde{h}_i}{\sum_{i \in I^j} p_j^k \tilde{h}_i}, \tag{11}$$

where I^j and I are the set of indexes for the individuals with major j and the set of all the individuals in the economy, respectively. Equation (11) calculates major-specific mismatch with fixed returns as average individual wage gains by allocating workers with major j to their top-occupation $k^*(j)$, relative to the total wages of workers with the same major j in the economy.

Having defined major-specific mismatch measure, the economy-wide individual mismatch can be defined as,

$$\mathcal{M} = \frac{\sum_{i \in I} p^* \tilde{h}_i - \sum_{i \in I} p_j^k \tilde{h}_i}{\sum_{i \in I} p_j^k \tilde{h}_i}, \tag{12}$$

average individual wage gains by allocating all workers to their top-occupation $k^*(j)$, relative to the total wages of all workers in the economy. This rate represents the average potential gains for all individuals if they were to *switch* to their top-occupation conditional on majors. By assumption returns $\{p_j^k\}$ are fixed even after the reallocation of workers. In reality, the reallocation of talent across occupations will result in a change in returns.

General Equilibrium Mismatch In general equilibrium, with frictions, the social planner will maximize societal income conditional on college major j for all individuals i,

$$\sum_i^I \max\{w_{i,k|j,t}\}_k \tag{13}$$

subject to the production functions (3) and (4). Thus workers will be relocated until Equation (7) is satisfied. Firstly, define the frictionless allocation of workers to occupations in year t as $\hat{\mathbf{H}}_t^k = \left(\hat{H}_{1,t}^k, \ldots, \hat{H}_{j,t}^k, \ldots \hat{H}_{J,t}^k\right)$ for all $k \in \mathcal{K}$. Obtaining the frictionless allocation $\hat{\mathbf{H}}_t^k$ for all occupations jointly requires solving $K \times J$ unknowns or the solution of the following $(K-1) \times J + J = K \times J$ equations,

$$\frac{\frac{\partial Y_t}{\partial Y_t^{k'}} \frac{\partial Y_t^{k'}}{\partial \hat{H}_{j,t}^{k'}}}{\frac{\partial Y_t}{\partial Y_t^{k}} \frac{\partial Y_t^{k}}{\partial \hat{H}_{j,t}^{k}}} = \frac{\alpha_{k',t} \left(\sum_{j=1}^J \alpha_{j,t}^{k'} \left(\hat{H}_{j,t}^{k'}\right)^{\sigma_{k'}}\right)^{\frac{\sigma}{\sigma_{k'}}-1} \alpha_{j,t}^{k'} \left(\hat{H}_{j,t}^{k'}\right)^{\sigma_{k'}-1}}{\alpha_{k,t} \left(\sum_{j=1}^J \alpha_{j,t}^{k} \left(\hat{H}_{j,t}^{k}\right)^{\sigma_{k}}\right)^{\frac{\sigma}{\sigma_{k}}-1} \alpha_{j,t}^{k} \left(\hat{H}_{j,t}^{k}\right)^{\sigma_{k}-1}} = 1, \tag{14}$$

and

$$\hat{H}_{j,t}^1 + \ldots + \hat{H}_{j,t}^k + \ldots + \hat{H}_{j,t}^K = \bar{H}_{j,t}, \quad \forall j \in \mathcal{J}, \quad \forall k \in \mathcal{K}, \tag{15}$$

where relative returns within major j across all occupations equal in equilibrium (see Equation (14)); and a market clearing condition that the optimal amount of labor across occupations within each major j equals the total supply of labor in major j, $\bar{H}_{j,t}$ (see Equation (15)). Equation (14) implicitly assumes that in a world without frictions or asymmetric information, all occupations should pay the same conditional on major. Denote the price in the optimal allocation in the general equilibrium as

$$\hat{p}_j^k = \frac{\partial Y_t}{\partial H_{j,t}^{k'}} \left(\hat{\mathbf{H}}_t^1, ..., \hat{\mathbf{H}}_t^K \right). \tag{16}$$

In the optimal allocation, the returns to each major are the same across occupations. Thus, we can define a unique price for each major j as

$$\hat{p}_j \equiv \hat{p}_j^1 = ... = \hat{p}_j^K. \tag{17}$$

Then, the general equilibrium mismatch for major j is defined as

$$m_j^{GE} = \frac{\sum_{i \in I^j} \hat{p}_j \tilde{h}_i - \sum_{i \in I^j} p_j^k \tilde{h}_i}{\sum_{i \in I^j} p_j^k \tilde{h}_i}. \tag{18}$$

The aggregate general equilibrium mismatch can be defined as

$$\mathcal{M}^{GE} = \frac{\sum_{i \in I} \hat{p}_j \tilde{h}_i - \sum_{i \in I} p_j^k \tilde{h}_i}{\sum_{i \in I} p_j^k \tilde{h}_i}. \tag{19}$$

The above definition represents the potential gains by moving to the frictionless allocations.[8] The concept is equivalent to the partial equilibrium mismatch, but with endogenous prices.

4 Data

We make use of the ALife dataset, a longitudinal dataset administered by the Australian Tax Office (ATO) (Carter et al., 2021). ALife contains an extract of personal income tax records from 1991 to 2019 for a 10 percent representative sample. The data provide accurate and detailed information on variables that are required to estimate individual tax liabilities. It collates employer records, salary information, and self-reported data from taxpayers. For our context, ALife also contains details on Higher Education loan information or the study records of terminal degrees.[9]

[8]The aggregate mismatch measure can be also defined as

$$\mathcal{M}_t^{GE} = \frac{F\left(F^1\left(\hat{\mathbf{H}}_t^1\right), ..., F^K\left(\hat{\mathbf{H}}_t^K\right)\right) - F\left(F^1\left(\mathbf{H}_t^1\right), ..., F^K\left(\mathbf{H}_t^K\right)\right)}{F\left(F^1\left(\mathbf{H}_t^1\right), ..., F^K\left(\mathbf{H}_t^K\right)\right)}, \tag{20}$$

where $\mathbf{H}_t^k = \left(H_{1,t}^k, ..., H_{j,t}^k, ... H_{J,t}^k\right)$ represents the current allocation of majors across occupations.

[9]For additional data details see Appendix A.

Given the administrative nature of ALife, the data has very little attrition, with on average, 96.5 percent of tax filers who lodge in a given year also lodge in the subsequent income year (Carter et al., 2021). The dataset contains around 300 variables from individual tax records, including some constructed variables, such as the presence of a spouse. Demographic information includes age at 30 June of a given year, gender, residential location, occupation, and whether the tax filer is a non-resident for tax purposes. A caveat, given that occupation information, is self-reported and prefilled in later tax years when the tax filing system moved to a predominantly online reporting basis, there is potential for misreporting and limited occupation switching (Hathorne and Breunig, 2022). To ameliorate the potential noise in occupation information, we restrict our sample to college-educated individuals aged 21 to 35.[10] We also restrict the sample to the workers who only choose one specific major in the college. This implicitly assumes that each individual has only one *top*-occupation which is solely determined by his/her unique college major.[11]

Table 1: Summary Statistics of Occupation Share by Major

	Occupations by Major									
	Share			Switch-in			Stay			
2-digit ASCED Major	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	No.Occ
Natural and Physical Sciences	0.038	0.005	0.325	0.005	0.001	0.026	0.034	0.004	0.299	26
Information Technology	0.050	0.004	0.471	0.006	0.001	0.036	0.044	0.004	0.435	20
Engineering and Related Technologies	0.042	0.002	0.582	0.003	0.000	0.022	0.039	0.001	0.561	24
Architecture and Building	0.062	0.012	0.443	0.005	0.001	0.023	0.058	0.010	0.420	16
Agriculture, Environmental and Related Studies	0.043	0.008	0.349	0.005	0.001	0.024	0.038	0.006	0.325	23
Health	0.037	0.001	0.654	0.002	0.000	0.019	0.035	0.001	0.635	27
Education	0.037	0.002	0.696	0.002	0.000	0.020	0.035	0.001	0.676	27
Management and Commerce	0.037	0.003	0.397	0.004	0.001	0.029	0.033	0.002	0.368	27
Society and Culture	0.037	0.005	0.240	0.004	0.001	0.017	0.033	0.004	0.223	27
Creative Arts	0.037	0.002	0.130	0.005	0.001	0.019	0.032	0.002	0.118	27

Notes: This table shows the summary statistics of the occupational share for each field of study. Columns (2)-(4) show the mean, minimum, and maximum of the occupational share within the field in the corresponding row. Columns (5)-(7) show the mean, minimum, and maximum of the share of staying in the observed occupation from the previous period. Columns (8)-(10) show the the mean, minimum, and maximum of switching into the observed occupation from the previous period. Column (11) show the number of occupations within the field. The product of the mean occupational share and number of occupations within the field is not 1, since estimation sample removes the field-by-occupation cells that contain less than 100 observations.

Table 1 provides a snapshot of college majors in our data. We consider all 11 ASCED 2-digit college majors, while occupations are classified based on 2-digit ANZSCO codes. The last column shows the number of occupations within each field, while the second column shows

[10] Hathorne and Breunig (2022) report similar findings across HILDA, a national representative sample, and ALife for the age group 25 to 38.

[11] Our results show that the proportion of workers with more than one major as classified under the 2-digit ASCED fields of study account for 19.6 percent of all workers.

the average of occupational shares within each major. We also report the share of individuals switching into occupations over time and the share that are stayers by occupation within their major.

As shown in Table 1, number of occupations has a large variance across majors. Management and Commerce, and Society and Culture majors match to 27 occupations, while Food, Hospitality, and Personal Services majors match to only 11 occupations. The majors with the lowest switching shares are Education and Health (as seen in column 5), consistent with the notion that educational and medical occupations are highly regulated and require specialized skills.

Observing large heterogeneity across fields of study in the number of occupational matches, employment shares, and switches, we study wage penalties from working in an occupation relative to the *top*-occupation of an individual's major. There are potentially many ways of defining the *top*-occupation. In our baseline, we use both wage and share information—by ranking occupations within fields of study by their wages subject to having at least a 10 percent share of the graduates of the major under consideration. We define the *top*-occupation as the occupation with the highest average wage among this set.[12] We use the 10 percent share, as we are not interested in comparing mismatch relative to a *superstar*, but relative to a typical college graduate.

Table 2: Summary Statistics of Top Occupation by Major

		Share of Occ. by Field			
2-digit ASCED Major	Top-Occupation	Total	Stay	Switch-in	Earnings
Natural and Physical Sciences	Design, Engineering, Science and Transport Professionals	0.325	0.299	0.026	76,348
Information Technology	ICT Professionals	0.471	0.435	0.036	87,449
Engineering and Related Technologies	Design, Engineering, Science and Transport Professionals	0.582	0.561	0.022	94,139
Architecture and Building	Engineering, Ict and Science Technicians	0.136	0.123	0.013	72,533
Agriculture, Environmental and Related Studies	Design, Engineering, Science and Transport Professionals	0.349	0.325	0.024	76,476
Health	Health Professionals	0.654	0.635	0.019	69,172
Education	Education Professionals	0.696	0.676	0.020	65,211
Management and Commerce	Business, Human Resource and Marketing Professionals	0.397	0.368	0.029	85,057
Society and Culture	Legal, Social and Welfare Professionals	0.240	0.223	0.017	74,620
Creative Arts	Business, Human Resource and Marketing Professionals	0.106	0.088	0.019	69,335

Notes: This table shows the summary statistics of the top occupational for each field of study. Column (3) shows the share of top occupation of the field in the corresponding row. Column (4) shows the share of staying in the top occupation from the previous period. Column (5) shows the share of switching into the top occupation from the previous period. Column (6) shows the mean earnings in the field-by-occupation combination.

Table 2 shows the top occupation by major and their employment share within a major. We also decompose the share by the share of stayers and switchers. The last column report

[12] We also perform robustness by: (1) ranking occupations based on the highest wage returns within majors with at least 5 percent share of graduates, and (2) based only on the highest wage returns within majors, and (3) based only on their highest share of employment within majors. *Top*-occupations remain largely unchanged in these permutations.

average yearly earnings of *top*-occupations. Comparing "Switch-in" columns between Table 2 and Table 1, *top*-occupation is typically the one that workers are most likely to switch in.

5 Estimation

The estimation has two parts: (1) estimating the returns to college majors by occupation, and (2) estimating the production function parameters. The second part requires us to have consistent estimates of the returns to college majors by occupation.

5.1 Returns to College Majors

To allow us to capture some of the richness in the data when estimating college returns, we introduce further characteristics to individual efficiency units of labor, a_i, in Equation (1). Suppose that, after graduating from college, a worker enters the labor market and provides college-major-specific human capital. His/her wage equation is then

$$w_{i,j,k,r,c,t} = p^k_{j,t}\tilde{h}^j_{i,t} = p^k_{j,t}h^j a_{i,t} = p^k_{j,t}h^j e^{X_{i,t}\gamma+\lambda_{r,t}+\nu_c+\epsilon_{i,j,k,c,r,t}}, \tag{21}$$

where individual abilities is defined by, $X_{i,t}$ a vector of time-varying individual characteristics in year t, $\lambda_{r,t}$ a region-by-year specific productivity, and ν_c a graduate-cohort specific productivity, and an idiosyncratic productivity shock $\epsilon_{i,j,k,c,r,t}$.

5.1.1 Occupational Choice Model conditional on College Major

Expanding on the occupation choice model from Equation (8), individual i, who graduates in major j of cohort c, chooses occupation k in calendar year t. The overall utility of choosing occupation k given major j is determined by log earnings $w_{i,j,k,c,r,t}$ and non-pecuniary utility $\eta_{i,(k|j),c,r,t}$ received from unobserved occupational characteristics (i.e., working conditions, taste, etc.). The utility from Equation (8) can be rewritten in terms of a conditional mean

$\tilde{U}_{i,(k|j),c,r,t}$ and an individual-specific shock $e_{i,(k|j),c,r,t}$ such that,

$$\begin{aligned} U_{i,(k|j),c,r,t} &= \underbrace{w_{i,j,k,c,r,t}}_{\text{Log Earnings}} + \underbrace{\eta_{i,(k|j),c,r,t}}_{\text{Non-pecuniary Utility}} \qquad \text{for each } j \in \{1,\ldots,J\} \quad (22) \\ &= \underbrace{\tilde{U}_{i,(k|j),c,r,t}}_{\text{Conditional Mean}} + \underbrace{e_{i,(k|j),c,r,t}}_{\text{Shock}} \\ &= \underbrace{E[w_{i,j,k,c,r,t}|W_{i,t}] + E(\eta_{i,(k|j),c,r,t}|W_{i,t})}_{\text{Conditional Mean: } \tilde{U}_{i,(k|j),c,r,t}} + \overbrace{\underbrace{\epsilon_{i,j,k,c,r,t}}_{\substack{\text{Errors in}\\ \text{Earnings Equation}}} + \underbrace{\kappa_{i,(k|j),c,r,t}}_{\substack{\text{Non-pecuniary Utility}\\ \text{Shock}}}}^{\text{Shock: } e_{i,(k|j),c,r,t}}, \end{aligned}$$

where $\tilde{U}_{i,(k|j),c,r,t}$ is the utility conditional on $W_{i,t}$, which includes a vector of individual observed characteristics by calendar year t and any exclusion restrictions for identifying the occupation preference from earnings. Utility shock of occupational choice $e_{i,(k|j),c,r,t}$ consists of earnings shock $\epsilon_{i,j,k,c,r,t}$ and non-pecuniary utility shock $\kappa_{i,(k|j),c,r,t}$. For maximizing the utility, individual i chooses occupation k in year t if and only if $U_{i,(k|j),c,r,t} \geqslant U_{i,(k'|j),c,r,t}$; $\forall\ k' \neq k$. We use indicator $M_{i,(k|j),c,r,t}$ to denote the event that we observe an individual who chooses occupation k given major j so that,

$$M_{i,(k|j),c,r,t} = 1 \text{ iff } e_{i,(k'|j),c,r,t} - e_{i,(k|j),c,r,t} \leqslant \tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(k'|j),c,r,t};\ \forall\ k' \neq k. \quad (23)$$

5.1.2 Correct Selection Bias: Parametric Correction Function

Specifically, individuals are observed in an j,k cell if and only if the rule (23), conditional on major j, is satisfied. There is potential selection bias—$E[\epsilon_{i,j,k,c,r,t}|M_{i,(k|j),c,r,t} = 1] \neq 0$, as the unexplained factor $\epsilon_{i,j,k,c,r,t}$ can be correlated with unobserved factors that are deterministic for the observations to appear in (j,k) cell. Directly using condition (23) to pin down the selectivity bias is infeasible because a complete specification of the joint distribution of error terms in the wage equation and error terms in K selection equations for each major j would be necessary, where K is the number of occupations for major j to choose. We follow the approach by Lee (1983) and Dahl (2002) to correct for the sample selection bias for observations in j,k cell by controlling for the correction function of selection probability—$g(\hat{p}_{i,j,k,c,r,t})$, where $\hat{p}_{i,j,k,c,r,t}$ is the estimated probability of choosing the observed occupation k conditional on graduating from major j, and $g(x)$ has a known inverse Mills ratio of the form: $\frac{-\phi(\Phi^{-1}(x))}{x}$. Details of the derivation of the selectivity bias can be found in Appendix B.1.

To this end, the expectation of log earnings after correcting for the selectivity bias becomes,

$$
\begin{aligned}
E[\ln w_{i,j,k,c,r,t}|M_{i,(k|j),c,r,t} = 1] \\
&= \sum_{j'}^{J}\sum_{k'}^{K}\beta_{j'}^{k'} D_{j',k'} + X_{i,t}\gamma + \lambda_{r',t'} + \nu_{c'} + \iota_{i'} + E[\epsilon_{i,j,k,c,r,t}|M_{i,(k|j),c,r,t} = 1] \\
&= \sum_{j'}^{J}\sum_{k'}^{K}\beta_{j'}^{k'} D_{j',k'} + X_{i,t}\gamma + \lambda_{r',t'} + \nu_{c'} + \iota_{i'} + \sum_{j'}^{J}\sum_{k'}^{K}\rho_{j'}^{k'} \cdot g(\hat{p}_{i,j,k,c,r,t}), \quad (24)
\end{aligned}
$$

where $w_{i,j,k,r,t}$ is annual earnings, $\beta_{j'}^{k'}$ is the rate of return to major-j-specific human capital in occupation k (or logarithm of the price ($p_{j'}^{k'}$) of major j' in occupation k'), $D_{j',k'}$ is a dummy variable that takes 1 if $k' = k$ and $j' = j$, and $\epsilon_{i,j,k,c,r,t}$ is the error term. Depending on the specification we allow returns to vary by region and/or year.

The earnings model that is corrected for selection bias becomes,

$$
\ln w_{i,j,k,c,r,t} = \sum_{j'}^{J}\sum_{k'}^{K}\beta_{j'}^{k'} D_{j',k'} + X_{i,t}\gamma + \lambda_{r',t'} + \nu_{c'} + \iota_{i'} + \sum_{j'}^{J}\sum_{k'}^{K}\rho_{j'}^{k'} \cdot g(\hat{p}_{i,j,k,c,r,t}) + \xi_{i,j,k,c,r,t}
$$
$$
\text{where } E[\xi_{i,j,k,c,r,t}|M_{i,(k|j),c,r,t} = 1] = 0, \quad (25)
$$

where $g(\hat{p}_{i,j,k,c,r,t})$ is the selection correction function for individuals appearing in cell j,k, with a known functional form of the single estimator $\hat{p}_{i,j,k,c,r,t}$. In total, we use $J \times K$ correction functions, with the effect of each one on log earnings captured by the j,k specific parameter, $\rho_{j,k}$.

As the occupation choice set varies considerably across majors (see Table 1), to make estimation feasible, we reduce the number of correction functions into $J \times 10$. Specifically, for all individuals with major j, we partition the entire observed (major-specific) occupational set into ordered 10 sets based on the descending order of occupational share within the major: with (1) the first set comprising the *top*-occupation for major j (i.e., the occupation with the highest wage conditional on a 10 percent share of the major j); and (2) the remaining nine even sets of occupations for major j, ordered with a decrease of the occupational share to major j.[13] We estimate the individual's occupational choice probabilities conditional on the field of study j with the individual characteristics and an exclusion restriction—a one period lagged of local

[13] Occupations are approximately evenly placed in each set based on the descending order of occupational share within the major, so that the second occupation set is more major-related than the third set, so on so forth. We also perform robustness exercises with 8 and 12 occupation groups, respectively. For the major of 'Food, Hospitality and Personal Services', there are in total 11 occupations in the sample, and thus there are 11 sets in total in robustness check, in which a 12-occupation-group partition is conducted. Results are similar across the three versions.

market demand shock—by a random forest algorithm (see Appendix B.2 for more detail).[14]

5.2 Estimating Production Function Parameters

To solve for the frictionless allocation through Equations (14) and (15), and to define the mismatch measures through Equation (19), we need to know the parameter values of the production function. We first derive the reduced-form expressions from the first-order conditions of the representative firm profit maximization problem. We then exploit the region-time variations to estimate the production function parameters.

5.2.1 Deriving the Reduced Form Expressions

We take the first-order condition for Equation (3) that yields,

$$\frac{p^k_{j',r,t}}{p^k_{j,r,t}} = \frac{\partial F^k/\partial H^k_{j',r,t}}{\partial F^k/\partial H^k_{j,r,t}} = \frac{\alpha^k_{j',t}}{\alpha^k_{j,t}}\left(\frac{H^k_{j',r,t}}{H^k_{j,r,t}}\right)^{\sigma^k-1}. \tag{26}$$

Taking the logarithm on both sides yields,

$$\underbrace{\ln\left(p^k_{j',r,t}/p^k_{j,r,t}\right)}_{\beta^k_{j',r,t}-\beta^k_{j,r,t}} = \ln\left(\alpha^k_{j',t}/\alpha^k_{j,t}\right) + \left(\sigma^k - 1\right)\ln\left(H^k_{j',r,t}/H^k_{j,r,t}\right). \tag{27}$$

Equation (27) is the first equation we estimate with the data. To address the endogeneity issue of labor supplies, we use a Bartik-style instrument.

Next, the first-order condition for Equation (4) is,

$$\frac{p_{k',r,t}}{p_{k,r,t}} = \frac{\partial F/\partial Y_{k'}}{\partial F/\partial Y_k} = \frac{\alpha_{k'}}{\alpha_k}\left(\frac{Y_{k',r,t}}{Y_{k,r,t}}\right)^{\sigma-1}. \tag{28}$$

[14] In a robustness exercise we also consider a version with individual fixed effects, ι_i, such that human capital equals, $\tilde{h}^j_{i,t} = h^j e^{X_{i,t}\gamma+\lambda_{r,t}+\nu_c+\iota_i}$. in Equation (1). When estimating Equation (25) with individual fixed effects, major-occupation returns are identified from individuals that switch occupations at least once in their life (roughly one-third of our sample, and smaller for skill-specific majors such as Education and Health). To recover all returns, β^k_js, and individual fixed effects, ιs, we estimate relative returns by college major and then recover absolute returns, such that $\beta^1_j = \sum_i \frac{\iota_i}{N_j}$, where N_j are all individuals who studied major j and $\beta^k_j = \sum_i \frac{\iota_i}{N_j} + \beta^k_j$ for all $k > 1$.Individual fixed effects that feed into $\tilde{h}_{i,t}$ (see Equation (1)) are thus demeaned, $\hat{\iota}_i = \iota_i - \sum_i \frac{\iota_i}{N_j}$.

By taking the logarithm of Equation (28), we obtain,

$$\ln\left(\frac{p_{k',r,t}}{p_{k,r,t}}\right) = \ln\left(\frac{\alpha_{k',t}}{\alpha_{k,t}}\right) + (\sigma - 1)\ln\left(\frac{Y_{k',r,t}}{Y_{k,r,t}}\right), \tag{29}$$

where Y_k is given by Equation (3). Using the equilibrium profit condition, $p_{k,r,t}Y_{k,r,t} = \sum_j p^k_{j,r,t}H^k_{j,r,t}$, we solve for the price of occupation k in region r at time t,

$$p_{k,r,t} = \frac{1}{A_{k,t}}\left(\sum_j \left(\alpha^k_{j,t}\right)^{-\frac{1}{\sigma_k - 1}} \left(p^k_{j,r,t}\right)^{\frac{\sigma_k}{\sigma_k - 1}}\right)^{(1-\frac{1}{\sigma_k})}. \tag{30}$$

Substituting Equation (30) for Equation (29) yields,

$$\ln\left(\frac{\left(\sum_j \left(\alpha^{k'}_{j,t}\right)^{-\frac{1}{\sigma_{k'}-1}} \left(p^{k'}_{j,r,t}\right)^{\frac{\sigma_{k'}}{\sigma_{k'}-1}}\right)^{(1-\frac{1}{\sigma_{k'}})}}{\left(\sum_j \left(\alpha^{k}_{j,t}\right)^{-\frac{1}{\sigma_{k}-1}} \left(p^{k}_{j,r,t}\right)^{\frac{\sigma_{k}}{\sigma_{k}-1}}\right)^{(1-\frac{1}{\sigma_{k}})}}\right) = \ln\left(\frac{\alpha_{k',t}}{\alpha_{k,t}}\right) + \ln\left(\frac{A_{k',t}}{A_{k,t}}\right) + (\sigma - 1)\ln\left(\frac{Y_{k',r,t}}{Y_{k,r,t}}\right). \tag{31}$$

This is the second equation we estimate.

Note technology parameters, $\ln\left(\frac{\alpha_{k',t}}{\alpha_{k,t}}\right) + \ln\left(\frac{A_{k',t}}{A_{k,t}}\right)$ are not separately identifiable, but this is not crucial for our exercise. A typical assumption in the literature (see Katz and Murphy (1992)) is to assume a linear time trend in technological change, e.g., the expression can be approximated by,

$$\ln\left(\frac{\alpha_{k',t}}{\alpha_{k,t}}\right) + \ln\left(\frac{A_{k,t}}{A_{k',t}}\right) = \gamma_{0,kk'} + \gamma_{1,kk'}t,$$

when using time-series data or simply $\gamma_{0,kk'}$ if using only cross-sectional variation. Note, throughout the analysis we assume that the production technologies are the same across all regions in Australia, but potentially vary over time.

5.2.2 Exploiting Time and Regional Variations

For the occupational-specific production functions, estimating parameters $\{\alpha^k_j\}$ and σ_k from Equation (29), we exploit variations in $p^k_{j'}$ and $H^k_{j'}$. For the aggregate production function, estimating $\{\alpha_k\}$ and σ from Equation (31), we again require variations in $Y^{k'}$ and $p^{k'}$. For the baseline estimation of the production function parameters, we assume $\{\alpha^k_j\}$, $\{\sigma_k\}$, $\{\alpha_k\}$, and σ are fixed across time and regions and exploit the region and year variations.

Due to the sampling limitations, we group years into 7 year groups to estimate Equations (29) and (31). Therefore, both general and partial equilibrium evolution in mismatches are computed by Equations (12) and (19) over 7 year groups for comparison purposes. We leave further detail of parameters estimation to Appendix C. Table D1 in Appendix C shows the estimation results for occupation-specific production function parameters. Our results on the elasticity of substitution between occupations, $\frac{1}{1-\sigma} = 1.830$, are comparable to Burstein et al. (2019) and Caunedo et al. (2021), who estimate a model with capital.

6 Results

We illustrate the results from our framework using Australian data. We mainly focus on individual life-cycle profiles to discuss and contrast findings from the literature so far, both with fixed returns and under our newly developed general equilibrium framework. We additionally, also show the aggregate evolution of mismatch over time and discuss what drives the observed trends within the context of Australia. These results together provide a broad picture of the usability of our measure and skill mismatch in the economy. They especially contrast the large difference between the general equilibrium and typical fixed-return framework.

6.1 Base Estimates

Table 3 shows average major-specific mismatch based on Equation (11) and overall mismatch in the aggregate economy based on Equation (12) for the average worker from 2003 to 2019, with standard errors calculated using the delta method.[15] Results shown are of both OLS and selection corrected returns, estimates of β_j^k, from Equation (25).[16] Comparing column 3 with column 5, we find that correcting for selection increases the average mismatch for typical STEM-related majors (in particular, Natural and Physical Sciences, Information Technology, and Engineering and related Technologies), Management and Commerce, Society and Culture, and Creative Arts. These are also the majors with large average mismatch. The estimated baseline mismatch in the aggregate economy, calculated using Equation (12), is 28.3 percent.

[15] Specifically, mismatch estimates in equations (11) and (12) are functions of vector $\boldsymbol{\beta}$—i.e., $h(\boldsymbol{\beta})$, where $\boldsymbol{\beta} = (\beta_j^1, ..., \beta_j^{k^*}, ..., \beta_j^k)$. Based on the multivariate delta method, we compute the variance of $h(\boldsymbol{\beta})$ using, $Var(h(\boldsymbol{\beta})) = \nabla h(\boldsymbol{\beta})^T Cov(\boldsymbol{\beta}) \nabla h(\boldsymbol{\beta})$. To reduce the computational burden, we consider $\tilde{h}_i$ to be constant.

[16] In Appendix E we report all wage differences βs relative to the top-match occupation across college majors (see Figure F1). Based on these major-occupation returns, Section C, details the estimation procedure for the production parameters and Table D1 provides the estimated parameters.

Table 3: Individual Major-Specific Mismatch

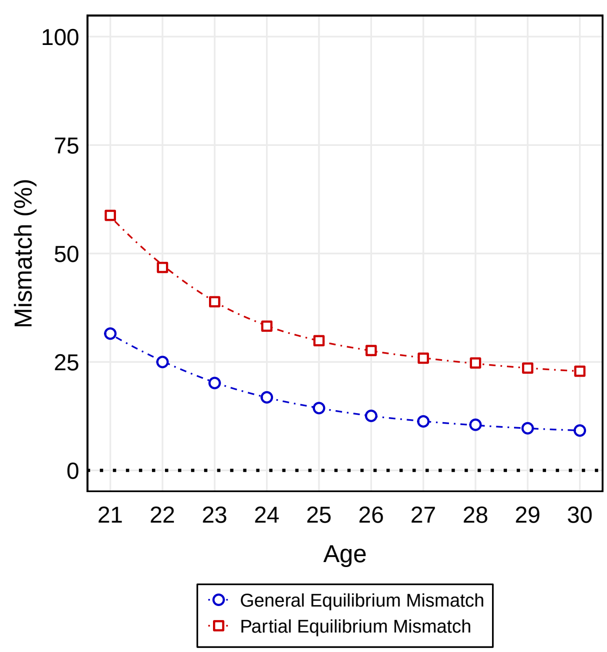
ASCED 2-digit	Major	OLS	S.E	Correction	S.E	GE
1	Natural and Physical Sciences	0.150	0.004	0.219	0.019	0.033
2	Information Technology	0.133	0.003	0.293	0.016	0.099
3	Engineering and Related Technologies	0.107	0.001	0.306	0.010	0.083
4	Architecture and Building	0.071	0.010	0.044	0.020	-0.009
5	Agriculture, Environmental and Related Studies	0.152	0.005	0.175	0.022	0.115
6	Health	0.056	0.001	0.054	0.004	0.055
7	Education	0.076	0.001	0.120	0.005	0.070
8	Management and Commerce	0.151	0.001	0.433	0.009	0.228
9	Society and Culture	0.199	0.003	0.397	0.019	0.119
10	Creative Arts	0.237	0.007	0.372	0.018	0.167
	Aggregate	0.127	0.003	0.270	0.012	0.125

Notes: This table shows the individual mismatch for each major. Column (3)-(4) show the mismatch estimates generated by OLS estimation approach and corresponding standard error. Column (5)-(6) show mismatch estimates generated by correction function estimation approach and corresponding standard error. Standard errors are computed using delta method. The last rows show the aggregated individual-level mismatch.

6.2 Life-Cycle Skill Mismatch

Skill mismatch appears substantially smaller when considering how general equilibrium forces adjust occupation-major returns with the reallocation of talent. How this cost is distributed across different individuals should guide any policy intervention that aims at reallocating talent or retraining workers. A thorough life-cycle analysis can shed light on these issues.

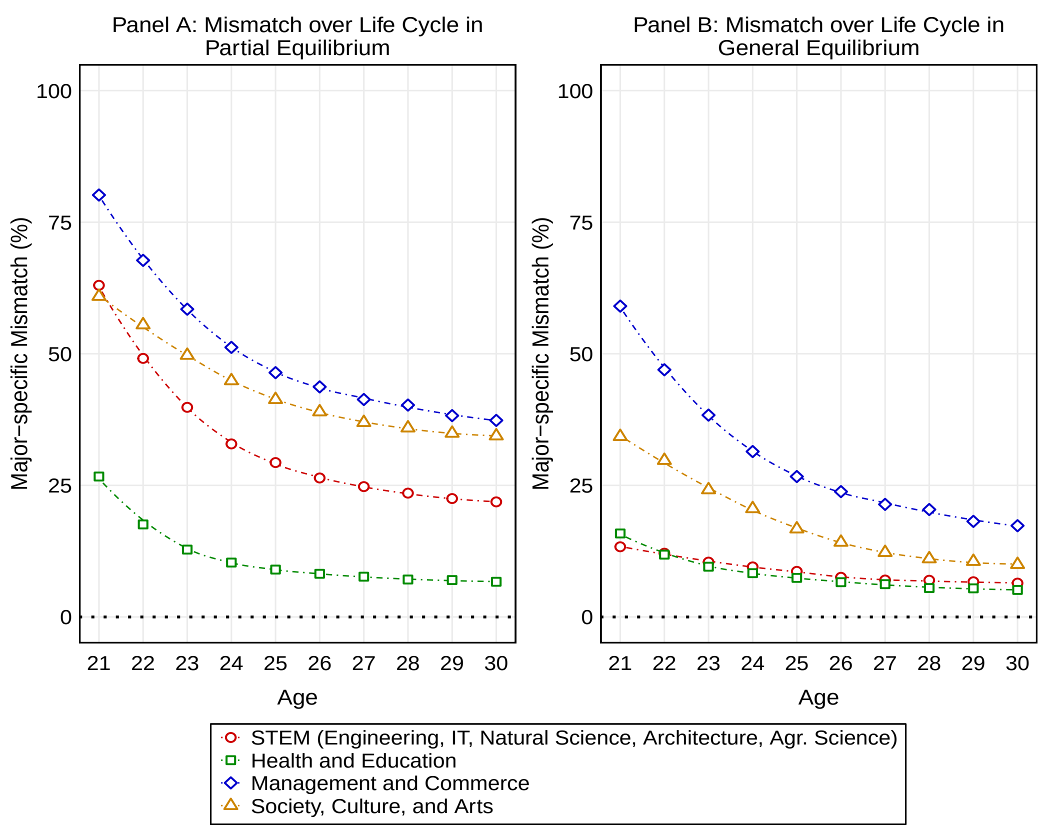
Figure 1: Life-cycle Mismatch in the General and Partial Equilibrium



Notes: This figure plots mismatch in the partial and general equilibrium using the Equation (12) over the life cycle.

We compute mismatch within each college major by age in two ways, (i) using Equation (11)[17] and (ii) by again using Equation (18), but replacing prices with equilibrium prices from Equation (17) after the reallocation of workers. We start with fixed returns, Figure 1 plots the result for five major groups. Panel (a) shows the life-cycle evolution with fixed returns, and panel (b) shows the share of individuals in the top-matched occupation over the life-cycle. The three groupings of STEM, Management and Commerce, and Society, Culture, and Arts majors have the highest mismatch at the typical age of graduation (age 21), but mismatch declines significantly as workers age. The level of mismatch is less than half in Health and Education compared to Commerce and Social Science and Arts majors. Moreover, the slope of the decline of mismatch is flatter in Health and Education majors than in STEM; Management and Commerce; and Society, Culture, and Arts majors. Lastly, mismatch is relatively flat in Food, Hospitality and Personal Services majors over the life-cycle.[18]

Figure 2: The Major-Specific Mismatch in Partial and General Equilibrium over the Life-Cycle



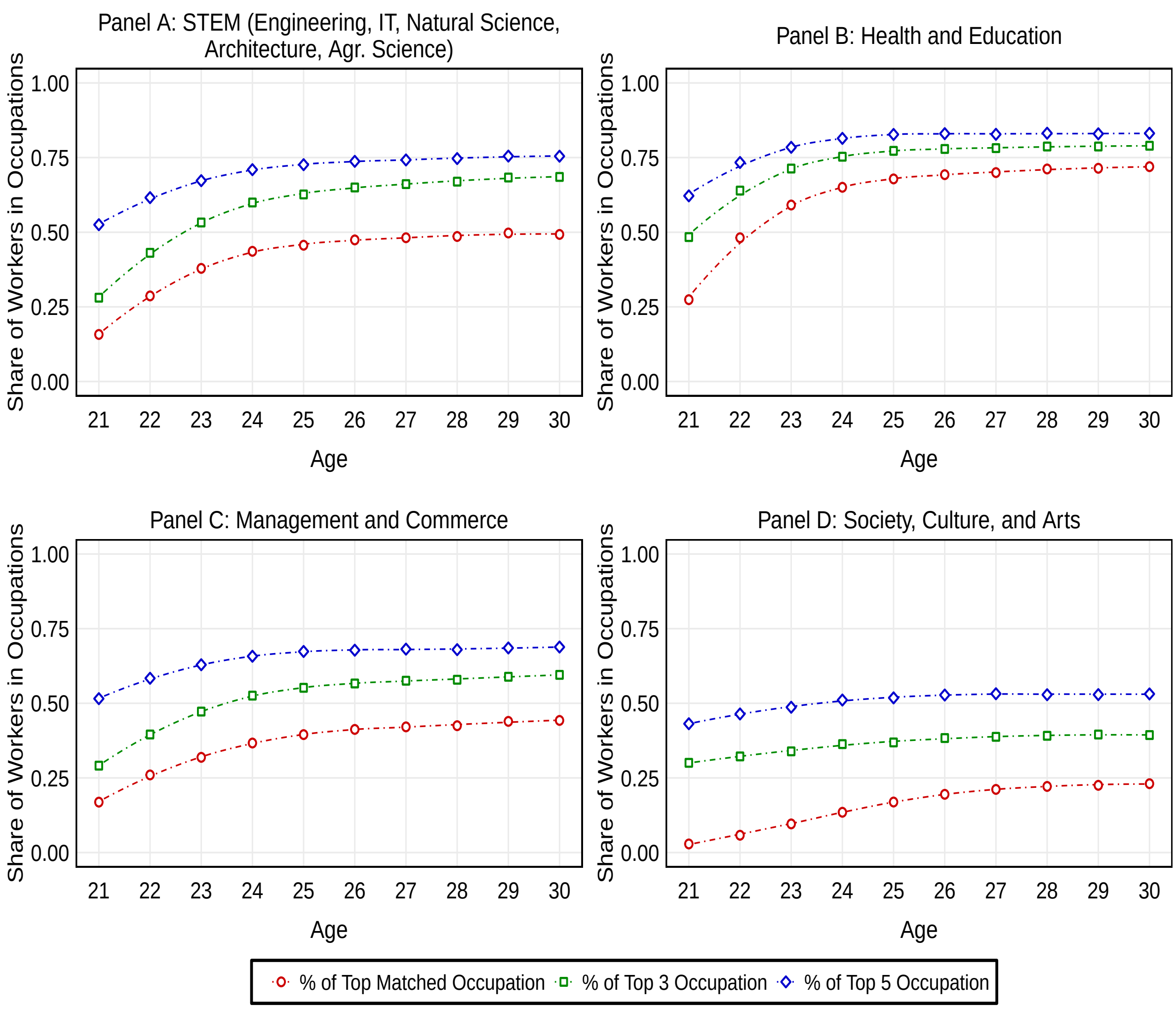
Notes: Panel A and Panel B plot the mismatch in the partial and general equilibrium for aggregate fields of study using the Equation (12) over the life-cycle.

[17] Specifically, we compute mismatch for workers who studied major j and worked at age a—that is, replacing $i \in I^j$ with $i \in I^j_a$ in Equation (11).

[18] We plot the difference in major-specific mismatch between general and partial equilibrium in the Figure F3.

Part of the falling mismatch with age is due to improved sorting in the labor market as seen in Panel B of Figure 1, which plots the share of individuals in the top-occupation. The finding is consistent with the previous literature on US data (Guvenen et al., 2020). As is evident in all majors but Hospitality, Food, and Personal Service, the share of workers in top-paying occupations increases with age. Notable is that these shares are lowest in Commerce, Social Science and Arts and highest in Health and Education consistent with a concept of general versus specific skills.

Figure 3: Share of the Top Occupation within Major over the Life-Cycle



Notes: This figure plots the share of top-matched occupation, top 3 and top 5 occupations within each field of study over the life cycle.

In contrast, Panel A of Figure 2 plots the same life-cycle profiles with general equilibrium occupation-major returns. Panel B plots the difference between the general equilibrium and fixed return results. A number of striking patterns emerge. First, in general equilibrium, the cost of skill mismatch is magnitudes smaller for all majors but Health and Education. Second, skill mismatches decreases to less than 10% by age 30 irrespective of major. Comparing the general equilibrium and fixed return results (Panel B) shows that mismatch is roughly 25 percentage points larger for most ages for STEM, Management and Commerce and Society, Culture and Arts majors.

The results suggest estimates of skill mismatch that ignore general equilibrium reallocation effects will be overstated. This is especially true, as our baseline results represent the upper bound of the cost of skill mismatch as we attribute all mismatch to frictions and none to taste. Nonetheless, for some groups skill mismatch exists even in the general equilibrium. Policies concerned about skill mismatch should specifically target young workers, as over time, when considering general equilibrium forces, all the cost of mismatch disappears. However, if active policies to improve the allocation of talent are beneficial for society and individuals' lifetime welfare is beyond the scope of this paper.

6.3 Evolution of Aggregate Mismatch

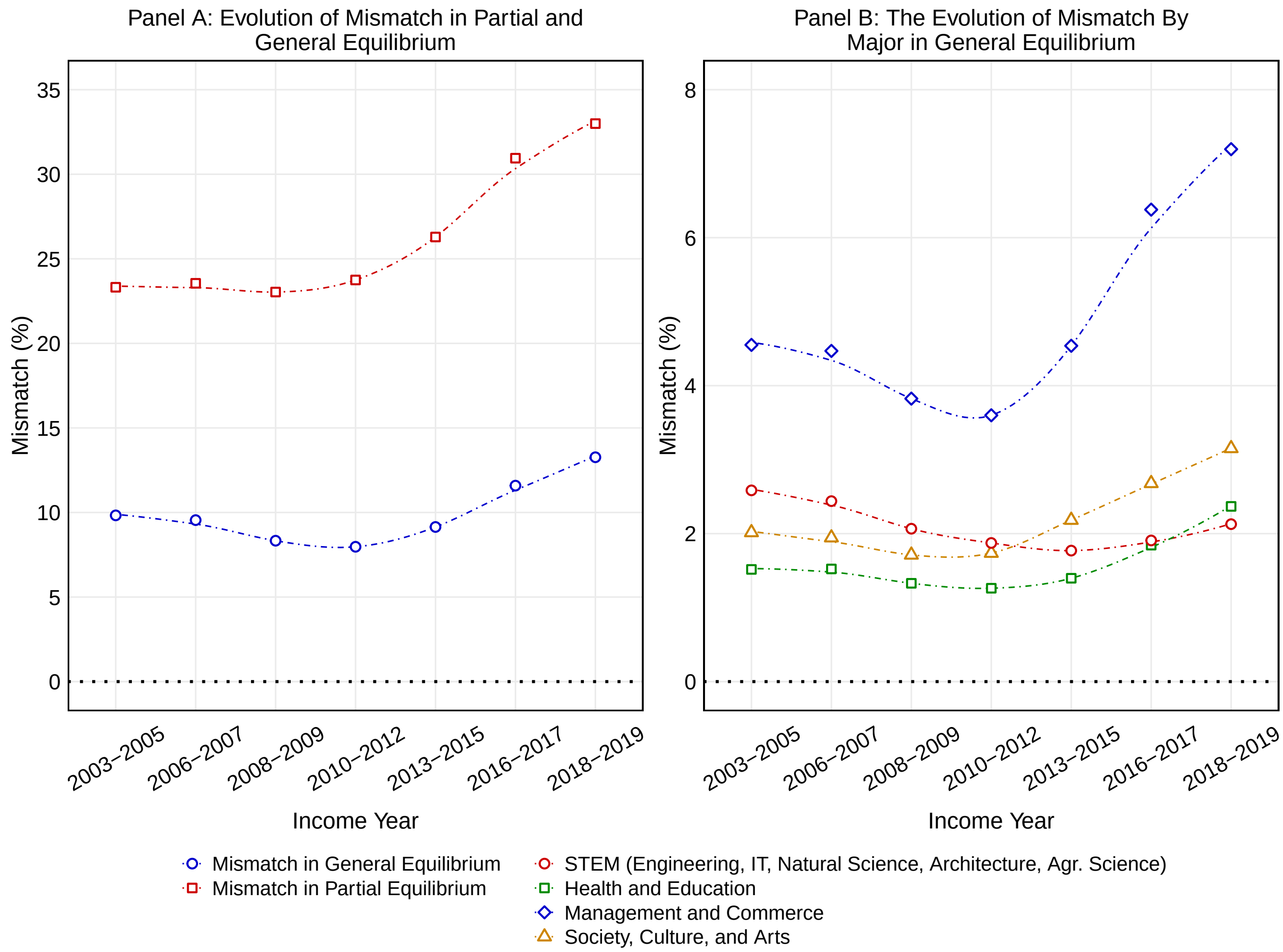
Using estimated selection-corrected return and production function parameters, Panel A of Figure 4 illustrates the evolution of aggregate mismatch for both the average worker from Equation (12) and general equilibrium evaluated from Equation (19) at each year (group). While the average individual mismatch is 28.3 percent, the equivalent aggregated mismatch in general equilibrium over the same period, as per Equation (19), is 10.0 percent.[19]

The evolution of mismatch is relatively stable until 2012, but increasing thereafter. Although the time evolution follows a similar trend with fixed returns and in the general equilibrium, the relative rise in aggregate output loss over time is roughly one-third in general equilibrium. This suggests that although workers may demand higher wages due to a perceived shortage of talent, the forces of the general equilibrium will partially offset these pressures. In other words, as talent is reallocated within the economy to optimize output, wages will adjust accordingly.

Panel B of Figure 4 shows that further decomposing the output loss by major group shows

[19]These baseline estimates of mismatch in general equilibrium assume a production function with linear time trends in production shares—α_j^ks and α_ks over time. For details on the estimation procedure see Appendix C.

Figure 4: The Evolution of Mismatch in the General Equilibrium



Notes: Panel A plots the evolution of estimated output loss in the partial and general equilibrium from year group 2003-2005 to 2018-2019 computed by Equations (12) and (19), respectively. Panel B plots the evolution of estimated output loss due to misallocation disaggregated into five major categories from year group 2003-2005 to 2018-2019. Each major group mismatch is computed by Equation (32) for each year group.

some heterogeneity among college majors.[20][21] We find that all majors experience a relatively

[20]We decompose the aggregate mismatch for major j in terms of GDP per capita loss by:

$$m_{j,t}^{GE} = \frac{F\left(F^1\left(\hat{\mathbf{H}}_t^1\right),\ldots,F^K\left(\hat{\mathbf{H}}_t^K\right)\right) - F\left(F^1\left(\hat{\mathbf{H}}_{-j,t}^1\right),\ldots,F^K\left(\hat{\mathbf{H}}_{-j,t}^K\right)\right)}{F\left(F^1\left(\hat{\mathbf{H}}_{-j,t}^1\right),\ldots,F^K\left(\hat{\mathbf{H}}_{-j,t}^K\right)\right)} \tag{32}$$

where $\hat{\mathbf{H}}_{-j,t}^k = \left(\hat{H}_{1,t}^k,\ldots,H_{j,t}^k,\ldots\hat{H}_{J,t}^k\right)$ means that all majors are at their optimal allocation across occupations, except for major j remaining at their current allocation. Summing all five major groups' mismatch evolution yields the aggregate mismatch in Panel A of Figure 4. We show major-specific mismatch in partial and general equilibrium using Equation (11) over years in the companion Figure F4.

[21]For readability of the figure, we aggregate 11 ASCED 2-digit majors into 5 major groups: (1) STEM-related majors include Natural and Physical Sciences, Information Technology, Engineering and Related Technologies, Architecture and Building, and Agriculture, Environmental and Related Studies; (2) Health and Education; (3)

flat or slightly decreasing trend in output loss until 2012, which increases (in some cases sharply) thereafter. Management and Commerce, and STEM majors exhibit the largest mismatch pre-2012. Management and Commerce major also has the largest increase over time post-2012. STEM, and Society, Culture and Arts majors show a small decrease, which is again undone by a subsequent increase until 2019. Similarly, output loss due to the misallocation of Health and Education majors also increases markedly post-2012. Overall, aggregated mismatch in general equilibrium is mainly driven by Management and Commerce majors (3.5 percent), Society, Culture, and Arts majors (2.5 percent), and STEM majors (2.4 percent). While Health and Education has an average output loss of less than half (1.2 percent).

6.3.1 Decomposing the Rise in Mismatch

There are two potential drivers of the rise in mismatch post-2012. A change in the relative supply of college graduates across majors (see Figure F5 for the number of new college graduate workers by major over time) or a change in the efficiency of matching of workers to top occupations post-2012 (see Figure F6 for the life-cycle of mismatch across different cohorts).[22]

We show through two counterfactuals, in the context of our general equilibrium framework, how to quantify a change in (i) relative supplies, and (ii) sorting. To do so we compare each year-bin post-2012 with a base period, the average of 2003-2009.

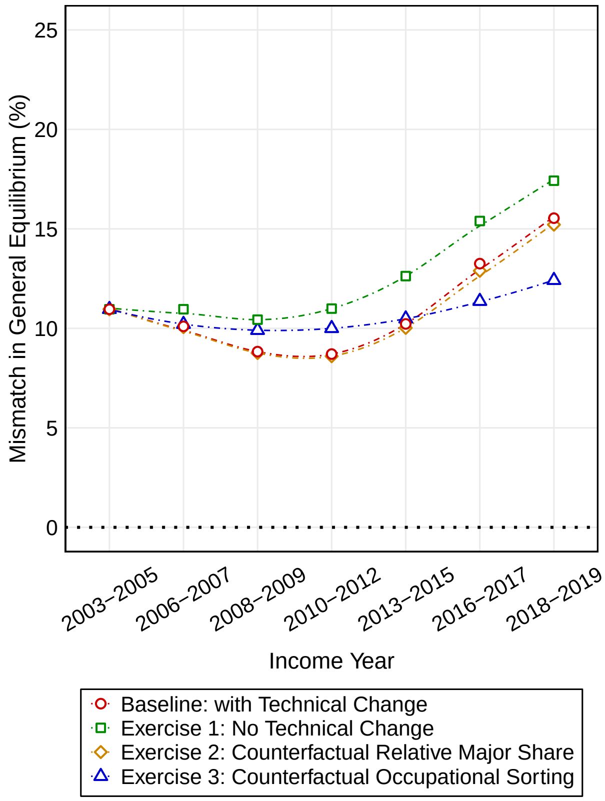
For the first counterfactual, given the production function is homogeneous of degree one, we quantify mismatch for a counterfactual labor supply, that has the same relative major shares as the actual labor supply in the base period, but sorts according to labor market conditions post-2012. The optimal reallocation of counterfactual talent $\hat{H}c^{k'}_{j,t}$ at time t satisfies Equations (14) and (15) and the counterfactual mismatch is computed through Equation (19). This counterfactual mismatch (see Figure 5 yellow-diamond line) shows how mismatch falls once relative supplies across majors are kept constant.

The second counterfactual quantifies mismatch for a counterfactual labor supply, that has the same relative occupation shares conditional on major as the actual labor supply in the base period, but graduations shares from the current time period. Thus, this counterfactual

Management and Commerce (as original); (4) Society, Culture, and Arts includes Society and Culture, and Creative Arts; and (4) Food, Hospitality, and Personal Services (as original).

[22]For context, 2009 marked a major change to the subsidies of the Australian tertiary education system as the federal government decided to uncap the Commonwealth Supported Places (CSPs). CSPs are university places that are partly funded by the government. Until 2009, the government set the number of CSPs by discipline available each year. The uncapping was done gradually from 2010 onward, but by 2012 all domestic students had the ability to access a subsidized university spot if they were admitted to a university. Previous studies have suggested that this policy resulted in an increase of 31.5 percent in CSPs places from 2009 to 2016 (Department of Education and Training, 2017). This subsequent increase in admissions was uneven across majors and Universities and coincided with the Financial Crisis.

Figure 5: Counterfactual Analysis

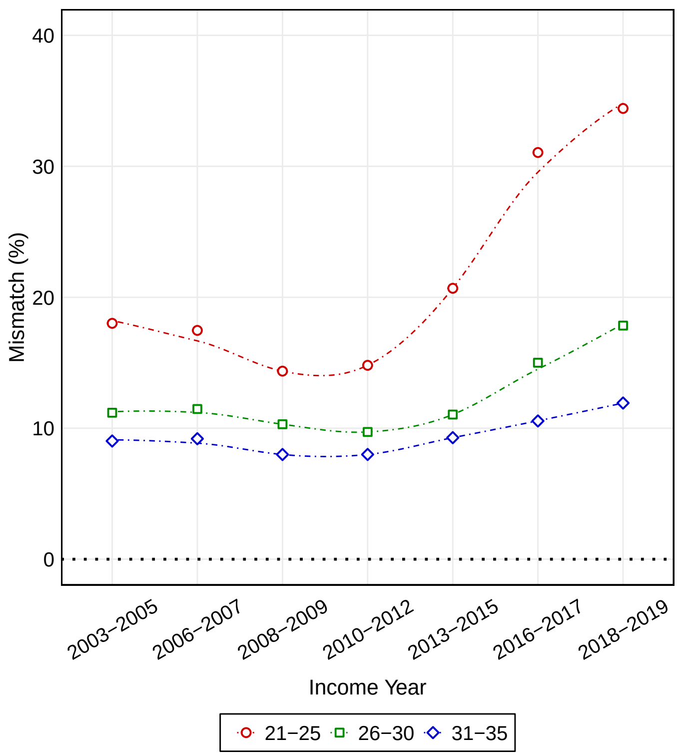


Notes: The figures shows 3 counterfactual, to study the rise in the evolution of skill mismatch in general equilibrium. The red circular line plots the baseline with technical change (changing α's) as in Figure 4. The green square lines shows the counterfactual when fixing α's to the base year, 2003-2005. The yellow diamond line fixed relative major supplies to 2003-2005 and, finally, the blue triangle line fixes the sorting of workers to occupations by major to the 2003-2005 base year.

mismatch (see Figure 5 blue triangle line) shows how mismatch falls in the post-2012 period once new cohorts sort into top occupations at the same rate and timing as cohorts in the base period.

Virtually all change since 2003 can be attributed to sorting. To understand if sorting is driven by a change in technology and affects all workers equally we consider two further exercises. First, we consider if technical change, through changes in α's can explain some of the evolution in skill mismatch. However, if α's are fixed at 2003-2005 values and shutting down any time trend in technology leads to an even larger increase in skill mismatch (see Figure 5 green square line). If anything this suggests that workers have responded to technical change in matching with a different occupation set over time. Secondly, we consider if all workers' skill matches have deteriorated equally, to do so we decompose the evolution of skill mismatch by three age groups – 21-25, 26-30, and 31-35-year-olds. As is evident from Figure 6, for workers aged 31-35 skill mismatch is flat during the entire period. Almost all of the increase in skill

Figure 6: Decomposition by Age Group



Notes: The figure plots the decomposition of mismatch in general equilibrium by age groups of 21-25, 26-30, and 31-35 over years. Specifically, we compute the mismatch measure by using equation (20) for each age group and year-bin.

mismatch is driven by the youngest cohort (workers aged 21-25) consistent with our life-cycle results.

7 Conclusion

This paper proposes a framework to study skill mismatch consistently at the individual level and in general equilibrium. We apply the framework to study mismatch across college major-occupation combinations. At the individual level, we measure mismatch through wage loss at fixed skill returns, in the general equilibrium, we measure mismatch through output loss. We estimate our mismatch measure with Australian administrative tax panel data using employment history and university degree information. We find that Commerce, Social Sciences, Arts, and STEM-related fields are the main drivers of mismatch in the aggregate. At the individual level, only Commerce, Social Sciences and Arts majors show that mismatch imposes costly and persistent scarring effects, as STEM majors seem to be able to leverage multiple skills across occupations in the labor market. Our results indicate that skill mismatch is overestimated when ignoring general equilibrium effects. While on average there are significant individual gains by reallocating workers to occupations best suited for their majors, the effect completely disappears as individuals age and occupation-major returns are allowed to adjust accordingly. In summary, our results highlight that the estimated general equilibrium mismatch is substantially smaller than found in previous studies. Our findings suggest that identifying the causes of mismatch and informing policy decisions is an important direction for future research.

References

Altonji, J. G., P. Arcidiacono, and A. Maurel (2016). The analysis of field choice in college and graduate school: Determinants and wage effects. In *Handbook of the Economics of Education*, Volume 5, pp. 305–396. Elsevier.

Altonji, J. G., E. Blom, and C. Meghir (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annu. Rev. Econ. 4*(1), 185–223.

Altonji, J. G., L. B. Kahn, and J. D. Speer (2014). Trends in earnings differentials across college majors and the changing task composition of jobs. *American Economic Review: Papers & Proceedings 104*(5), 387–93.

Altonji, J. G., L. B. Kahn, and J. D. Speer (2016). Cashier or consultant? entry labor market conditions, field of study, and career success. *Journal of Labor Economics 34*(S1), S361–S401.

Andrews, D., N. Deutscher, J. Hambur, and D. Hansell (2020). The career effects of labour market conditions at entry.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of econometrics 121*(1-2), 343–375.

Atalay, E., P. Phongthiengtham, S. Sotelo, and D. Tannenbaum (2020). The evolution of work in the united states. *American Economic Journal: Applied Economics 12*(2), 1–34.

Autor, D. H. and M. J. Handel (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of labor Economics 31*(S1), S59–S96.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics 118*(4), 1279–1333.

Bacolod, M. P. and B. S. Blum (2010). Two sides of the same coin us "residual" inequality and the gender gap. *Journal of Human resources 45*(1), 197–242.

Burstein, A., E. Morales, and J. Vogel (2019). Changes in between-group inequality: Computers, occupations, and international trade. *American Economic Journal: Macroeconomics 11*(2), 348–400.

Carter, A., T. Abhayaratna, and S. Johnson (2021). The ato longitudinal information files (alife): Individuals-a new dataset for public policy research. *Tax and Transfer Policy Institute, working paper 13*.

Caunedo, J., D. Jaume, and E. Keller (2021). Occupational exposure to capital-embodied technical change. *CEPR Discussion Paper*. No. DP15759.

Dahl, G. B. (2002). Mobility and the return to education: Testing a roy model with multiple markets. *Econometrica 70*(6), 2367–2420.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics 132*(4), 1593–1640.

Deming, D. J. and K. Noray (2020). Earnings dynamics, changing job skills, and stem careers. *The Quarterly Journal of Economics 135*(4), 1965–2005.

Department of Education and Training (2017). The higher education reform package.

Eckardt, D. (2019). Are chemists good bankers? returns to the match between training and occupation. *mimeo*.

Edmond, C. and S. Mongey (2021). Unbundling labor. *Working Paper*.

Fredriksson, P., L. Hensvik, and O. N. Skans (2018). Mismatch of talent: Evidence on match quality, entry wages, and job mobility. *American Economic Review 108*(11), 3303–38.

French, E., C. Taber, et al. (2011). Identification of models of the labor market. *Handbook of Labor Economics 4*, 537–617.

Gathmann, C. and U. Schönberg (2010). How general is human capital? a task-based approach. *Journal of Labor Economics 28*(1), 1–49.

Groes, F., P. Kircher, and I. Manovskii (2015). The u-shapes of occupational mobility. *The Review of Economic Studies 82*(2), 659–692.

Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2020). Multidimensional skill mismatch. *American Economic Journal: Macroeconomics 12*(1), 210–44.

Hathorne, C. and R. V. Breunig (2022). Occupational mobility in the alife data: how reliable are occupational patterns from administrative australian tax records?

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019). The allocation of talent and us economic growth. *Econometrica 87*(5), 1439–1474.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics 124*(4), 1403–1448.

Ingram, B. F. and G. R. Neumann (2006). The returns to skill. *Labour economics 13*(1), 35–59.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963–1987: Supply and demand factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Kinsler, J. and R. Pavan (2015). The specificity of general human capital: Evidence from college major choice. *Journal of Labor Economics 33*(4), 933–972.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics 131*(3), 1057–1111.

Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica: Journal of the Econometric Society*, 507–512.

Lemieux, T. (2014). Occupations, fields of study and returns to education. *Canadian Journal of Economics/Revue canadienne d'économique 47*(4), 1047–1077.

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

Liu, K., K. G. Salvanes, and E. Ø. Sørensen (2016). Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review 84*, 3–17.

McGowan, M. A. and D. Andrews (2015). Labour market mismatch and labour productivity: Evidence from piaac data. *OECD Economic Department Working Papers* (1209), 0_1.

Patterson, C., A. Şahin, G. Topa, and G. L. Violante (2016). Working hard in the wrong place: A mismatch-based explanation to the uk productivity puzzle. *European Economic Review 84*, 42–56.

Pellizzari, M. and A. Fichen (2013). A new measure of skills mismatch: Theory and evidence from the survey of adult skills (piaac). *OECD Social, Employment, and Migration Working Papers* (153), 0_1.

Poletaev, M. and C. Robinson (2008). Human capital specificity: evidence from the dictionary of occupational titles and displaced worker surveys, 1984–2000. *Journal of Labor Economics 26*(3), 387–420.

Rathelot, R. v. R. et al. (2017). Rethinking the skills gap. *IZA World of Labor*.

Rendall, M. (2010). Brain versus brawn: the realization of women's comparative advantage. *Working Paper*.

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

Şahin, A., J. Song, G. Topa, and G. L. Violante (2014). Mismatch unemployment. *American Economic Review 104*(11), 3529–3564.

Yamaguchi, S. (2012). Tasks and heterogeneous human capital. *Journal of Labor Economics 30*(1), 1–53.

Appendix

A Data

A.1 HECS/HELP Data

We use Higher Education Loan Program (HELP) data to obtain an individual's post-secondary educational information. This data provides us with the worker's college course name—i.e., Bachelor of Engineering—which suggests both degree level and field of study; institutional types such as university, Technical and Further Education (TAFE)/College/Academy, or school/others; and yearly student loan debt. We design a keyword extraction algorithm to classify thousands of different course names recorded in HELP into the 4-digit level of the Australian Standard Classification of Education (ASCED) as well as into different degree levels.[23] Since degree completion indicators are not available in the administrative data, we follow the approach by Andrews, Deutscher, Hambur, and Hansell (2020) and impute the completion indicator by observing when students stop incurring new HECS/HELP debt for each degree level. As such, the graduation cohort is defined as the last year when the worker incurs the new debt.

A.2 Australian Longitudinal File on Individuals (ALife)

ALife tax record provides us with annual pre-tax salary and wage, occupational track record (classified by ANZSCO), and place of residence for 10% Australian population. We aggregate workers' occupations at ANZSCO 2-digit level. We construct an individual's work experience based on his or her years of full-time employment, which is calculated as the number of years between the first year of observed pay and the current year, provided that the annual salary is higher than the annual minimum wage.[24] We inflate wages to 2019 Australian dollars.

A.3 Estimation Sample and Restriction

We link workers' labor market outcomes from ALife to their college major and restrict observations to the period after the year of graduation. To ensure that we have sufficient samples within each major-occupation pair, we use 2-digit levels of ASCED and ANZSCO to classify

[23] Specifically, degree level is aggregated based on ASCED into three broad groups: (1) non-award degree; (2) certificate/diploma and bachelor; and (3) graduate certificate/diploma and postgraduate degree.

[24] We compute annual minimum wage as minimum annual wage = minimum wage rate/2 $\times$ 35hours $\times$ 48weeks.

majors and occupations, respectively. We focus on the undergraduate sample, which is limited to individuals between the ages of 21-35 years old who held at least one type of academic degree, such as a diploma, certificate, or bachelor's degree, regardless of institution type. To ensure that wage reflects the skill prices and to alleviate the concern of the impact of career promotion on wage, we remove observations that are self-employed or work in managerial or government servant positions. It is not uncommon to find a worker may have studied two different majors, and/or obtained more than one degree level throughout their higher education history. To ensure that occupations are matched to only one field of study and degree level, we restrict the sample to individuals with only one major at the 2-digit classification and the most recent degree level.

B Details on Estimation Procedure

B.1 Correct Selection Bias

Individuals are observed in an j,k cell if and only if the Rule (23), conditional on major j, is satisfied,[25]

$$M_{i,(k|j),c,r,t} = 1, \text{ iff } e_{i,(k|j),c,r,t} - e_{i,(k'|j),c,r,t} \leqslant \tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(k'|j),c,r,t};\ \forall k' \neq k \tag{B.1}$$

Directly using (B.1) to pin down the selectivity bias is infeasible because we would need a complete specification of the joint distribution of error terms in the wage equation and error terms in K selection equations for each major j. To address the estimation challenge, we follow Lee (1983) to reduce of the problem by only looking at the maximum order statistic. Equivalently, (B.1) can be rewritten as:

$$M_{i,(k|j),c,r,t} = 1, \text{ iff } \max_{k'}(e_{i,(k'|j),c,r,t} - e_{i,(k|j),c,r,t} - (\tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(k'|j),c,r,t})) \leqslant 0 \tag{B.2}$$

Given this, the observation in cell (j,k) is observed ($M_{i,(k|j),c,r,t} = 1$) if the maximum order statistics is less than 0. To get the probability of this event, let $F^e_{k|j}$ be the joint CDF of selection error terms $(e_{i,(1|j),c,r,t} - e_{i,(k|j),c,r,t}, \ldots, e_{i,(K|j),c,r,t} - e_{i,(k|j),c,r,t})$. We can derive the CDF of the maximum

[25] Note, we allow for regional variation on selection probabilities giving us additional variation in the data.

order statistics, conditional on sub-utility differences, as the following:

$$\begin{aligned}
&H_{k|j}(t \mid \tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(1|j),c,r,t}, ..., \tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(K|j),c,r,t}) \\
=&Pr(\max_{k'}(e_{i,(k'|j),c,r,t} - e_{i,(k|j),c,r,t} - (\tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(k'|j),c,r,t})) \leqslant t \mid \\
&\tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(1|j),c,r,t}, ..., \tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(K|j),c,r,t}) \\
=&Pr(e_{i,(1|j),c,r,t} - e_{i,(k|j),c,r,t} \leqslant \tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(1|j),c,r,t} + t, ..., \\
&e_{i,(K|j),c,r,t} - e_{i,(k|j),c,r,t} \leqslant \tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(K|j),c,r,t} + t) \\
=&F^e_{j,k}(\tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(1|j),c,r,t} + t, ..., \tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(K|j),c,r,t} + t) \qquad \text{(B.3)}
\end{aligned}$$

If $t = 0$, then

$$\begin{aligned}
&Pr(\max_{k'}(e_{i,(k'|j),c,r,t} - e_{i,(k|j),c,r,t} - (\tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(k'|j),c,r,t})) \leqslant 0 \\
&|\tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(1|j),c,r,t}, ..., \tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(K|j),c,r,t}) \\
=&F^e_{j,k}(\tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(1|j),c,r,t}, ..., \tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(K|j),c,r,t}) \\
=&F^e_{j,k}(\vec{U}) \qquad \text{(B.4)}
\end{aligned}$$

where the vector of sub-utility difference is $\vec{U} = (\tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(1|j),c,r,t}, ..., \tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(K|j),c,r,t})$.

To this end, above Equation (B.4) gives the probability that the maximum order statistics $\max_{k'}(e_{i,(k'|j),c,r,t} - e_{i,(k|j),c,r,t} - (\tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(k'|j),c,r,t}))$ is less than 0. That is the probability of observing an individual in occupation k conditional on major j. However, these random variables indexed over i are not identically distributed because the sub-utility differences might be systematically different.

Following Lee (1983), we propose a new random variable that can be constructed using the transformation,

$$v^*_{i,j,c,k,r,t} = \Gamma^{-1}_{j,k}\{F^e_{j,k}(\vec{U})\}, \qquad \text{(B.5)}$$

where $\Gamma^{-1}_{j,k}$ can be any univariate continuous strictly increasing function, so that $\Gamma_{j,k}$ can be defined as a marginal distribution of the new random variable $v_{i,j,k,c,r,t}$.[26] Therefore, the probability of a random variable $v_{i,j,k,c,r,t}$ that is less or equal to the value $v^*_{i,j,k,c,r,t}$ can well summarize the probability of the event that the maximum order statistics is less than 0, and thereby $M_{i,(k|j),c,r,t} = 1$. To see this,

$$Pr(v_{i,j,k,c,r,t} \leqslant v^*_{i,j,k,c,r,t}) = \Gamma_{j,k}(\Gamma^{-1}_{j,k}\{F^e_{j,k}(\vec{U})\}) = F^e_{j,k}(\vec{U}) \qquad \text{(B.6)}$$

[26]Lee's implicit assumption for the above transformation is that the same transformation is applied regardless of the specific values of sub-utility differences.

Hence, using (B.5), we can rewrite (B.2) as the following:

$$M_{i,(k|j),c,r,t} = 1, \text{ iff } \upsilon_{i,j,k,c,r,t} \leqslant \Gamma^{-1}_{j,k}\{F^{e}_{j,k}(\vec{U})\} \tag{B.7}$$

Following Lee (1983), we assume that the error in wage equation $\epsilon_{i,j,k,c,r,t}$ and this new transformed random variable $\upsilon_{i,j,k,c,r,t}$ are independently and identically distributed with a joint distribution function $G_{j,k}(...)$. With this assumption, we can use the two-stage Heckman estimation for correcting selectivity bias. Specifically, we assume that: (1) $\Gamma_{j,k}$ is a univariate standard normal CDF (Φ); (2) $G_{j,k}(...)$ is a bivariate standard normal CDF.

To this end, we can pin down the self-selection bias $E[\epsilon_{i,j,k,c,r,t}|M_{i,(k|j),c,r,t} = 1]$ based on the above distributional assumption in the following way:

$$\begin{aligned}
&E[\epsilon_{i,j,k,c,r,t}|M_{i,(k|j),c,r,t} = 1] \\
&\quad = E[\epsilon_{i,j,k,c,r,t}|\upsilon_{i,j,k,c,r,t} \leqslant \Phi^{-1}_{j,k}\{F^{e}_{j,k}(\vec{U})\}] \\
&\quad = \frac{Cov(\epsilon_{i,j,k,c,r,t}, \upsilon_{i,j,k,c,r,t})}{var(\upsilon_{i,j,k,c,r,t})} E(\upsilon_{i,j,k,c,r,t}|\upsilon_{i,j,k,c,r,t} \leqslant \Phi^{-1}_{jk}\{F^{e}_{j,k}(\vec{U})\}) \\
&\quad = \frac{Cov(\epsilon_{i,j,k,c,r,t}, \upsilon_{i,j,k,c,r,t})}{\sqrt{var(\epsilon_{i,j,k,c,r,t})var(\upsilon_{i,j,k,c,r,t})}} \frac{\sqrt{var(\epsilon_{i,j,k,c,r,t})}}{\sqrt{var(\upsilon_{i,j,k,c,r,t})}} E(\upsilon_{i,j,k,c,r,t}|\upsilon_{i,j,k,c,r,t} \leqslant \Phi^{-1}_{j,k}\{F^{e}_{j,k}(\vec{U})\}) \\
&\quad = \rho_{j,k} \times 1 \times E(\upsilon_{i,j,k,c,r,t}|\upsilon_{i,j,k,c,r,t} \leqslant \Phi^{-1}_{j,k}\{F^{e}_{j,k}(\vec{U})\}) \\
&\quad = \rho_{j,k} \frac{-\phi(\Phi^{-1}(F^{e}_{j,k}(\vec{U})))}{\Phi(\Phi^{-1}(F^{e}_{j,k}(\vec{U})))} \\
&\quad = \rho_{j,k} \frac{-\phi(\Phi^{-1}(F^{e}_{j,k}(\vec{U})))}{F^{e}_{j,k}(\vec{U})} \\
&\quad = \rho_{j,k} \lambda_{j,k}(F^{e}_{j,k}(\vec{U}))
\end{aligned} \tag{B.8}$$

where $\vec{U} = (\tilde{U}_{i,(k|j)c,r,t} - \tilde{U}_{i,(1|j),c,r,t}, ..., \tilde{U}_{i,(k|j),c,r,t} - \tilde{U}_{i,(k'|j),c,r,t})$, $\rho_{j,k}$ is the correlation between $\epsilon_{i,j,k,c,r,t}$ and $\upsilon_{i,j,k,c,r,t}$, and $\lambda_{j,k}(\cdot)$ is the parametric form of correction function.[27] Following Dahl (2002) with the index sufficiency assumption, we can show $F^{e}_{j,k}(\vec{U}) = p_{i,j,k,c,r,t}$.

B.2 Estimation of Selection Probability

To estimate occupational selection probability, we follow Eckardt (2019) and adopt a flexible machine learning approach—random forest—to avoid functional form assumption of the un-

[27] We assume the correlation only differs among j, k combinations.

derlying subutility function. The advantage of the random forest lies in its flexibility in incorporating a large number of predictors into the model and its prediction accuracy. Meanwhile, it avoids the usual drawbacks of the conditional logit model that suffers from the independence of irrelevant alternatives property.

We predict occupational choice conditional on each major by using a random forest algorithm. The explanatory variables for predicting occupational choice include graduation cohort dummies, age, gender, working experience, balance of debt, and a set of Bartik shift-share instruments capturing exogenous change in occupational demand in each year and region. Specifically, Bartik instrument $B_{k'|r,t}$ for each occupation k is constructed as:

$$B_{k'|r,t} = \underbrace{z_{k'|r,(t=0)}}_{\substack{\text{Share: Initial Occ. } k' \text{ share} \\ \text{in region } r}} \times \underbrace{m_{k',t}}_{\substack{\text{Shift: Growth of Occ. } k' \\ \text{at the national level}}} \tag{B.9}$$

where $z_{k'|r,(t=0)}$ is the initial share of occupation k' in region r at the base year $t = 0$ (i.e., year 1994—the first year when occupation code is observed), and $m_{k',t}$ is the yearly growth rate of occupation k' share in year t at the national level. Both initial occupation shares in the local region and yearly national-level shift in occupation shares are assumed exogenous.[28] That is, $B_{k'|r,t}$ approximates the exogenous change in the share of occupation k in region r in year t.[29]. As the exclusion restriction should affect the probability of choosing an occupation rather than directly affecting wages, we use a one-year lagged Bartik instrument, $B_{k'|r,t-1}$, as a predictor, to mitigate the issue.

[To be complete]

C Estimation of Production Function Parameters

For baseline estimation of production function parameters, we assume no technical change and use the fixed $\{\alpha_j^k\}$ and $\{\alpha_k\}$ across time. For occupational-specific production function, estimating parameters $\{\alpha_j^k\}$ and $\{\sigma_k\}$ in specification (27) needs to exploit the region and time variation in $\ln(p^k_{j',r,t}/p^k_{j,r,t})$ and $\ln(H^k_{j',r,t}/H^k_{j,r,t})$. In order to obtain sufficient variation and to address the sampling limitations, we group 17 years into two year-bins (hereafter denoted as b) and obtain the price of major-occupation combination $p_{j',r,b}$ and efficiency labor units $H_{j',r,\mathcal{T}}$

[28]Since regional occupation shares in the initial period (1994) are a decade ahead relative to the beginning of the estimation period (2003) we consider this a reasonable assumption.

[29]To precisely gauge the growth of occupational share in the economy, we only use the occupation information in the tax data, without imposing any sample restriction.

by estimating the specification (25) conditional on each region r and year-bin $\mathcal{T}$.[30] We then estimate the following specification for each occupation,

$$\underbrace{\ln\left(p^k_{j',r,\mathcal{T}}/p^k_{j,r,\mathcal{T}}\right)}_{\beta^k_{j',r,\mathcal{T}}-\beta^k_{j,r,\mathcal{T}}} = \ln\left(\alpha^k_{j'}/\alpha^k_j\right) + \left(\sigma^k - 1\right)\ln\left(H^k_{j',r,\mathcal{T}}/H^k_{j,r,\mathcal{T}}\right); \quad \text{for each } k \in K, \tag{C.1}$$

where $p^k_{j',r,\mathcal{T}}$ is the return to major-occupation combination conditional on the region r and year-bin $\mathcal{T}$; and $H^k_{j',r,\mathcal{T}} = \sum_{i \in I^k_{j',r,\mathcal{T}}} \tilde{h}_i$ is total efficiency labor units. To this end, for each occupation k, we obtain the estimates of σ_k and $\{\alpha^k_j\}$ with the constraint of $\sum_{j \in J} \alpha^k_j = 1$.

For the aggregate production function, estimating $\{\alpha_k\}$ and σ from Equation (31) needs variation in $p_{k',r,t}$ and $Y_{k',r,t}$. We continue to exploit cross-regions and year-bin variation. By Equation (3), we obtain $Y^k_{r,\mathcal{T}}$ using estimated $p^k_{j',r,\mathcal{T}}$ and $H^k_{j',r,\mathcal{T}}$. Equation (31) can then be estimated at the level of region and year-bin,

$$\ln\left(\frac{\left(\sum_j \left(\alpha^{k'}_j\right)^{-\frac{1}{\sigma_{k'}-1}} \left(p^{k'}_{j,r,\mathcal{T}}\right)^{\frac{\sigma_{k'}}{\sigma_{k'}-1}}\right)^{(1-\frac{1}{\sigma_{k'}})}}{\left(\sum_j \left(\alpha^{k}_j\right)^{-\frac{1}{\sigma_{k}-1}} \left(p^{k}_{j,r,\mathcal{T}}\right)^{\frac{\sigma_{k}}{\sigma_{k}-1}}\right)^{(1-\frac{1}{\sigma_{k}})}}\right) = \ln\left(\frac{\alpha^{k'}}{\alpha^k}\right) + (\sigma - 1)\ln\left(\frac{Y^{k'}_{r,\mathcal{T}}}{Y^k_{r,\mathcal{T}}}\right). \tag{C.2}$$

Note technology parameters, $\ln\left(\frac{\alpha_{k',t}}{\alpha_{k,t}}\right) + \ln\left(\frac{A_{k',t}}{A_{k,t}}\right)$ are not separately identifiable, but this is not crucial for our exercise. A typical assumption in the literature (see Katz and Murphy (1992)) is to assume a linear time trend in technological change, e.g., the expression can be approximated by:

$$\ln\left(\frac{\alpha_{k',t}}{\alpha_{k,t}}\right) + \ln\left(\frac{A_{k,t}}{A_{k',t}}\right) = \gamma_{0,kk'} + \gamma_{1,kk'}t$$

when using time-series data or simply $\gamma_{0,kk'}$ if using only cross-sectional variation.

To this end, we obtain the estimates σ and $\{\alpha^k\}$ by the constraint of $\sum_{k \in K} \alpha^k = 1$. Therefore, all production function parameters $\{\alpha^k_j\}$, $\{\sigma_k\}$, $\{\alpha^k\}$, and σ are estimated. Table D1 provides the estimated production function parameters.

[30] We aggregate the years 2003-2019 into 7 two-year-bins, 2003-2005, 2006-2007, 2008-2009, 2010-2012, 2013-2015, 2016-2017, and 2018-2019. To obtain precise estimates of the price of major-occupation combinations, we restrict the number of observations within the cell of major-occupation-year-bin-and-region to at least five and restrict the number of majors with occupation to be at least two. In total, this leaves us 233 major-occupation combinations, the same number as the original estimation sample without restriction, while the caveat is that it loses 44 percent of the number of cells of major-occupation-region-and-year-bin.

Table D1: Estimates of Production Function Parameters

ANZSCO 2-digit	Occupation Title	$alpha_{k}$	Elas	Major Weights Within Occupation									
				1	2	3	4	5	6	7	8	9	10
21	Arts and Media Professionals	0.027	1.753	0	0	0	0	0	0.208	0.182	0.193	0.183	0.235
22	Business, Human Resource and Marketing Professionals	0.094	2.25	0.076	0.084	0.095	0.057	0.07	0.074	0.066	0.324	0.092	0.063
23	Design, Engineering, Science and Transport Professionals	0.084	1.983	0.114	0.067	0.279	0.093	0.109	0.066	0.049	0.073	0.072	0.077
24	Education Professionals	0.046	1.796	0.061	0.045	0.051	0	0.052	0.088	0.517	0.081	0.078	0.028
25	Health Professionals	0.04	1.758	0.055	0	0.039	0	0.044	0.684	0.034	0.057	0.048	0.038
26	ICT Professionals	0.067	1.984	0.072	0.239	0.162	0	0.094	0.072	0.065	0.159	0.068	0.068
27	Legal, Social and Welfare Professionals	0.034	1.877	0.079	0	0.104	0	0	0.086	0.058	0.152	0.485	0.037
31	Engineering, ICT and Science Technicians	0.044	1.878	0.115	0.107	0.125	0.123	0.089	0.086	0.065	0.111	0.095	0.083
34	Electrotechnology and Telecommunications Trades Workers	0.035	1.856	0.085	0.119	0.149	0.076	0.081	0.087	0.081	0.126	0.099	0.096
36	Skilled Animal and Horticultural Workers	0.021	1.612	0.12	0	0.083	0.102	0.156	0.118	0.097	0.131	0.107	0.085
3X	Automotive, Engineering, and Construction Trades Workers	0.03	1.768	0.08	0.096	0.152	0.111	0.088	0.091	0.096	0.132	0.1	0.054
3Y	Food Trades, Other Technicians, and Trades Workers	0.026	1.69	0.094	0.084	0.133	0.099	0.12	0.113	0.093	0.135	0.096	0.034
41	Health and Welfare Support Workers	0.019	2.303	0.068	0	0	0	0	0.283	0.143	0.167	0.223	0.115
42	Carers and Aides	0.031	1.857	0.084	0.096	0.086	0	0.082	0.151	0.174	0.123	0.155	0.049
43	Hospitality Workers	0.025	1.93	0.094	0.082	0.093	0.086	0.069	0.097	0.09	0.149	0.111	0.129
45	Sports and Personal Service Workers	0.025	1.727	0.111	0	0.12	0	0.104	0.148	0.136	0.181	0.163	0.038
52	Personal Assistants and Secretaries	0.02	1.779	0.096	0	0	0	0.117	0.121	0.125	0.261	0.176	0.103
53	General Clerical Workers	0.036	1.648	0.086	0.076	0.092	0.084	0.076	0.113	0.116	0.198	0.126	0.033
54	Inquiry Clerks and Receptionists	0.025	1.997	0.108	0.101	0.118	0	0.084	0.114	0.111	0.184	0.133	0.047
55	Numerical Clerks	0.045	1.787	0.066	0.092	0.091	0.075	0.078	0.077	0.096	0.194	0.112	0.119
56	Clerical and Office Support Workers	0.021	1.666	0.135	0.1	0.096	0	0	0.122	0.111	0.185	0.166	0.085
59	Other Clerical and Administrative Workers	0.035	1.711	0.088	0.081	0.084	0.068	0.088	0.079	0.083	0.201	0.142	0.087
61	Sales Representatives and Agents	0.041	1.85	0.07	0.088	0.092	0.075	0.082	0.106	0.084	0.231	0.11	0.06
62	Sales Assistants and Salespersons	0.033	2.056	0.087	0.099	0.095	0.093	0.077	0.102	0.105	0.153	0.125	0.063
63	Sales Support Workers	0.03	1.789	0.083	0.108	0.096	0.116	0.095	0.081	0.101	0.148	0.089	0.082
7X	Machinery Operators and Drivers	0.031	1.826	0.1	0.094	0.121	0.089	0.105	0.098	0.094	0.151	0.114	0.034
8X	Labourers	0.037	1.949	0.091	0.082	0.128	0.098	0.09	0.096	0.083	0.153	0.114	0.065
	Elasticity of Substitution of Occupational Products		2.332										

Notes: This table shows the estimated production function parameters Columns 5-15 shows the major weights within each occupation, with the number 1-11 in the top panel representing for the 2-digit ASCED major.

D Robustness

D.1 Mismatch with Occupational Tenures 1

Consider aggregate human capital for a major j and an occupation k,

$$\hat{H}_j^k = \sum_{i \in I_j^k} \tilde{h}_{i,k} e^{\tau_i} = \sum_{i \in I_j^k} \tilde{h}_{i,k} \frac{\sum_{i \in I^k} \tilde{h}_{i,k} e^{\tau_i}}{\sum_{i \in I_j^k} \tilde{h}_{i,k}} = H_j^k T_j^k \tag{D.1}$$

where $T_j^k = \frac{\sum_{i \in I_j^k} \tilde{h}_{i,k} e^{\tau_i}}{\sum_{i \in I_j^k} \tilde{h}_{i,k}} = E_{\tau_j^k}\left[e^{\tau_j^k}\right]$ is the expected value of the exponential of tenure length. Rewrite the production function as

$$\begin{aligned} Y^k &= F^k(\hat{H}_1^k, ..., \hat{H}_j^k, ..., \hat{H}_J^k) \\ &= F^k(H_1^k T_j^k, ..., H_j^k T_j^k, ..., H_J^k T_j^k). \end{aligned} \tag{D.2}$$

Now, assume that the planner CANNOT choose a type of worker when reallocating so that T_j^k doesn't change (i.e. the distribution of tenures within occupation k doesn't change). Then, the reallocation problem becomes

$$Y^k = F^k\left(\left[H_1^k - R_1^k(-)\right] T_1^k + R_1^k(+), ..., \left[H_j^k - R_j^k(-)\right] T_j^k + R_1^k(+), ..., \left[H_J^k - R_J^k(-)\right] T_J^k + R_J^k(+)\right), \tag{D.3}$$

for all occupation k, where $R_j^k(-) \geq 0$ is the number of major-j workers reallocated out of occupation k and $R_j^k(+) \geq 0$ is that into occupation k, which satisfy the resource constraints,

$$\sum_k R_j^k(-) = \sum_k R_j^k(+), \tag{D.4}$$

for all major j. In addition, if $R_j^k(-) > 0$ holds, then $R_j^k(+) = 0$, or vice versa.

D.2 Mismatch with Occupational Tenures 2

Now assume that the planner CAN choose whom to reallocate. Let G_j^k, distribution of tenure length τ_j^k for major j in occupation k. In this situation, it is optimal for the planner to choose the workers with the lowest tenure first to reallocate, because

Then, the reallocation problem becomes

$$Y^k = F^k\left(\left[H_1^k - R_1^k(-)\right]\int_{G^{-1}\left(R_1^k(-)\right)}^{\infty} \tau_1^k dG_1^k + R_1^k(+), \ldots, \left[H_j^k - R_j^k(-)\right]\int_{G^{-1}\left(R_j^k(-)\right)}^{\infty} \tau_j^k dG_j^k + R_j^k(+), \ldots, \left[H_J^k - R_J^k(-)\right]\int_{G^{-1}\left(R_J^k(-)\right)}^{\infty} \tau_J^k dG_J^k + R_J^k(+)\right), \tag{D.5}$$

for all occupation k, where $R_j^k(-) \geq 0$ is the number of major-j workers reallocated out of occupation k and $R_j^k(+) \geq 0$ is that into occupation k, which satisfy the resource constraints,

$$\sum_k R_j^k(-) = \sum_k R_j^k(+), \tag{D.6}$$

for all major j.

D.3 Mismatch with Self-Selection Effects

From Equation (25), we could estimate the effect of selection on workers' wages. Let k' the worker's current occupation. Define

$$\tilde{h}'_{i,k} = \begin{cases} \tilde{h}_{i,k} \times \exp\left[\hat{\rho}_{j,k} \cdot g\right] & \text{if } k = k', \\ \tilde{h}_i & \text{otherwise.} \end{cases} \tag{D.7}$$

This is major-specific human capital that includes selection effects. Note that this human capital depends on an individual's occupational choice and thus has the subscript k.

We can now define a measure of major-occupation mismatch in a partial equilibrium with selection effects. Suppressing time subscripts, our mismatch measure for major j can be defined as:

$$m_j^{PE} = \frac{\sum_{i \in I^j} p_j^{k^*(i)} \tilde{h}'_i - \sum_{i \in I^j} p_j^k \tilde{h}'_{i,k}}{\sum_{i \in I} p_j^k \tilde{h}'_{i,k}}, \tag{D.8}$$

where

$$k^*(i) = \arg\max_k \left[p_j^k \tilde{h}'_{i,k} - p_j^{k'} \tilde{h}'_{i,k'}\right]. \tag{D.9}$$

Equation (D.9) defines individuals' top occupation taking into account the selection effect. could lose their selection effect, captured by $\tilde{h}'_{i,k}$. Having defined major specific mismatch measure, the aggregate economy-wide mismatch is now defined as,

$$\mathcal{M}^{PE} = \sum_{j \in J} m_j^{PE} = \frac{\sum_{i \in I} p_j^{k^*(i)} \tilde{h}'_i - \sum_{i \in I} p_j^k \tilde{h}'_i}{\sum_{i \in I} p_j^k \tilde{h}'_i}. \tag{D.10}$$

E Figure Appendix

Figure F1: Major-Occupation Mismatch Estimates

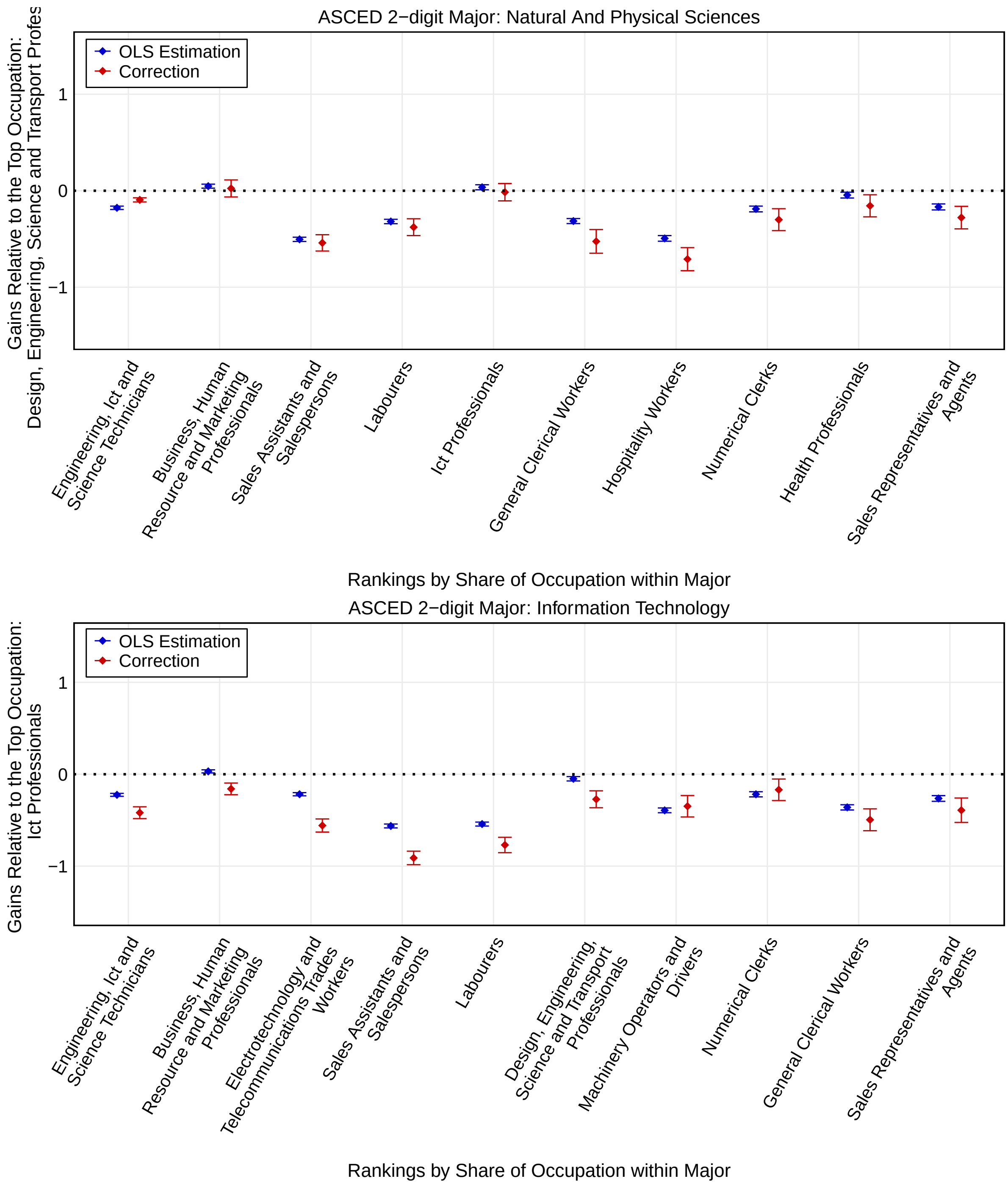
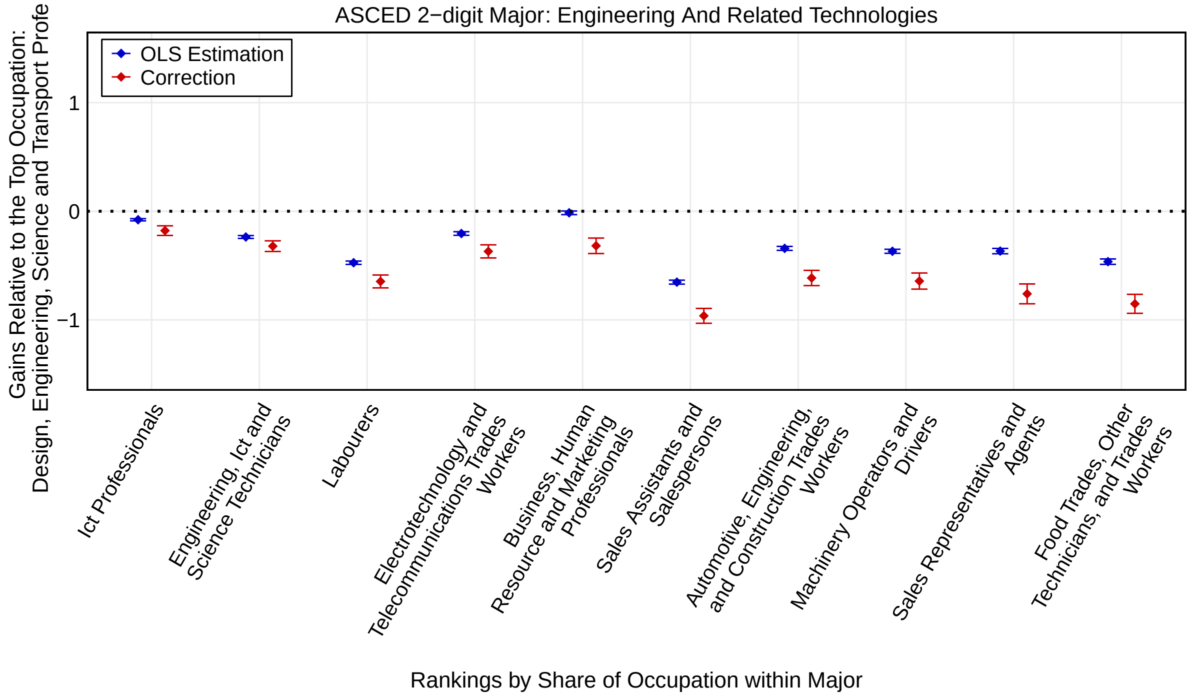


Figure Continued



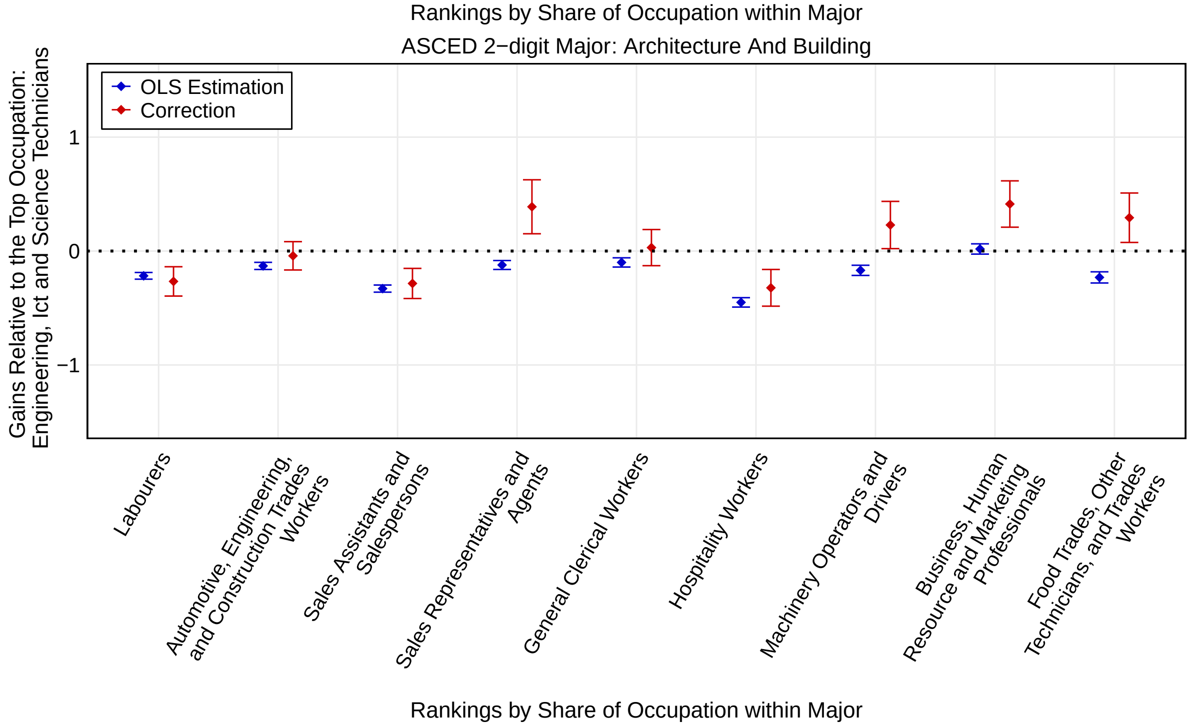
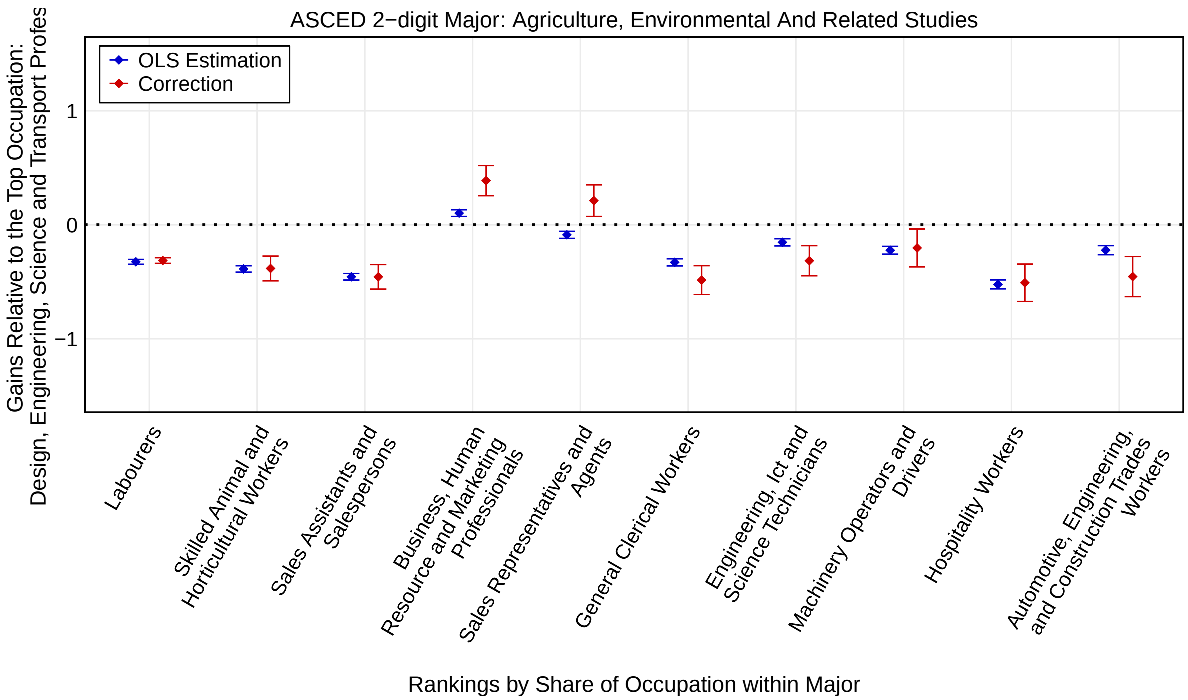


Figure Continued



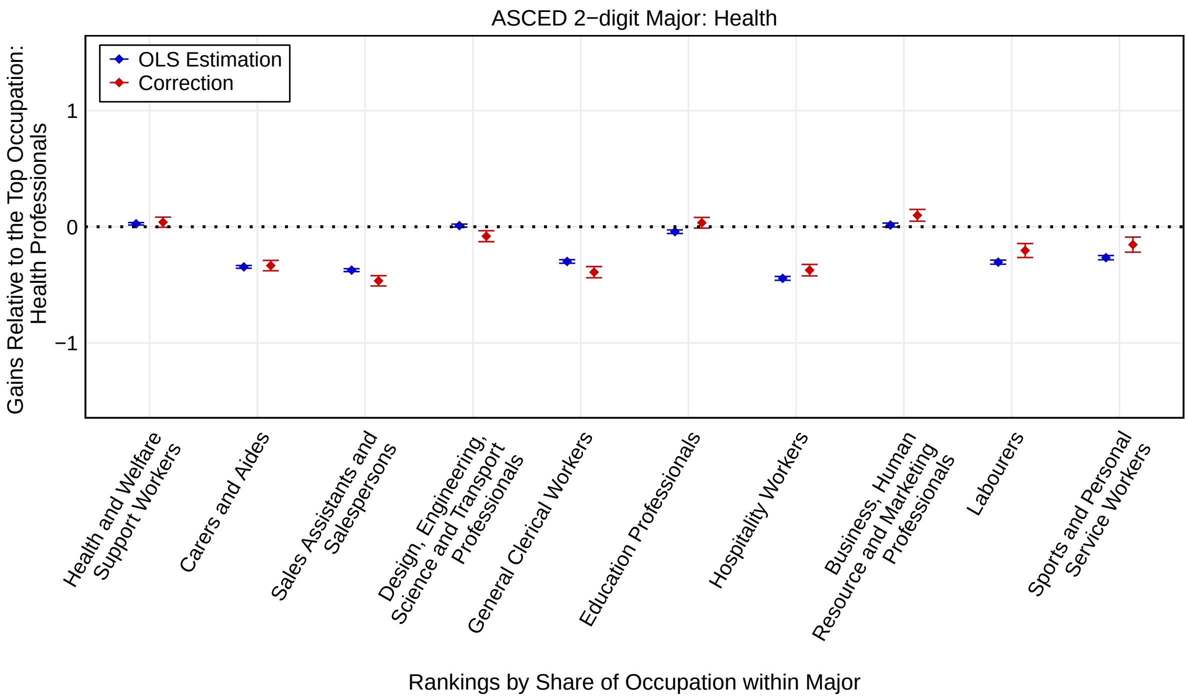
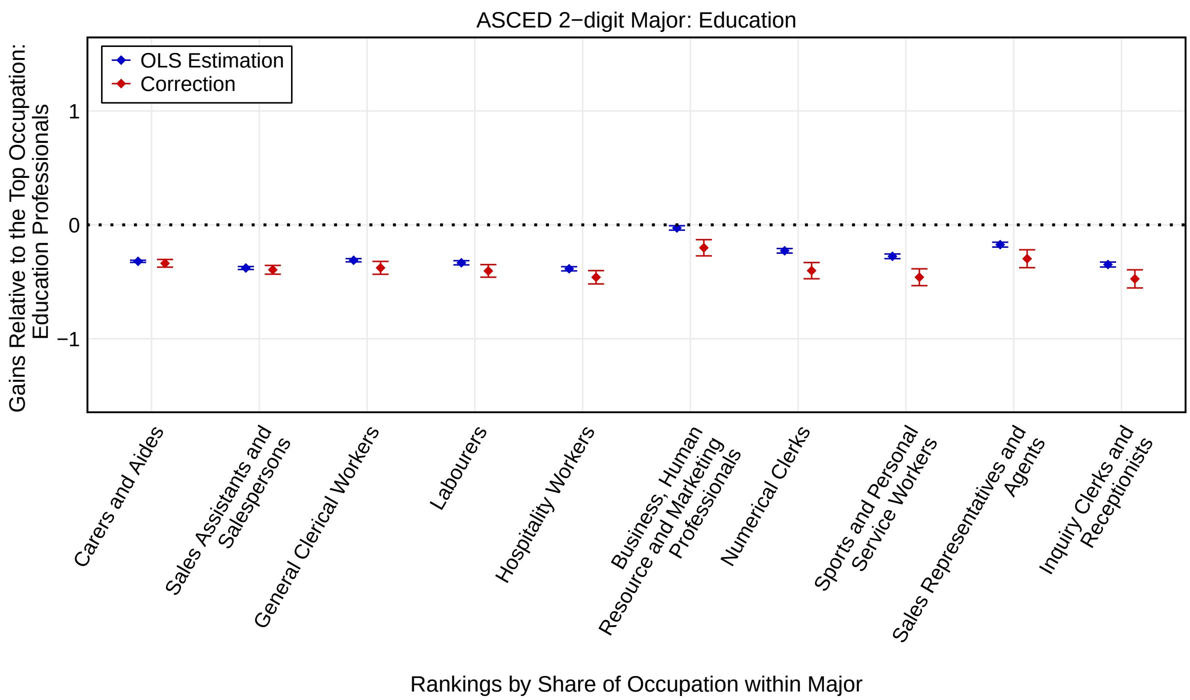


Figure Continued



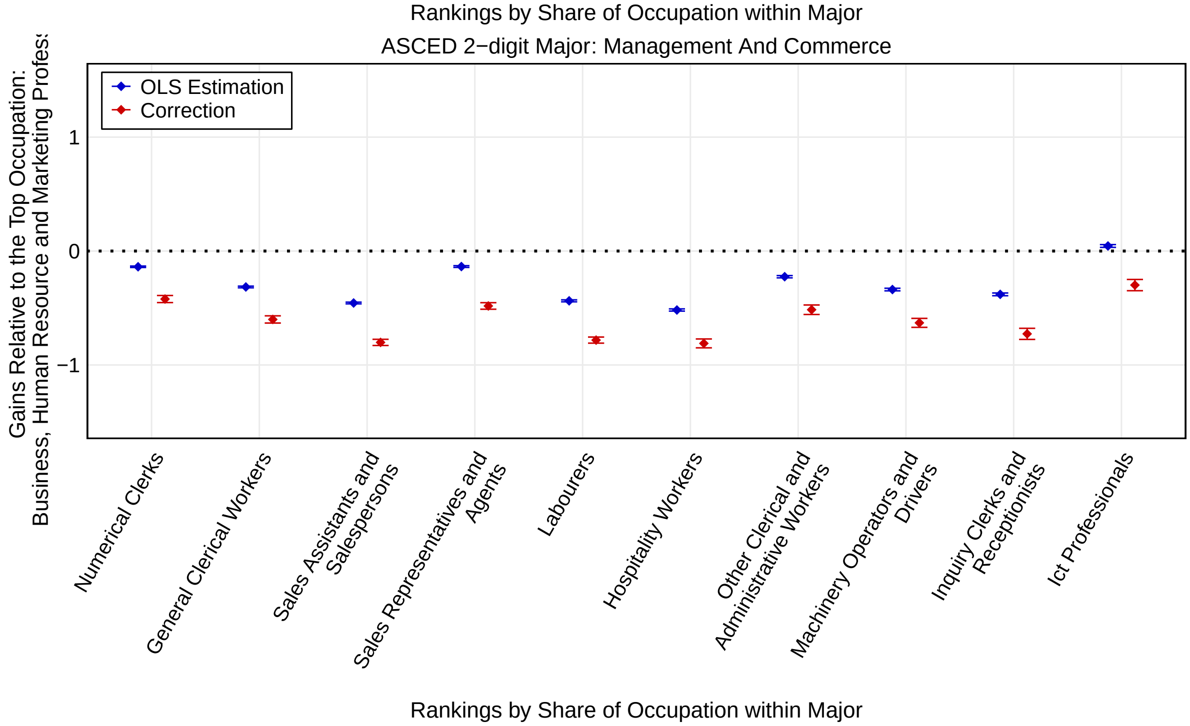
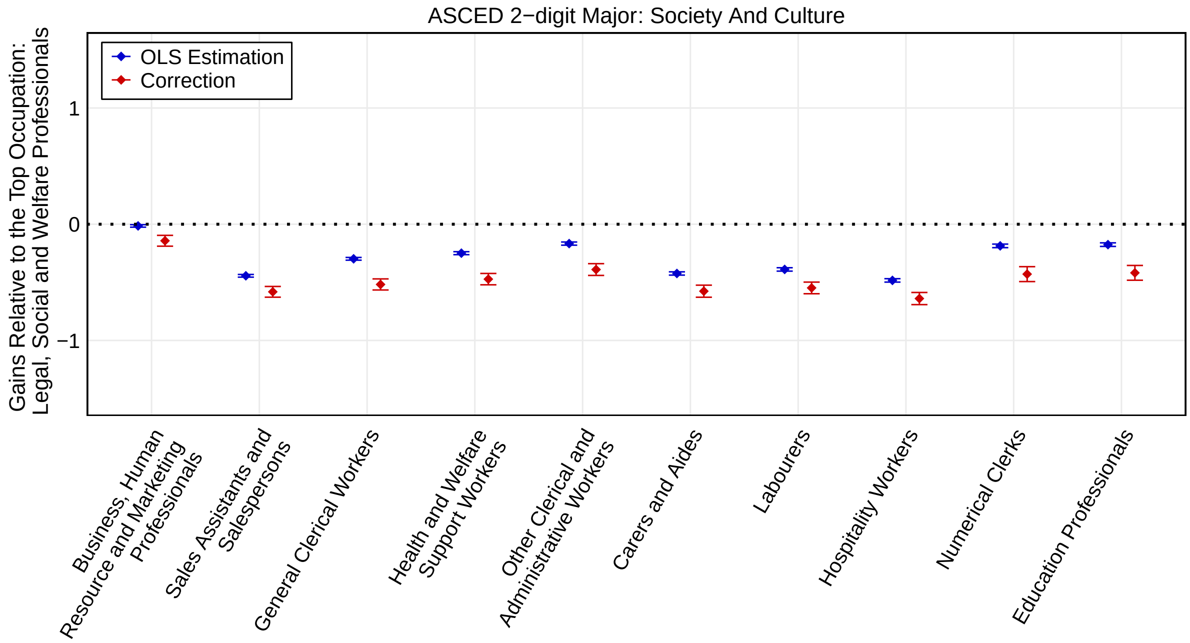
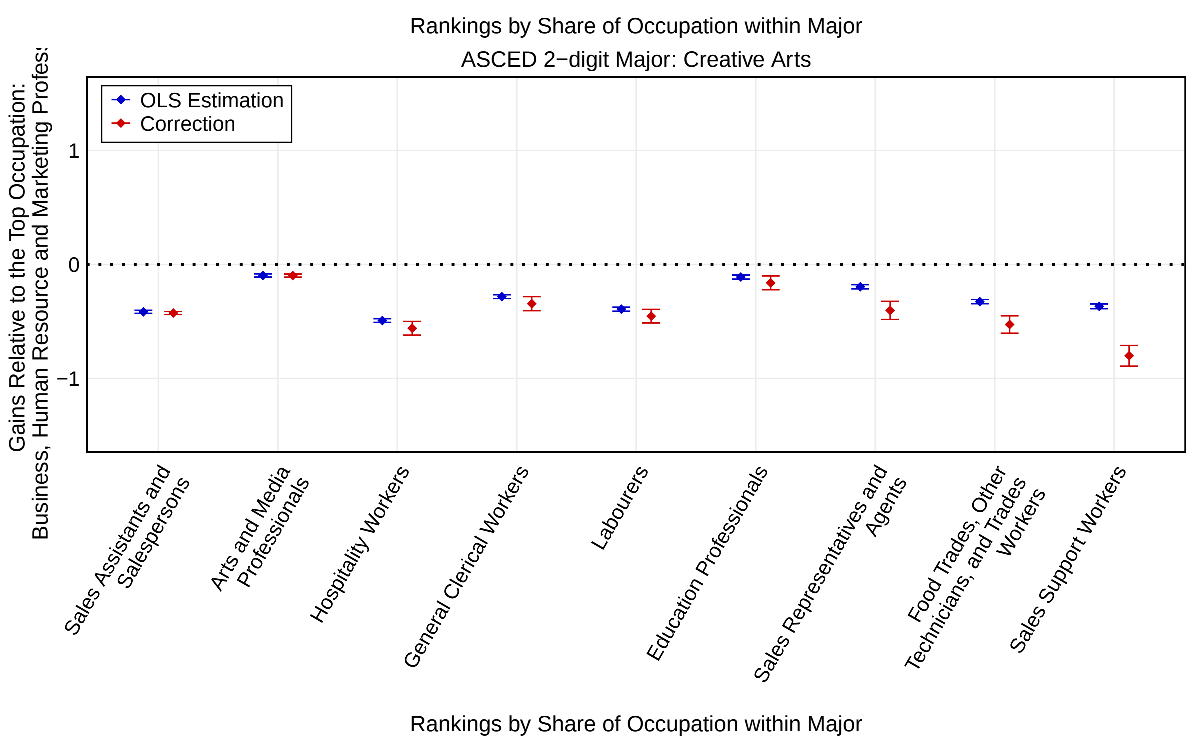


Figure Continued





Notes: These figures plot the wage penalties from not choosing the field top-occupation (as shown in y-axis) at each observed occupational choice (as shown in x-axis) for each field of study. The order of occupations in the x-axis is based on the descending order of the share of occupations, ordered from Left to Right. The scale of y-axis represents for the relative gain (or loss) in log earnings from not choosing the top-occupation. We show both of the OLS and corrected point estimate of the wage penalties $\zeta_{(k|j)}$ from Equation (25) and 90% confidence interval. To avoid cluttering the figure, we only show the penalties of the occupations whose occupational share is among top 10 of all occupations within the major.

Figure F2: Occupational Distribution by Age in Data, and Reallocated GE

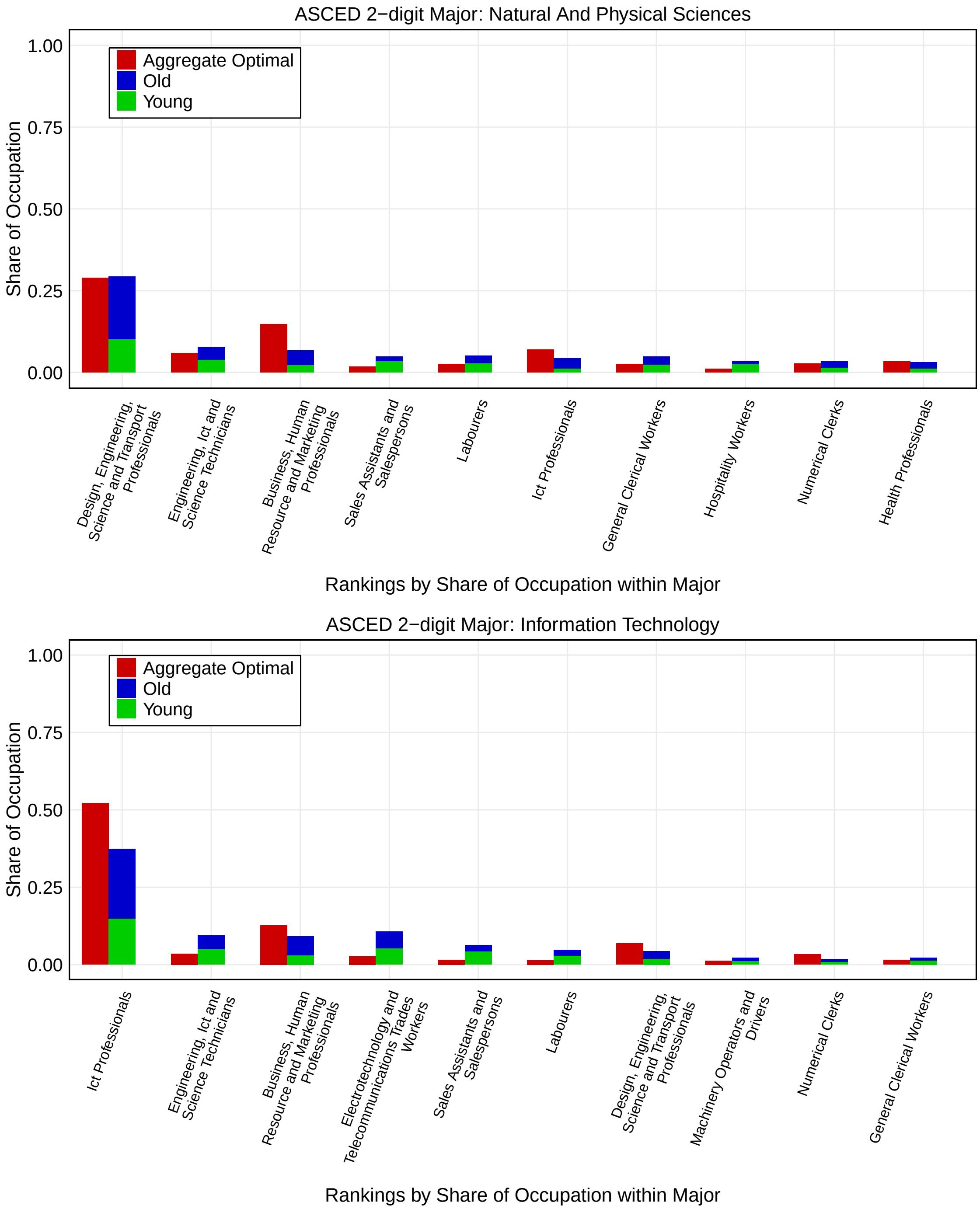
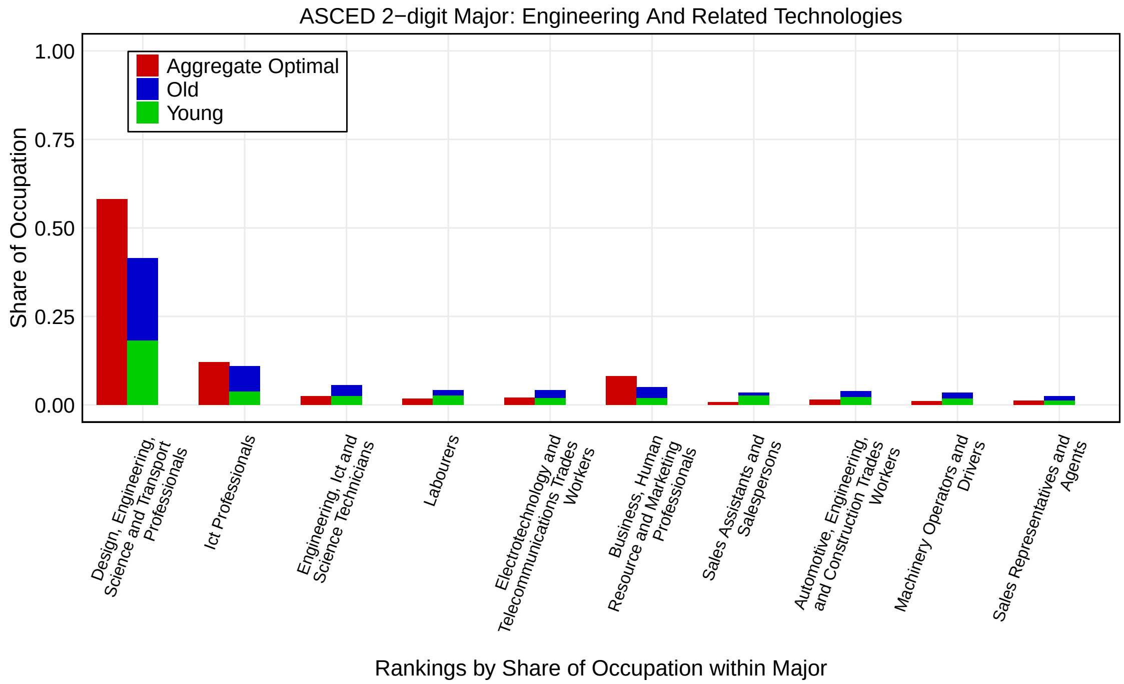


Figure Continued



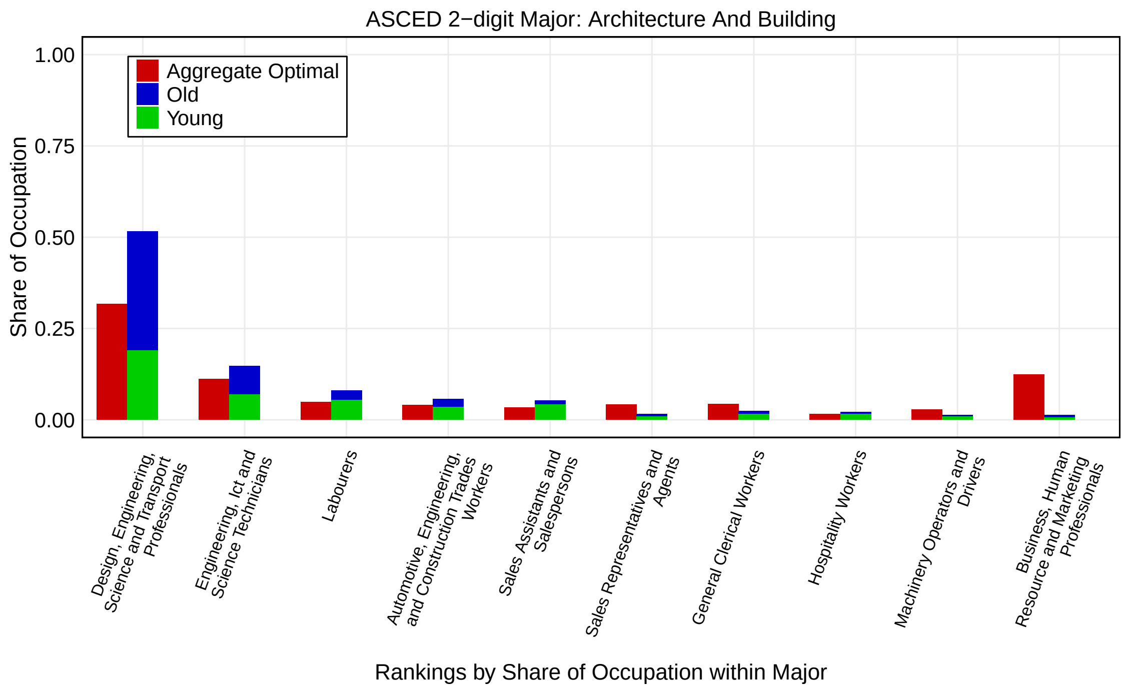
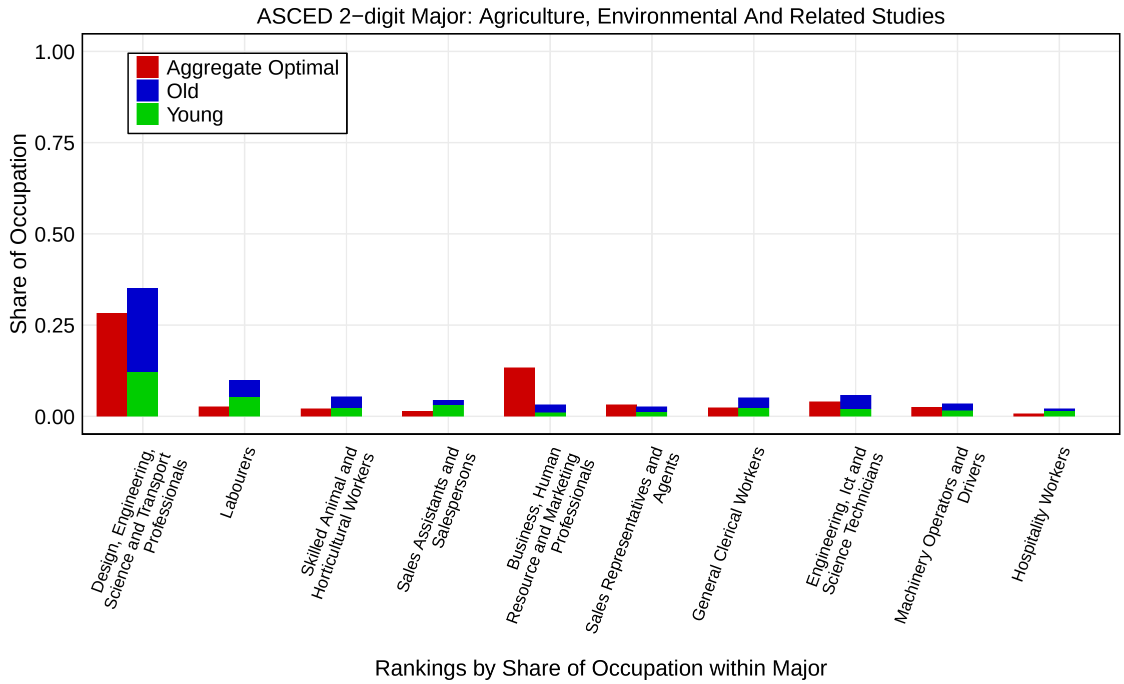


Figure Continued



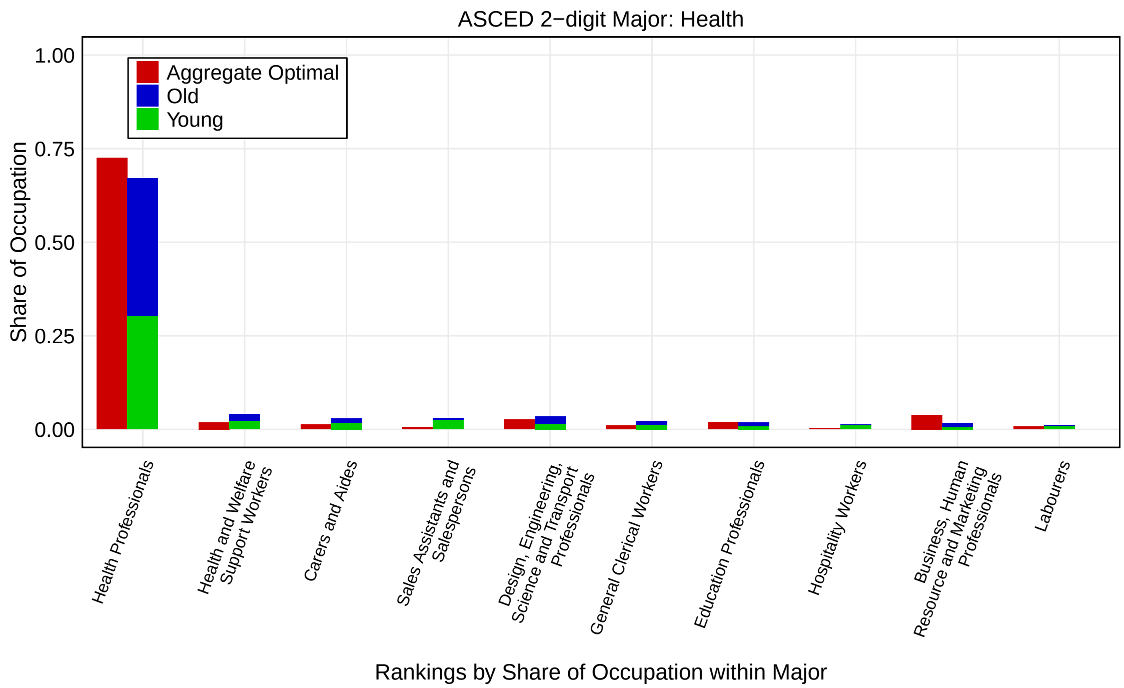
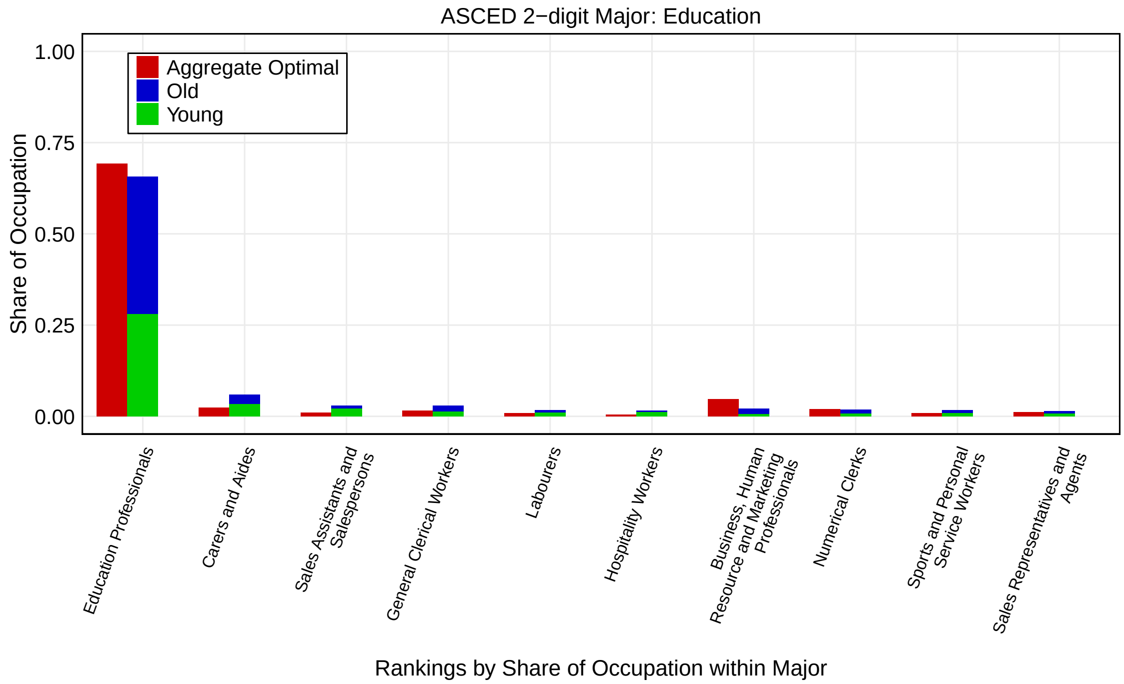


Figure Continued



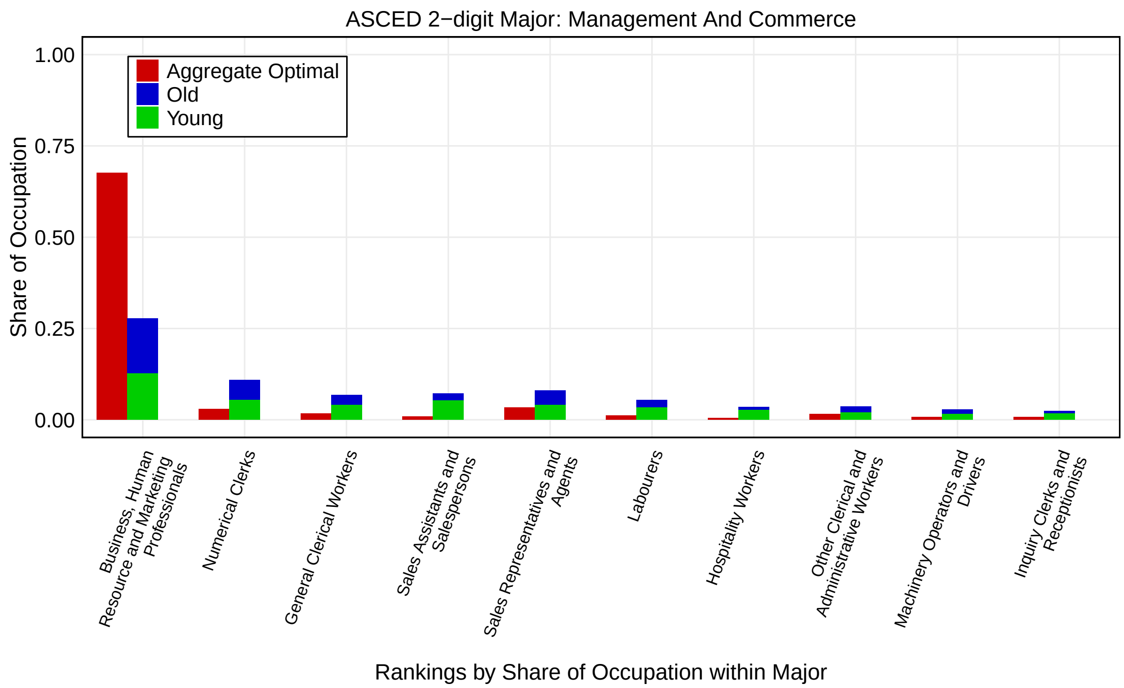
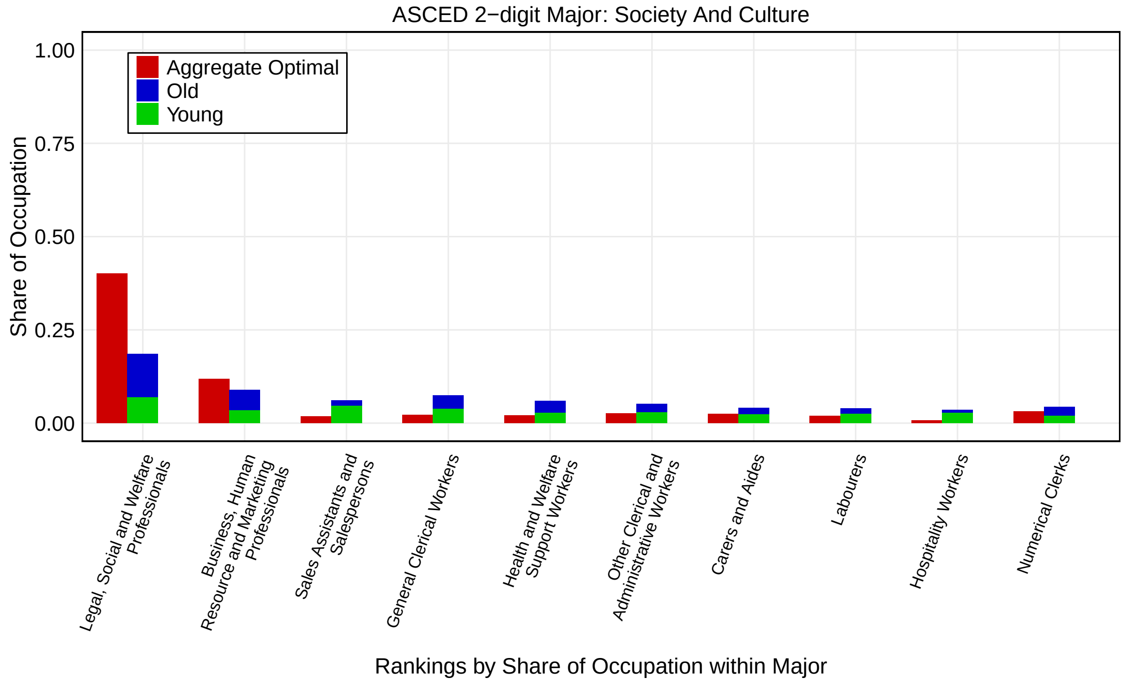
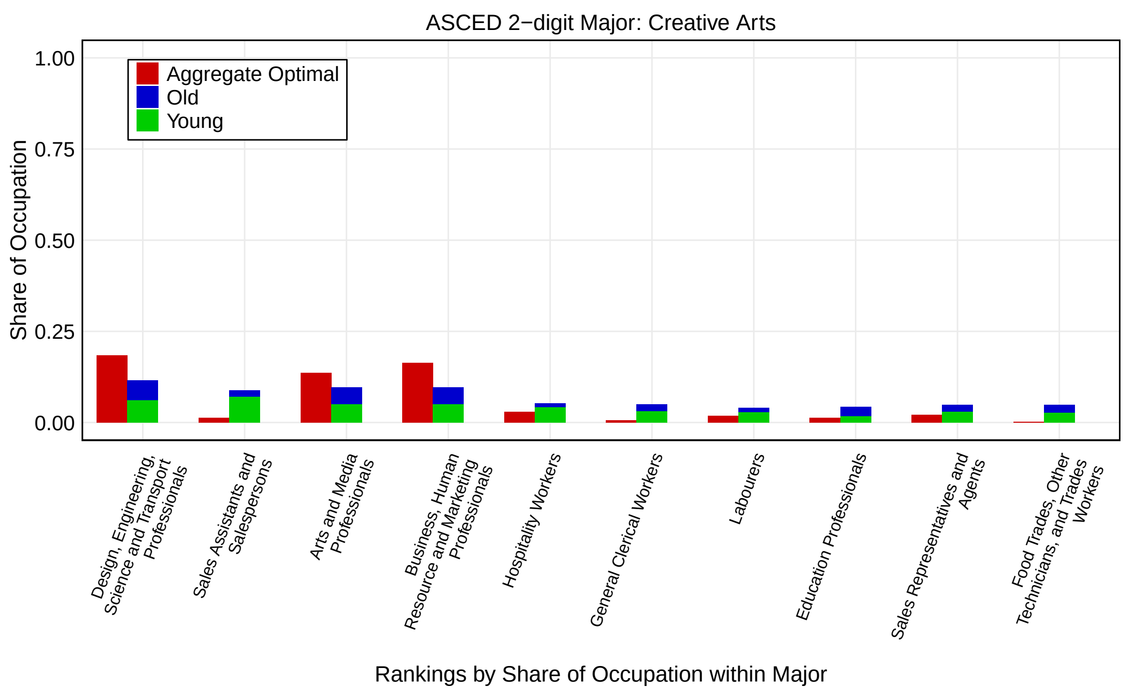


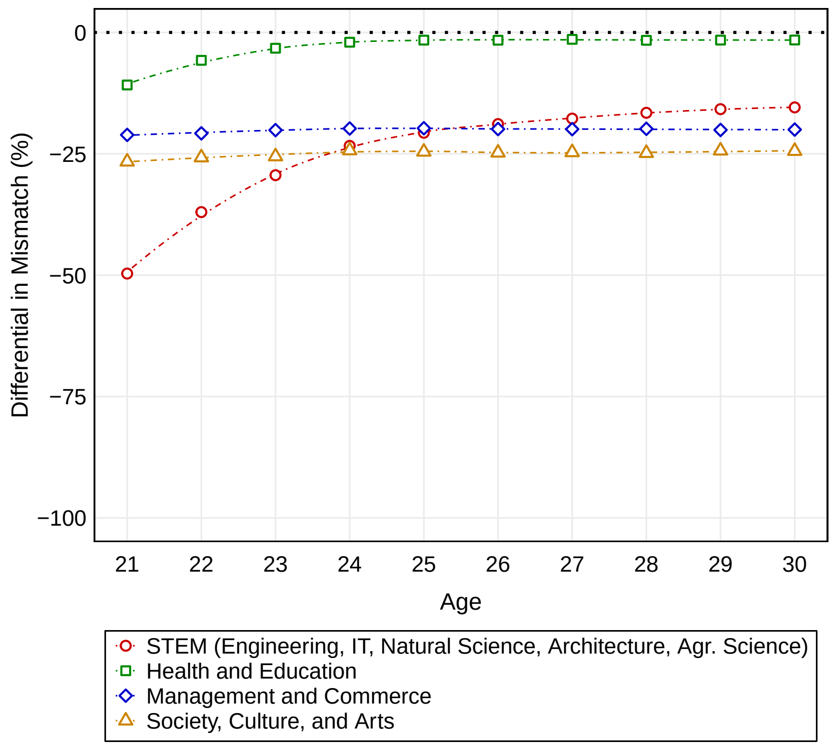
Figure Continued





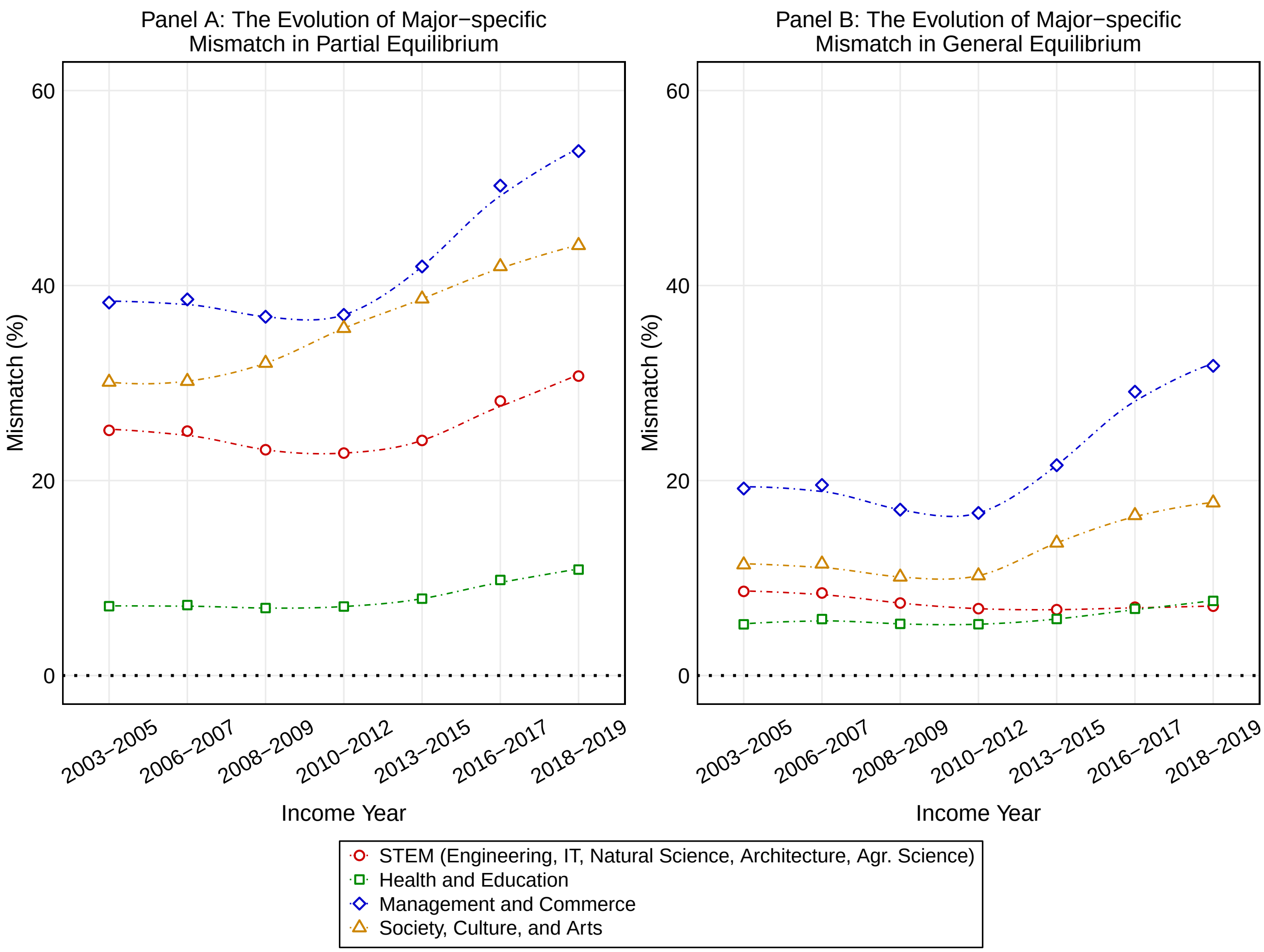
Notes: These figures plot the occupational distribution by young workers (age of 21-27) and old workers (age of 28-35) based on data, and the aggregate optimal share after reallocation of workers for each field of study. The order of occupations in the x-axis is based on the descending order of the share of occupations, ordered from Left to Right. To avoid cluttering the figure, we only show the penalties of the occupations whose occupational share is among top 10 of all occupations within the major.

Figure F3: The Difference between Major-specific Mismatch in Partial and General Equilibrium over Life Cycle



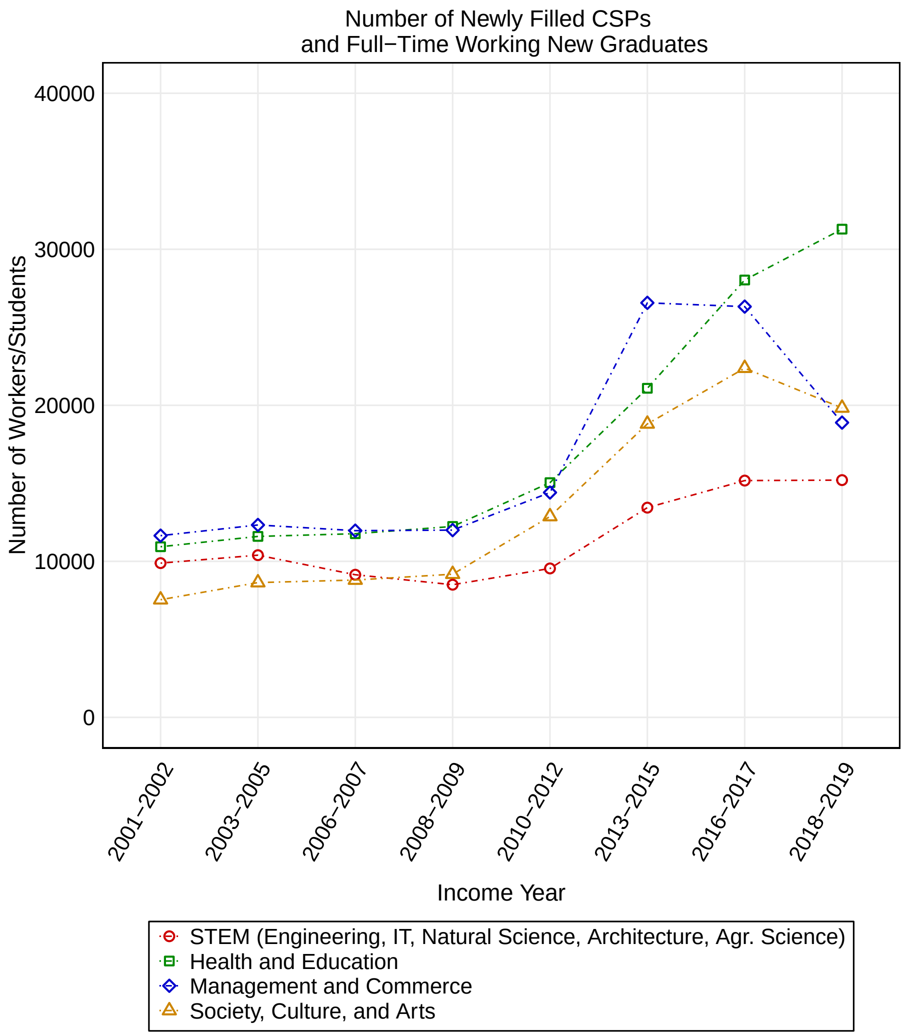
Notes: This figure plots the difference in major-specific mismatch between partial and general equilibrium over the life cycle.

Figure F4: The Evolution of Major-specific Mismatch in Partial and General Equilibrium



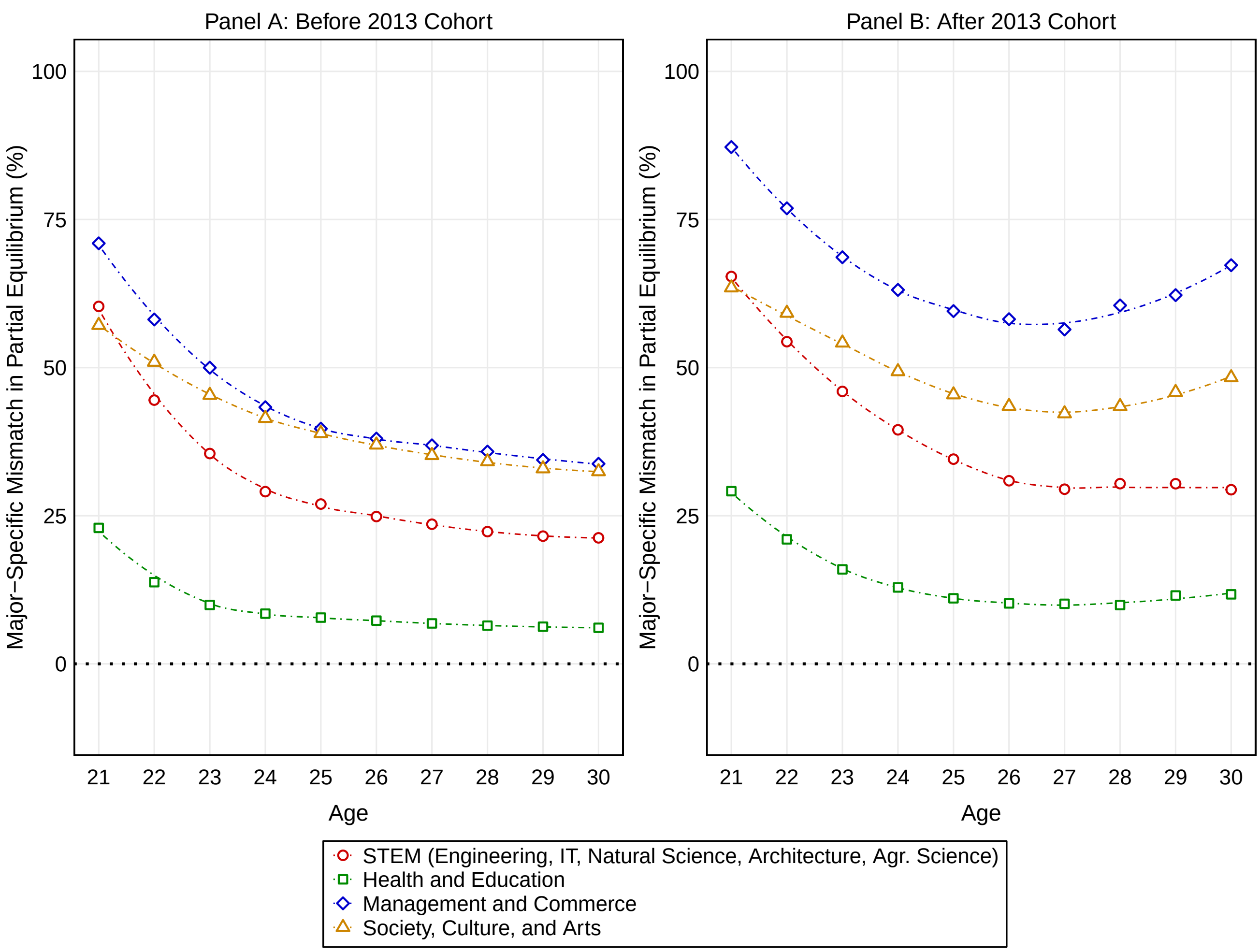
Notes: This figure plots the evolution of major-specific mismatch in partial and general equilibrium from year group 2003-2005 to 2018-2019 computed by Equations (11).

Figure F5: Number of New (Full-time Working) Graduates By Majors



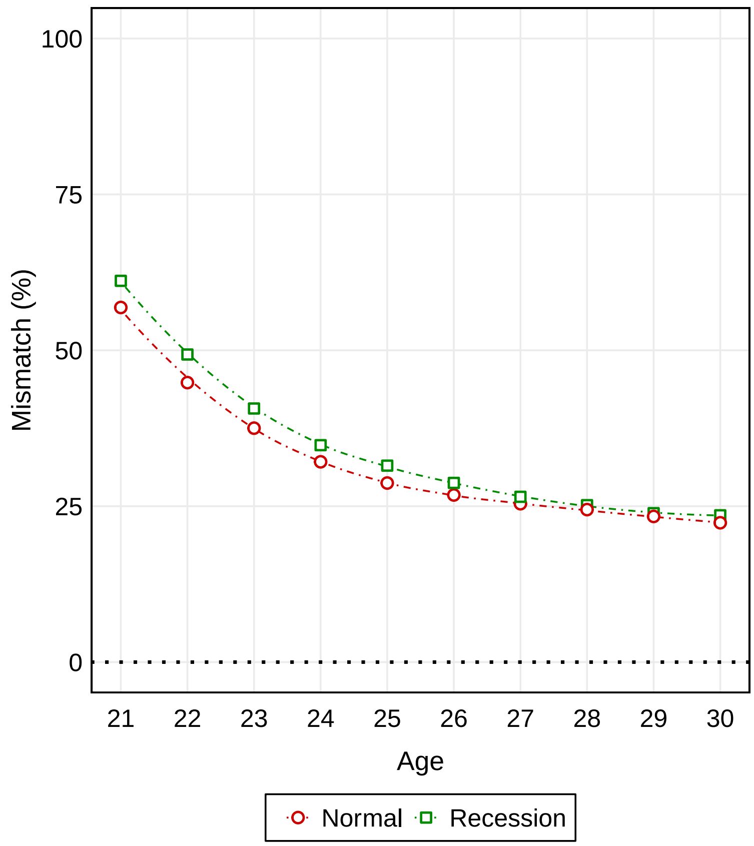
Notes: This figure plots the number of newly full-time working graduates by each major over years.

Figure F6: The Average Mismatch in Partial Equilibrium over the Life-Cycle by Cohorts



Notes: This figure plots the evolution of occupational mismatch for aggregate fields of study over the life-cycle. Cohorts are split pre- and post-policy intervention in 2009, such that cohorts graduating from 2013 onward are assumed to have studied under an uncapped system of Commonwealth-supported places. The estimate of field-specific occupational mismatch is computed using Equation (11) for each age by each cohort group.

Figure F7: The Average Mismatch in Partial Equilibrium over the Life-Cycle by Recession and Normal Period



Notes: This figure plots the evolution of occupational mismatch for aggregate fields of study over the life-cycle over the recession and normal periods. The estimate of field-specific occupational mismatch is computed using Equation (11) for each age.