WORKING PAPER 2025:7

Worker specialization and the consequences of occupational decline

Simon Ek

IFAU
INSTITUTE FOR EVALUATION OF LABOUR MARKET AND EDUCATION POLICY

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

Worker specialization and the consequences of occupational decline*

Simon Ek†

Abstract: Are workers with poor outside opportunities less responsive and more susceptible to negative demand shifts in routine occupations? To answer this, I create and estimate an occupation specialization index (OSI) using Swedish register data and machine learning tools. It measures the expected utility difference between a worker's occupation and his best outside option. This determines the loss he is willing to tolerate to avoid switching. Low-OSI generalists disproportionately left routine work. Their future wage growth was comparable to similar workers initially in non-routine occupations. By contrast, routine specialists largely stayed put and experienced lower wage growth than generalists and non-routine specialists.

Keywords: Multidimensional skills; Occupational structure changes.

JEL codes: J23; J24; J31; J62.

*I am grateful to my previous supervisors Peter Fredriksson and Georg Graetz for their continued support. I also thank Adrian Adermon, Teodora Borota Milicevic, Michael Böhm, Lars Calmfors, Mikael Carlsson, Tiernan Evans, Maarten Goos, Christoph Hedtrich, Magnus Henrekson, Caroline Hall, Sofia Hernnäs, Edvin Hertegård, Raoul van Maarseveen, Alan Manning, Guy Michaels, Oskar Nordström Skans, Per Skedinger, Joacim Tåg, Yaroslav Yakymovych, Olof Åslund, Özge Öner, and participants at workshops/seminars at the Department of Economics at Uppsala University, the Centre for Economic Performance at London School of Economics, the Institute for Evaluation of Labour Market and Education Policy (IFAU), the Research Institute of Industrial Economics (IFN), the Department of Economics and Statistics at Linnaeus University, the Ratio Institute, and the Swedish Public Employment Service for valuable comments and discussions. Any errors are my own.

†simon.ek@ifau.uu.se. The Institute for Evaluation of Labour Market and Education Policy (IFAU) and Uppsala center for Labor Studies (UCLS).

1 Introduction

Many developed economies have experienced large occupational structure shifts in recent decades. Generally, routine-intensive occupations with codifiable tasks have declined in favor of non-routine work. This is partly due to advancements in labor-replacing technologies. The occupations that declined were commonly located in the middle of the wage distribution. These shifts have therefore contributed to wage polarization. Moreover, a recent literature finds that the growth over time in occupation wage premia and employment are positively correlated. This strongly suggests that these occupation structure shifts, including the substantial decline in routine work, were caused by changing demand.[1]

There are strong public concerns regarding how well incumbent workers will cope with future shifts in labor demand across occupations.[2] This paper develops a method grounded in theory for identifying which workers are particularly vulnerable to negative shifts in occupation labor demand.[3] It is based on estimating how much value a worker puts on his current occupation relative to his outside options. This difference is referred to as workers' degree of occupation specialization. I then ask two questions: First, how does specialization relate to the career consequences of incumbent employees following the historical decline in routine work? This query is important in its own right. But it is also a means of substantiating the general usefulness of my method for predicting worker susceptibility. Second, what is the nature of worker flows out of routine occupations? This is informative about the process by which these occupations decreased in size. It also speaks to whether mobility was voluntary for workers lacking good alternatives.

I begin by setting up a Roy-style discrete choice model with deterministic (which depend on workers' characteristics) and idiosyncratic utility terms. When a demand shock lowers the wage premium in a worker's occupation, his utility loss is determined by the difference between his utility in that occupation and his best non-shocked outside option, i.e., his spe-

[1] See, e.g., Goos et al. (2014) and Goos et al. (2019) for occupation structure shifts across countries; Autor et al. (2003), Goos and Manning (2007), Autor et al. (2008), and Adermon and Gustavsson (2015) for the literature on routine-biased technological change and wage polarization; Cortes (2016), Böhm (2020), Cavaglia and Etheridge (2020), and Böhm et al. (2023) for studies on occupational wage premium and employment growth.

[2] See, e.g., Brynjolfsson and McAfee (2014), Mokyr et al. (2015), Frey and Osborne (2017), and OECD (2019). For example, future technology is envisioned to be able to perform many tasks previously considered impossible, such as writing news articles and driving cars.

[3] There exists many projections (forecasts by, e.g., the U.S. Bureau of Labor Statistics Occupation Projections) and ample work (e.g., The OECD Future of Work initiative, Frey and Osborne 2017, Arntz et al. 2017, and Webb 2019) on future occupation employment. I am not aware of any predictions at the worker level. This paper provides a framework for making such projections.

cialization. Workers with little specialization move to a now-more attractive option. Highly specialized workers remain, losing more utility. I hypothesize that this effect is working partly through wages rather than only through amenities. For workers in occupations that did not experience any negative demand shift, specialization should be less important for future outcomes. Next, I show that under Gumbel distributed idiosyncratic preferences, expected specialization is a function solely of a worker's *ex ante* probability of working in a non-shocked outside option. I name this expected value the occupation specialization index (OSI). I also demonstrate that ordering workers by the OSI is equivalent to ordering them by the expected utility loss from a negative wage premium shock of any (unknown) size.

To construct the OSI empirically, I grow a large number of decision trees. The trees predict occupation choice probabilities using Swedish register data on male worker characteristics. These include multidimensional abilities (collected during the Swedish enlistment process), educational attainment, region of residence, age, and industry-specific experience. The reason for focusing on male workers is twofold: Men were disproportionately exposed to the historical decline in routine occupations; and the multidimensional ability data cover the vast majority of males from certain cohorts. Moreover, the ability data have been shown to be important determinants of, e.g., wages and occupational sorting (see, e.g., Fredriksson et al. 2018). I use Autor and Dorn (2013)'s routine task intensity index (RTI) to classify occupations as either routine (above) or non-routine (below-median RTI). Between 2001 and 2013, non-routine occupations remained stable or grew while routine occupations declined (by 6 percentage points on average).

The OSI is expected to be more negatively related to occupation switching and wage growth for routine than non-routine workers. To test this, the OSI is related to the long-run (up to twelve years into the future) career outcomes of workers observed in 1997-2001 separately by initial routine and non-routine low- to middle-skilled occupations.[4] The probability of leaving routine work depends strongly on initial OSI. I interpret this as switching typically being voluntary, guided by the attractiveness of workers' outside options. The wage growth penalty of initially working in a routine occupation increases in the OSI. On average, routine specialists experienced lower wage and earnings growth than both low-OSI workers in either type of occupation and non-routine specialists. These results are consistent with the prediction that low-OSI "generalists" are better able to avoid losses from declining demand by

[4]With only one exception, the routine occupations are classified as low- or middle-skilled. To obtain a more comparable comparison group of non-routine occupations, I exclude higher-skilled occupations in the main analysis. But these are included in robustness checks. The low- to middle-skilled routine occupations are mainly concentrated in manufacturing while the non-routine are found in, e.g., services, construction, and transportation.

transitioning to more attractive occupations.

The paper contributes to the literature on routine-biased technological change. Several previous papers in this literature highlight that susceptibility to negative demand shifts is determined by the difference between a worker's current utility and potential utility in his non-shocked options. My addition is to show that, under certain assumptions, my occupation specialization index measures the expected value of this difference. The index is simply a monotone transformation of the *ex ante* propensity for working in a non-shocked outside option. To the best of my knowledge, this closed-form solution has not been utilized before. I also demonstrate that the index predicts well which incumbents lost from the historical decline in demand for routine work. Hence, the Roy (1951) model seemingly provides an informative representation of who loses when occupation labor demand declines. Moreover, the observable worker attributes at hand can be used to characterize workers' occupation options well.

There exists a few studies on the consequences of working in declining or routine occupations. Edin et al. (2022a) follow workers in occupations that later experienced arguably unanticipated decline using both Swedish and U.S. data. Ross and Ukil (2021) relate future industry employment to the future earnings of workers in the NLSY. Böhm (2020) and Jaimovich et al. (2021) use cross-sectional AFQT data to study changes in outcomes of workers with skill bundles fit for routine work. Cortes et al. (2017) describe which demographic groups in the U.S. contributed to the decline in routine employment. Bachmann et al. (2019) demonstrate that working in routine occupations in Germany is associated with low job stability and a high risk of unemployment. Cortes et al. (2020) find that the decline in routine work can to a large extent be accounted for by changing transition rates from non-employment to routine occupations.

More generally, there is a large literature explaining occupation switching behaviour (e.g., Jovanovic and Nyarko 1997; Gathmann and Schönberg 2010; Groes et al. 2015; Cortes and Gallipoli 2018). I also relate to research using direct measures of worker attributes to infer job-specific match quality (e.g., Fredriksson, Hensvik, and Skans 2018; Guvenen et al. 2020; Lise and Postel-Vinay 2020) and to the literature on the importance of different types of skills (e.g., Lindqvist and Vestman 2011; Deming 2017; Roys and Taber 2022; Edin et al. 2022b). Finally, there are some studies of the role of occupation-specific human capital (often proxied by the distance between the task content of their initial and other occupations) for the ability to adjust to, e.g., mass layoffs (Robinson 2018) and trade shocks (Traiberman 2019; Eggenberger et al. 2022).

Section 2 describes the discrete choice model, how I create the specialization index, and the econometric framework for the empirical analysis. Section 3 presents the data. Section 4 reports the empirical results. Section 5 concludes. Appendix A presents the derivation of the specialization index. Additional empirical results are reported in Appendix B.

2 Conceptual framework

2.1 Deriving the occupation specialization index

A discrete occupation choice model. Consider a setting where workers are characterized by multiple attributes collected in the vector $\boldsymbol{x}$. There exists a finite number of occupations collected in the set K that are either classified as routine ($R \subset K$) or non-routine ($N \subset K$). Different workers are differentially suited for working in different occupations. The utility of worker i in occupation k is:

$$u_{ik} = \pi_k + m_k(\boldsymbol{x}) + q_k(\boldsymbol{x}) + \varepsilon_{ik}. \tag{1}$$

π_k represents the wage premium in k for a unit of skill, $m_k(\boldsymbol{x})$ determines the group-specific productivity in k and $q_k(\boldsymbol{x})$ captures any amenities. ε_{ik} is an idiosyncratic term at the individual $\times$ occupation level that may influence both the wage and amenities. Define deterministic utility as $u_k(\boldsymbol{x}) \equiv \pi_k + m_k(\boldsymbol{x}) + q_k(\boldsymbol{x})$. Workers choose the occupation with the highest utility. Denote a worker's initial choice by j.

Utility loss following a routine wage premium shock. Due to automation, there is a wage premium shock to all routine occupations of size $-\delta$, with $\delta > 0$. Define $d_j \equiv \delta \mathbb{1}[j \in R]$. Workers in R will choose whether to switch to another occupation.The initial occupation, j, yields a higher utility than any other occupation in R both before and after the shock. Therefore, the only relevant options are the non-shocked occupations in N. Since $d_j = 0$ for workers in N, their outside options do not matter. But if their occupation would have experienced the shock, the relevant option would be the best choice in the set of non-routine occupations excluding j, i.e., $N \backslash \{j\}$. I therefore focus on non-routine outside occupations when defining the relevant outside options and, later on, constructing the occupation specialization index, for workers initially in both routine and non-routine occupations.

The utility loss from staying in j is $-d_j$. The loss from switching is the difference between the initial utility in j and the best, non-shocked, outside option, i.e., his initial utility surplus. The worker will choose the option with the smallest associated loss. Thus, the change in utility is:

$$\Delta u_{ij} = \max\left\{-d_j,\ -\left(u_{ij} - \max_{n\in N\setminus\{j\}} u_{in}\right)\right\} \leq 0. \tag{2}$$

The loss, in absolute terms, is bounded above by workers' utility surplus relative to their best non-routine outside option. I call this their degree of occupation specialization. Specialists with a large surplus remain and tolerate the full effect of the wage premium shock. Workers with a small surplus will instead move and experience a smaller utility loss.

For routine workers, $N\setminus\{j\}$ equals N. Thus, the surplus is defined slightly different for routine and non-routine workers. But it captures the difference between current utility and the utility associated with the best non-routine outside option for both groups.

Gumbel distributed idiosyncratic terms. I henceforth assume that all ε_{ik} are IID standard extreme value type I (or Gumbel) distributed. Although the utility surplus in (2) is never observed, four key properties of the Gumbel distribution allow me to characterize the distribution of this surplus: First, the occupation choice probabilities follow the multinomial logit (see McFadden 1973). Second, the distribution of maximum utility before conditioning on occupation choice is also Gumbel distributed, with known location and scale. Third, using results from Hanemann (1984), the maximum utility conditional on any optimal choice j can be shown to be distributed the same as the unconditional maximum. Fourth, the difference between two same-scaled Gumbels is known to be logistic distributed. The details of characterizing the utility surplus distribution are reported in Appendix A.

The occupation specialization index. I now turn to describing the closed-form expression for the expected value of the utility surplus. I name this expression the occupation specialization index, or OSI. It reveals the expected utility a worker would lose if leaving his current occupation. Thus, one may interpret the OSI as how dependent, on average, workers are on their occupation for utility. The derivation of the OSI, and its properties, is described in Appendix A.

First, define the *ex ante* probability of working in a non-routine outside option as follows. It is determined by a worker's characteristics, his observed occupation, and the set of non-routine occupations:

$$\rho(\boldsymbol{x}, j, N) \equiv \sum_{n\in N\setminus\{j\}} p_n(\boldsymbol{x}). \tag{3}$$

Next, in Appendix A, I show that the expected value of the utility surplus is a monotonically decreasing function of $\rho(\boldsymbol{x}, j, N)$. This is the occupation specialization index:

$$\mathrm{OSI}(\boldsymbol{x}, j, N) \equiv \mathbb{E}\left[u_{ij} - \max_{n \in N \setminus \{j\}} u_{in} \,\middle|\, \boldsymbol{x}, j\right] = -\frac{\ln\left(\rho(\boldsymbol{x}, j, N)\right)}{1 - \rho(\boldsymbol{x}, j, N)}. \tag{4}$$

To the best of my knowledge, this metric has neither been used previously in any work on occupation decline or routine-biased technological change, nor been derived in the theoretical discrete choice literature. The OSI is more general than in my application: It can be used to infer the expected utility surplus of any observed choice relative to a subset of alternatives.

The expected value of the utility change following the wage premium shock in (2) is a function of only d_j and $\rho(\boldsymbol{x}, j, N)$:

$$\mathbb{E}\left[\Delta u_{ij} \,\middle|\, \boldsymbol{x},\ j\right] = -\frac{d_j - \ln\left(1 + \left(e^{d_j} - 1\right)\rho(\boldsymbol{x}, j, N)\right)}{1 - \rho(\boldsymbol{x}, j, N)}. \tag{5}$$

For any value of $d_j = \delta > 0$, it is monotonically increasing (i.e., the absolute loss becomes smaller) in $\rho(\boldsymbol{x}, j, N)$. Since the OSI is decreasing in $\rho(\boldsymbol{x}, j, N)$, ordering workers by the OSI is equivalent to ordering them by absolute expected loss for any shock size.

To summarize, the utility loss from an occupation wage shock is determined by the difference between a worker's utility in his occupation and his best non-shocked outside option. The expected value of this difference can be inferred from his *ex ante* probability of working in an outside, non-shocked occupation via the OSI. The OSI can also be used to order workers by expected loss from a shock of unknown size.

Intuitively, the OSI can be comprehended as follows. The extent to which a worker's peers with similar characteristics are observed in outside non-routine occupations carries a signal about his non-routine options. If no other workers with, say, a comparable skill set and education background works in an occupation other than his, these options are likely unattractive or unavailable to him. This idea is similar in spirit to the widely used revealed comparative advantage metric by Balassa (1965). It is also related to, e.g., Fredriksson, Hensvik, and Skans (2018) who measure match quality by the similarity of the skills of workers and their co-workers, Coraggio et al. (2025) who define match quality as the probability of being observed in an occupation-industry cell, and the outside options index developed by Caldwell and Danieli (2022).

Although this framework concerns utility, the empirical section deals with, e.g., wages and earnings. Workers that remain in occupations with decreasing wage premia will experience the effect on utility through wages. For switchers, however, utility may be influenced through wages or amenities. These are difficult to disentangle. But in Section 4.1, I show that specialization correlates positively with the wage surplus that a worker enjoys in his occupation.

Predictions. From the above framework, I highlight two predictions:

1. Expected loss: The expected loss from a negative demand shift in routine occupations increases in an incumbent's expected degree of specialization. Workers in non-routine occupations will be less affected by such a shock and their specialization matters less than for workers initially observed in routine work. The difference in the average loss between the two groups will therefore increase in specialization. I hypothesize that the loss in utility acts partly through wages rather than only through amenities.

2. Expected worker flows: There is negative selection on the OSI in who leaves the routine occupations: Highly specialized workers will to a larger extent remain and tolerate the full effect of the negative shock. Again, specialization should matter less for the probability of switching to another occupation for workers initially in non-routine work; It should primarily be routine generalists that engage in occupation switches.

2.2 Estimating the occupation specialization index

Decision tree classifier. To construct the OSI empirically, I begin by dividing all observations it in a given sample into groups, $\hat{g}(\boldsymbol{x}_{it})$, based on their observed characteristics, $\boldsymbol{x}_{it}$, and occupation choices, j_{it}, using a decision tree classifier. The decision tree is grown using a log-loss splitting rule, where observations are sequentially split into smaller and smaller groups in order to minimize $-\sum\limits_{it}\sum\limits_{k\in K}\mathbf{1}[j_{it}=k]\ln(p_k(\boldsymbol{x}))$. Thus, this machine learning process divides all observations into distinct groups to capture as much between-group heterogeneity in occupational sorting as possible. All potential values of all features in $\boldsymbol{x}_{it}$ are considered as the cutoff at each splitting decision. The estimated probabilities are simply the group-specific share of observations in each occupation in the estimation sample.

To diminish overfitting, I grow so called "honest trees" that completely separate the data used for deciding the model structure from the data used for estimation conditional on that structure (see, e.g., Athey and Imbens 2016). Approximately half of the sample is used to construct each tree, while the other half estimates group-specific probabilities for each occupation. I use a minimum leaf size of 250 observations.[5]

[5] This hyperparameter was selected by evaluating models with various minimum leaf sizes on both bootstrapped training data and out-of-bag data. A leaf size of 250 yields a high out-of-bag ROC-AUC with only a slight

Next, I compute $\hat{\rho}(\boldsymbol{x}_{it}, j_{it}, N)$ using the predicted probabilities by group $\hat{g}(\boldsymbol{x}_{it})$, which is then incorporated into the OSI formula in (4) to obtain $\widehat{\text{OSI}}(\boldsymbol{x}_{it}, j_{it}, N)$. To limit the influence of outliers, the OSI is censored at the 1st and 99th percentiles. Since the OSI is challenging to interpret directly, it is standardized to have a mean of zero and a standard deviation of one.

Standard errors. The OSI is an estimated metric, intended for use in subsequent empirical analyses. To obtain valid standard errors, I repeat the decision tree growing process on multiple bootstrapped samples for 1000 iterations. All regression models outlined in the section below are then estimated on each bootstrapped sample, using the groups and estimated OSI from each specific bootstrap. All regression analyses report either the standard deviation or the 5th and 95th percentile of the estimates across the bootstrapped samples.

2.3 Econometric framework

Motivation. The theoretical framework motivates a difference-in-differences-styled specification comparing the effect of the OSI on the outcomes of workers initially in routine and non-routine occupations. This holds constant any common effect of being specialized relative to the non-routine outside options by using non-routine workers as a comparison group to the "treated" routine workers. However, since the relationship between workers' characteristics and the OSI may, and in fact does, differ between routine and non-routine occupations, we also need to consider potentially systematic differences in workers' overall expected career outcomes. I account for this by either controlling for the underlying worker characteristics, or alternatively, by introducing fixed effects for all groups $\hat{g}(\boldsymbol{x}_{it})$ from the decision tree classifier, thus effectively utilizing variation in occupation specialization and outcomes within each group due to differences in sorting across occupations.

Main regression model. Define $r(j_{it}, N)$ as an indicator taking the value one if j_{it} is a routine occupation and zero otherwise, i.e., $r(j_{it}, N) \equiv \mathbb{1}\left[j_{it} \notin N\right]$. Next, let $y_{it+\tau}$ represent an outcome in year $t+\tau$ of individual i initially observed in t. The following model is considered the main specification:

$$y_{it+\tau} = \psi \widehat{\text{OSI}}\left(\boldsymbol{x}_{it}, j_{it}, N\right) + r\left(j_{it}, N\right)\left(\beta + \phi \widehat{\text{OSI}}\left(\boldsymbol{x}_{it}, j_{it}, N\right)\right) + \boldsymbol{\lambda}' \boldsymbol{z_{it}} + \epsilon_{it+\tau}. \tag{6}$$

Generally, $\boldsymbol{\lambda}' \boldsymbol{z_{it}}$ will incorporate controls for characteristics or fixed effects for the decision-tree classifier groups $\hat{g}(\boldsymbol{x}_{it})$. ψ captures the common effect of the OSI. β captures the difference

accuracy difference between training and out-of-bag data.

between routine and non-routine workers at average OSI, and the difference in the effect of the OSI between the two occupation groups is captured by ϕ. The model with group FE is equivalent to using the within estimator, i.e., using deviations from the group-specific averages in both the outcome and explanatory variables, to estimate ψ, β, and ϕ. Variations of the baseline model will also include occupation-fixed effects, but then β is no longer identified.

Semi-parametric model. To obtain non-parametric estimates of the effect of specialization, I will also estimate models with indicators for the quintile groups of the OSI distribution separately for routine and non-routine workers of the following form:

$$\begin{aligned} y_{it+\tau} = & \sum_{z\in\{2,\dots,5\}} \theta_z^N \left(1 - r\left(j_{it}, N\right)\right) q_z \left(\mathrm{OSI}(\hat{g}(\boldsymbol{x}_{it}), j_{it}, N)\right) \\ & + \sum_{z\in\{1,\dots,5\}} \theta_z^R r\left(j_{it}, N\right) q_z \left(\mathrm{OSI}(\hat{g}(\boldsymbol{x}_{it}), j_{it}, N)\right) + \boldsymbol{\lambda' z_{it}} + \epsilon_{it+\tau}. \end{aligned} \tag{7}$$

q_z symbolizes an indicator for belonging to the zth quintile group of the OSI distribution. The reference category is the lowest non-routine quintile group.

3 Data

3.1 Variables

Background characteristics. I collect data on individual characteristics from Swedish population-wide administrative registers. I use 13 different levels of educational attainment,[6] and 25 different fields of study,[7] according to the 2-digit categories of the Swedish Standard Classification of Education (SUN, based on ISCED). I also calculate work experience between $t-13$ and $t-1$ in 14 industries according to the Swedish Standard Industry Classification (SNI).[8]

Wage and occupation information. I use workers' full-time equivalent monthly wages from the Swedish Wage Structure Statistics (*Lönestrukturstatistiken*; WSS) survey. I also collect

[6]Preschool; compulsory <9 or 9-10 years; secondary <2, 2, or 3 years; post-secondary <2, 2, 3, 4 or ≥ 5 years; licentiate or similar degree; doctoral degree.

[7]Basic; literacy and numeracy; personal skills; teacher training and education science; arts and media; humanities; social and behavioural science; journalism and information; business and administration; law; life science; physical science; mathematics and statistics; computing; engineering and engineering trades; manufacturing and processing; architecture and building; agriculture, forestry, fishery; veterinary; health; social services; personal services; transport services; environmental protection; security services.

[8]Agriculture and related; mining and quarrying; manufacturing; electricity, gas and water supply; construction; wholesale and retail trade; hotels and restaurants; transport, storage and communication; financial intermediation; real estate and renting; public administration; education; health care and social services; other services.

information on occupation at the two-digit level of the Swedish Classification of Occupations (SSYK, based on ISCO). A few, very small, occupations are excluded. I also exclude managers and politicians. The final data include 22 occupations.[9]

Survey weights. Sampling in the WSS occurs at the firm/organization level. All public and almost 50 percent of private sector employees are sampled each year. The data include weights used to make any constructed moments representative of the full employee population.[10] To mitigate issues with extreme weights being put on a few very small firms in certain industries, I censor the weights at the 99th percentile. I use these survey weights for all empirical exercises, including growing the decision trees.

Multidimensional skills. I utilize information on cognitive and non-cognitive skills from the Swedish War Archive. These data were collected during the Swedish draft process in 1969-1994. They are available for around 90 percent of males born in 1951-1976 who underwent the draft at the age of 18 or 19.[11] The draftees performed four standardized cognitive tests on: Inductive reasoning; verbal comprehension; spatial ability, and; technical understanding. They also took part in a 25 minute interview with a psychologist. The psychologist evaluated the profile of the draftee and scored them along four dimensions. Mood et al. (2012) interprets these as: Social maturity; psychological energy, focus or perseverance; intensity or activation without pressure; emotional stability or tolerance to stress. The scaling of the scores varies by test type and draft cohorts. I standardize the skill measures within each cohort following Fredriksson, Hensvik, and Skans (2018) and Edin, Fredriksson, Nybom, and Öckert (2022b).

Variables used to estimate the OSI. To estimate the OSI, I incorporate each of the eight cognitive and non-cognitive abilities and variables measuring experience in the 14 industries in $\boldsymbol{x}$.[12] I also include age, education level (13 distinct values), education field (25 distinct values), and region of residence (21 distinct values). These variables proxy abilities at labor market entry, human capital acquired through education, experience and age, and differences in, i.a., occupation demand across local labor markets.

Occupation routine task intensity. To distinguish between routine and non-routine occupations, I use the routine task intensity (RTI) index from Autor and Dorn (2013). It is based on the five measures of task requirements in 1980 from the U.S. dictionary of occupation titles

[9]Occupation and full-time wages refer to a reference week in September for the private sector and November for the public sector.

[10]All firms with at least 500 employees as well as the whole public sector are sampled. In smaller firms, the sampling probability is positively related to size and stratified by industry.

[11]These skill measures are described in detail by Lindqvist and Vestman (2011).

[12]Individual occupation histories are not observable as occupation information is from a survey.

(DOT) used by Autor, Levy, and Murnane (2003): Eye-hand-foot coordination (classified as manual); set limits, tolerances and standards (routine cognitive) and finger dexterity (routine manual), the average of which is routine task requirement; direction control and planning, and GED math, the average of which is abstract task requirement. The RTI for occupation k is:

$$\text{RTI}_k = \ln(\text{routine input}_k) - \ln(\text{manual input}_k) - \ln(\text{abstract input}_k). \tag{8}$$

I classify occupations as either routine or non-routine based on the RTI_k relative to the median M at the worker level. Thus, the set of non-routine occupations is $N = \{k \in K \mid \text{RTI}_k \geq M\}$.

3.2 Sample

The sample used to grow decision trees and construct the OSI consists of approximately 1.7 million observations of male employees observed in 1997-2001, who are sampled in the WSS, for which information on skills is available and who are therefore aged 23-50. 1997-2001 can be thought of as a pre-period. 2001 is chosen as the final year somewhat arbitrarily. I need a sufficiently large sample to grow the decision trees. Moreover, aggregate statistics on occupation employment are published by Statistics Sweden from this year forward.

I then follow the individual associated with each of these observation up to twelve years forward in time and collect information on wages, annual earnings, employment and occupation.[13] To obtain better coverage, future occupation and log wage from the WSS are measured in $t+10$ through $t+12$. I use the most recent observation if available or move back otherwise. As survey weights, I use the inverse of the probability of being observed in t and at least in one year between $t+10$ and $t+12$.[14]

4 Empirical analysis

This section reports the empirical results. Section 4.1 presents validation exercises. Section 4.2 describes how different characteristics relate to the OSI. Section 4.3 reports the results on how the OSI relates to routine and non-routine workers' career outcomes.

[13] In one exercise, I also follow workers five years back in time from the pre-period year of observation and collect information on previous wages.

[14] More precisely, I use the observed initial weight_t and future $\text{weight}_{t+\tau}$ and calculate the inverse of the probability of being observed in both t and at least one year between $t+10$ and $t+12$ as $\text{weight}_t \times 1/\left(1-\left(1-\frac{1}{\text{weight}_{t+\tau}}\right)^3\right)$.

4.1 Validation exercises

Explanatory power of the decision tree classifiers. The decision tree classifiers have an average out-of-sample accuracy, i.e., assigns the highest probability to the actual occupation choice, of 45.6 percent. This is marginally better than the accuracy of a multinomial logit with flexible controls for all the features (43.5 percent). These numbers can be compared to when including only a constant in the vector $\boldsymbol{x}$. The guess would then be the largest occupation in the training data set (physical and engineering associate professionals), with an accuracy of 11.3 percent. The ROC-AUC score for the prediction sample is around 87.7 when making one-versus-rest predictions for all occupations and then calculating the unweighted average.

Figure B1 in Appendix shows which explanatory variables are the most important for the ability of the decision trees to explain occupation choices. Education level and field are the most important, by a large margin, at least partly because these variables contain very detailed information. These are followed by experience in manufacturing, in construction, and several other industries. region of residence, age and the cognitive skill metrics are typically found in the middle of the feature importance distribution, while the psychological evaluation results and experience in, e.g., hotels and restaurants and other industries are located closer towards the bottom. However, all features are utilized by nearly all the decision tree classifiers. Thus, they contain important information.

The accuracy and concordance metrics do not matter per se for the OSI. The critical aspect is instead the validity of the probabilities the classifiers assign to all potential choices. This is taken care of by the use of so called honest trees. But Figure B3 in the appendix shows that there is a very strong connection between the occupation choice probabilities in the sample used to grow the tree structure and the prediction sample (the latter of which is used when constructing the OSI). This indicates that overfitting is not a large concern when the trees are grown.

The theoretical model assumes that worker characteristics are intrinsically linked to occupation-specific utility and therefore occupation choices. One way to corroborate this is to study the choices of workers that transition from their original occupation. Figure B4 shows that the assigned probabilities are highly informative of which outside occupations are the most likely destinations of long-run switchers.

Relationship between routine task intensity and employment growth. Figure 1 ranks all occupations according to their routine task intensity. This rank is then plotted against growth in the employment share between 2001 and 2013 according to Statistics Sweden. The vertical line shows the routine/non-routine cutoff. The occupations that saw the lowest employment growth are routine-intensive. In fact, no routine occupation experienced an increasing employment share and only one non-routine occupation (teaching professionals)[15] saw declining relative employment.

The figure also plots the employment share growth by occupation for my sample of workers for which I observe occupation in both t and $t + 10$ to $t + 12$. The results are quite similar. With the exception of high-skilled physical and engineering professionals and stationary plant operators, all routine occupations that experienced negative employment growth also saw size decreases in my sample of incumbent workers. Overall, routine decline cannot be accounted for only by labor market entrants and leavers.

Education requirements of routine and non-routine occupations. In SSYK, occupations with the leading number 1-3, 4-8, and 9 can be considered high-, medium- and low-skilled, respectively. As evident from Figure 1, with only one exception, routine occupations are low- to middle-skilled. To be able to better compare these to non-routine occupations, I focus on workers initially observed in low- to middle-skilled occupations in Sections 4.2 and 4.3. These are listed in Table 1. Apart from clerks and manual labourers, the routine occupations are concentrated in manufacturing. The non-routine occupations are instead mainly found in construction, services, transportation, and agriculture.[16]

4.2 Characteristics associated with specialization

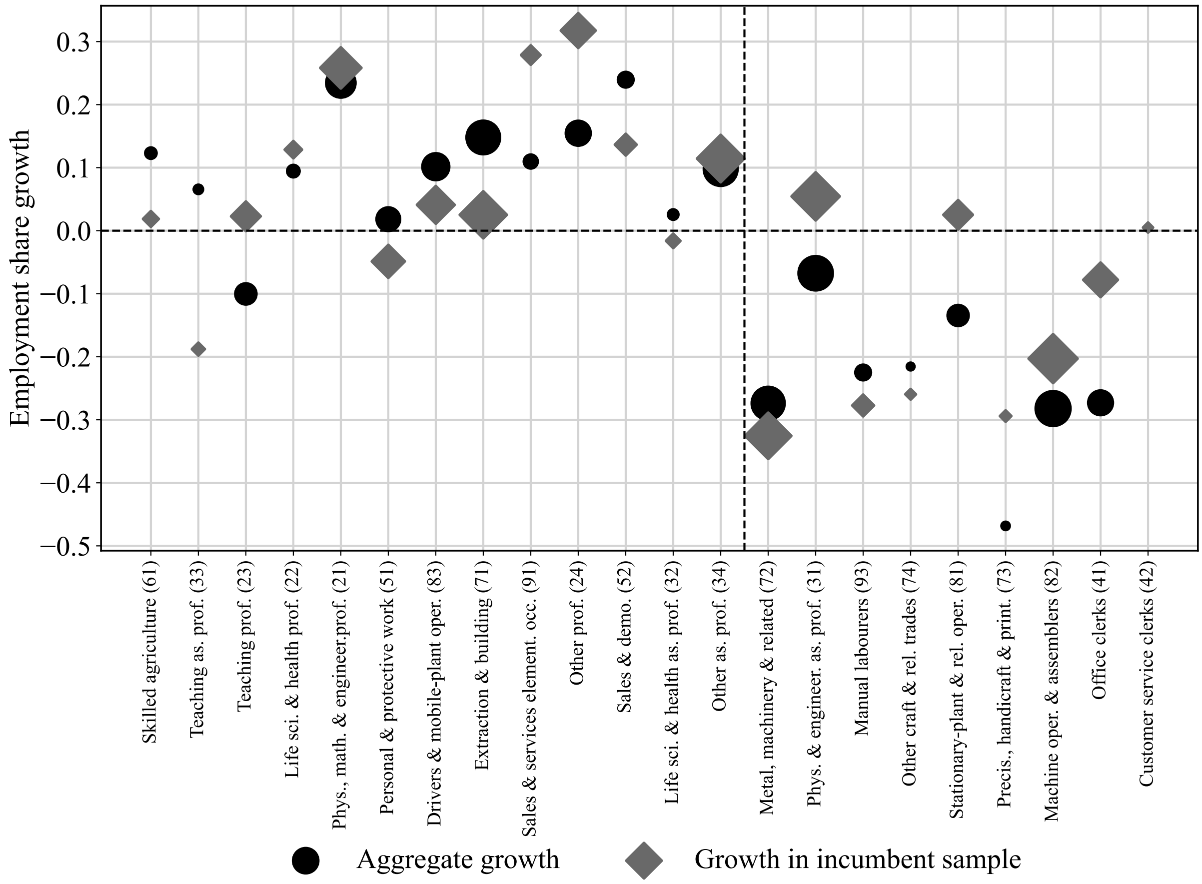
I now briefly describe which characteristics are associated with specialization. Figure 2 reports the average of the OSI by the values of the explanatory variables that are used to construct it for workers in low- to middle-skilled occupations.

For routine occupations, both cognitive and non-cognitive ability is relatively strongly negatively related to specialization. The story for educational attainment is similar. The education fields associated with the highest specialization are general (which includes primary) education and, unsurprisingly, fields related to manufacturing. Regarding regions, average OSI is relatively low in the Stockholm region and some of the other more densely populated areas. Specialization is first increasing and then decreasing somewhat in age. At the same time, experience in manufacturing and in the utilities sector are two of the strongest predic-

[15]However, associate teaching professionals exhibit a substantial increase in size, suggesting that workers' occupations may have been reclassified.

[16]Higher-skilled occupations still enter as outside occupations when calculating the OSI. In robustness checks, I show that the results are similar when including all occupations.

Figure 1: Occupation employment share growth by routine-intensity rank



Notes: The figure plots the employment share growth at the occupation level on the vertical axis. On the horizontal axis, the occupations are ranked by their routine task intensity (from low to high). The vertical dashed line represents the cutoff between routine and non-routine occupations. "Aggregate growth" refers to relative employment growth between 2001 and 2013 according to Statistics Sweden. "Growth in incumbent sample" instead refers to my sample of workers for which I observe occupation in both t (1997-2001) and $t + 10$ to $t + 12$ (depending on when workers are observed in the WSS). The size of the markers is determined by initial employment share.

Table 1: Low- to middle-skilled routine and non-routine occupations

Routine			Non-routine		
SSYK	Name	N	SSYK	Name	N
74	Other craft & related trades	4,544	61	Skilled agriculture work	13,322
42	Customer service clerks	6,800	52	Sales & demo.	22,376
73	Precision, handicraft & printing	7,099	91	Sales & services elementary occupations	31,339
93	Manual labourers	42,068	83	Drivers & mobile-plant operators	76,120
81	Stationary-plant & related operators	81,090	51	Service, care & protective work	110,831
41	Office clerks	95,091	71	Extraction & building	131,126
72	Metal, machinery & related	145,631			
82	Machine operators & assemblers	167,267			

Notes: The table reports the number of observations in the pre-period sample for all low- to middle-skilled occupations (excluding the high-skilled categories 1-3 at the broadest level of SSYK) separately by occupations below and above median routine intensity.

tors of specialization for routine workers.

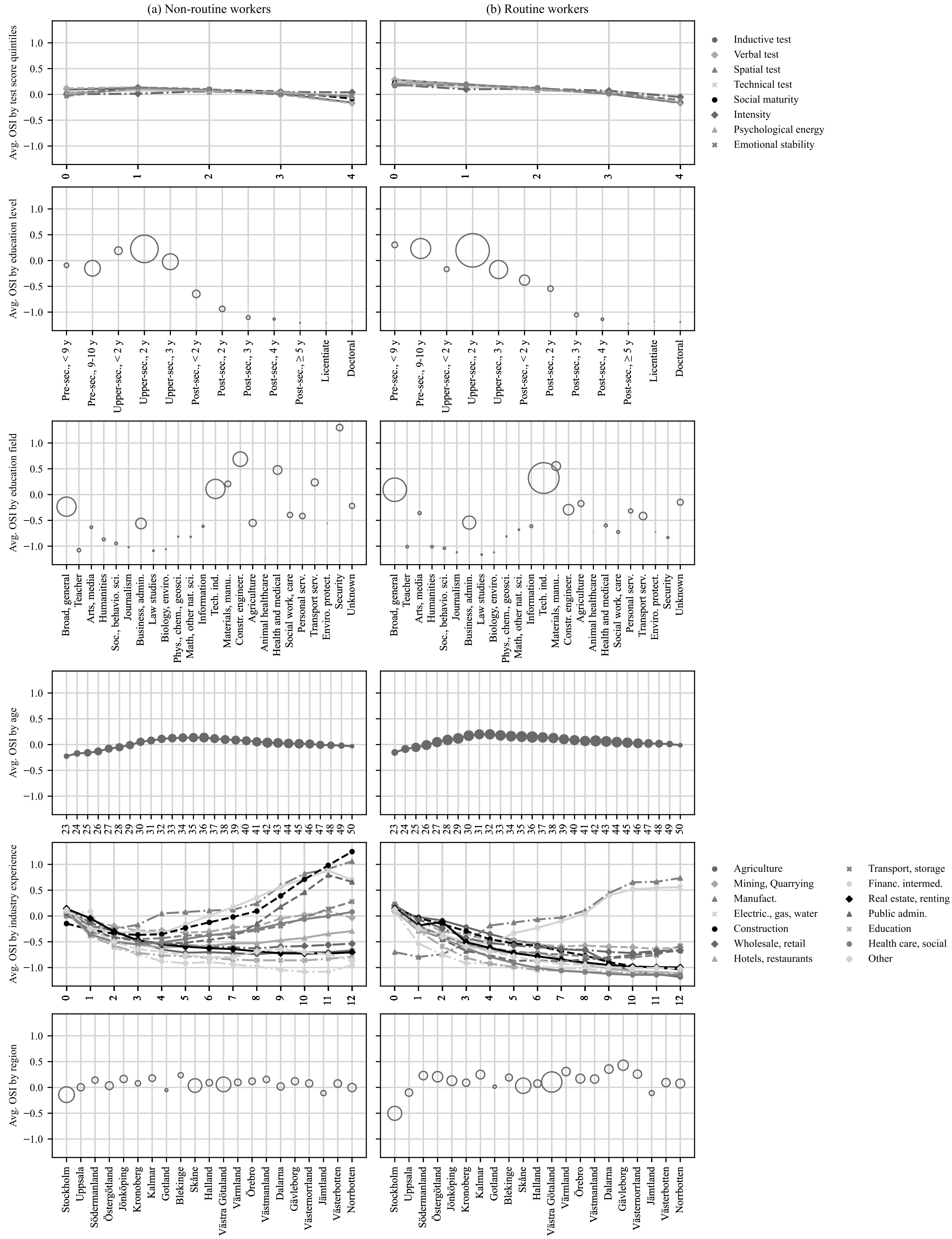
For non-routine workers, the ability relationships are not as stark as for routine occupations. But the relationship with education level and age is similar. It is worth noting, however, that low-educated workers in non-routine occupations can be considered more generalist than those in routine work. Moreover, specialization is high for workers with an education related to health care, transportation, security, and construction (as well as manufacturing). The results for region of residence are similar to, but less pronounced than, those for routine workers. Finally, workers with experience in construction, manufacturing, transportation, public administration, utilities and health care all exhibit high levels of specialization.

4.3 Specialization and career outcomes

How does the OSI relate to the future career outcomes of workers in routine and non-routine occupations? From the theory, I expect routine workers to be more likely to switch to a non-routine outside option than non-routine workers. This difference should be caused primarily by low-OSI workers. Moreover, I expect average wage/earnings growth to be lower for workers in routine than non-routine occupations. This growth penalty should be increasing in the OSI.

Non-parametric estimates. This subsection reports results from estimating versions of the semi-parametric model expressed in equation (7). The coefficients are presented in Figure 3. I

Figure 2: Relationship between characteristics and the occupation specialization index



Notes: The figure plots the average of the occupation specialization index across all bootstrap iterations by the features used to grow the decision tree classifiers. This is done separately by low- to middle-skilled routine and non-routine occupations. The size of all markers, except those showing industry-specific experience, is governed by the number of observations.

begin by relating the OSI to initial log wages separately for routine and non-routine low- to middle-skilled occupations. Two occupation outcomes are then analyzed: making any occupation transition between t and $t+10$ to $t+12$, and transitioning to a non-routine outside option. Next, I analyze three additional outcomes: log wage growth between t and $t+10$ to $t+12$, annual earnings growth in t to $t+12$ relative to the initial level, i.e., $\Delta_{t,t+12}\text{earnings}/\text{earnings}_t$, and an employment indicator for $t+12$. Earnings and wages are adjusted for CPI.

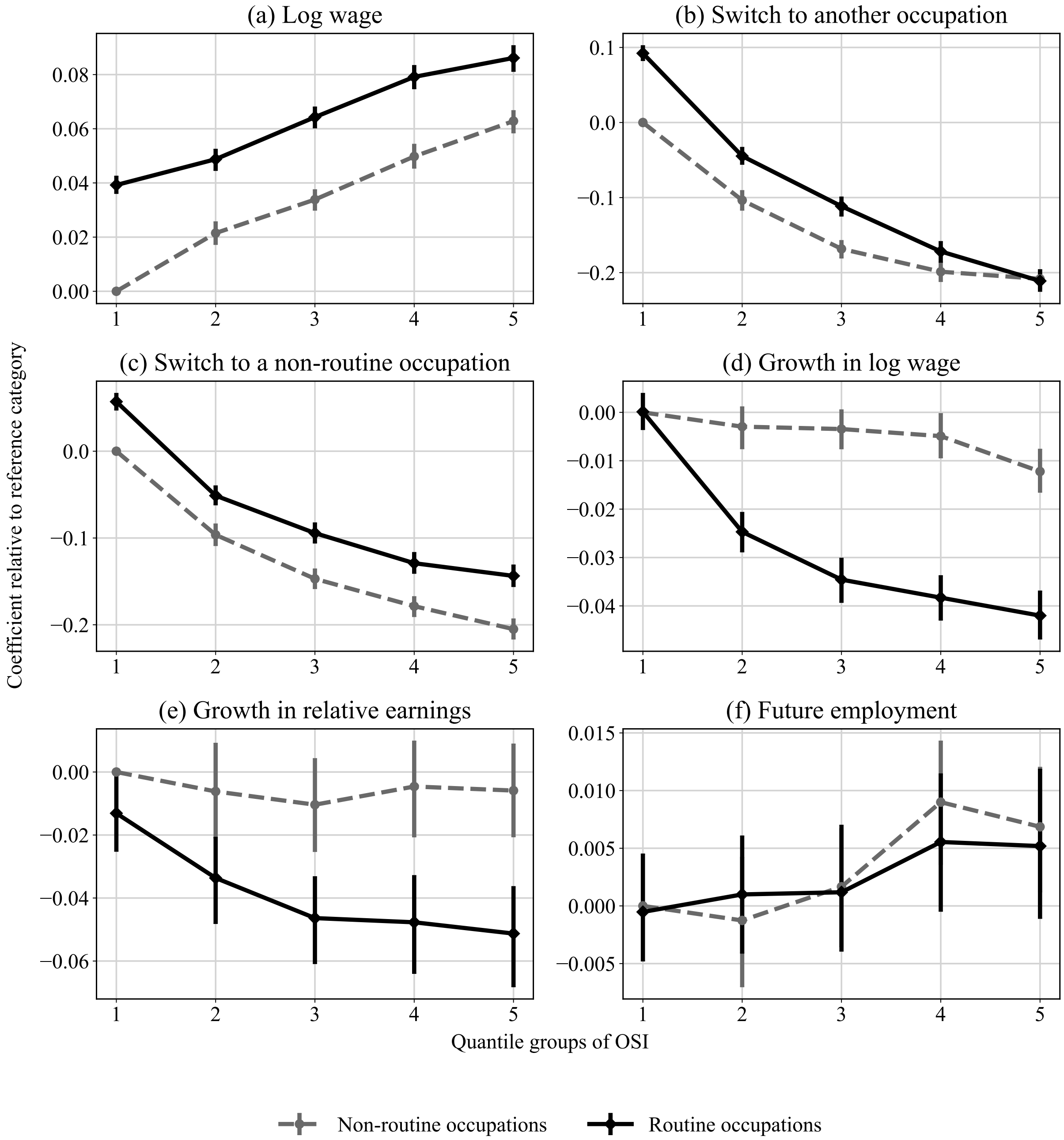
Both routine and non-routine workers have a clear, and roughly equally strong relationship between the occupation specialization index and wage level (see panel (a)). Thus, being in an occupation where your peers with similar traits are concentrated—or well-matched according to the reasoning and findings in, e.g., Fredriksson, Hensvik, and Skans (2018)—appears to pay off. Moving from the lowest to the highest specialization quintile is associated with around a 4-6 percent wage increase. Interestingly, the wage surplus appears to be higher on average in routine compared to non-routine occupations: compared to their within-group peers, and controlling for how likely you were to initially work in an outside option, routine occupations pay more.

Leaving the initial occupation is strongly negatively related to specialization for both routine and non-routine occupations (panel (b)). Moving from the lowest to the highest specialization quintile decreases the probability of leaving with about 20-30 percentage points. The patterns are similar when only considering switches to other non-routine occupations in panel (c). For reference, almost half of the least specialized workers in routine occupations left for a non-routine occupation. In (b), the relationship is somewhat stronger for routine compared to non-routine workers. But this is to a lesser extent the case in (c), which shows that routine workers have a higher probability of leaving for a non-routine outside option than non-routine workers across the specialization distribution. Thus, the OSI indeed strongly predicts which workers leave the set of routine occupations. But the hypothesis that primarily routine (and not non-routine) generalists are engaged in occupation switches is not consistently borne out.

Specialization is negatively related to wage and earnings growth (panels (d) and (e)). These relationships are markedly stronger for workers in routine compared to non-routine work. At the bottom of the OSI distribution, there is no discernible difference in wage and earnings growth between the two occupation groups. At the top of the distribution, the difference in wage (earnings) growth is around four log points (five percentage points). These results indicate that routine specialists experienced substantial negative consequences of the demand shift.[17] By contrast, routine workers with the lowest OSI levels appear to have been largely

[17]This is consistent with the findings for workers who stay in routine occupations in Cortes (2016).

Figure 3: Non-parametric estimates of specialization on labor market outcomes



Notes: The figure relates the OSI to initial log wage, an indicator for moving to another occupation between t and $t + 10$ to $t + 12$, an indicator for moving to another non-routine occupation during the same time period, log wage growth between t and $t + 10$ to $t + 12$, growth in annual earnings between t and $t + 12$ divided by initial earnings, and the probability of being classified as employed in $t+12$. The vertical axis reports the quintile group coefficients from equation (7). All regressions include controls for a second-order polynomial in the eight skill dimensions, age and 14 previous industry experience variables as well as education and field, and region fixed effects.

shielded from it. There is a slight positive relationship between the OSI and future employment for both occupation groups (panel (f)). This relationship appears marginally stronger for non-routine workers, but the difference in the effect is never significant. The absence of any substantial employment effect suggests that the quality of workers' future jobs is the important margin for future outcomes.

Alternative specifications, samples and specialization metrics. Next, I explore the robustness of my results to the choice of specification and specialization metric. To this end, I estimate different versions of the main regression model from equation (6).

The first five panels of Table 2 report results using the main OSI metric while varying the regression controls. The first model reports the raw, unadjusted relationships. The second model uses the same controls as in Figure 3, i.e., flexible controls for all variables used to estimate the OSI. Instead of controlling for the features, the third model incorporates fixed effects for the group that each observation belongs to according to the decision tree classifier. The fourth model controls for occupation FE in addition to the group FE, thus relying on variation within each occupation. The fifth model adds a second-order polynomial in initial log wage fully interacted with the routine indicator on top of the group FE. This allows for comparing workers with similar absolute productivity yet different relative occupation advantages. Table 2 finally reports results for when incorporating all observations, and thus not excluding workers in higher-skilled occupations, in a regression model with group FE.

In all models, the OSI is more strongly negatively associated with wage and earnings for workers in routine than non-routine work, as captured by the interaction term. The effects decrease somewhat when introducing occupation FE or controlling for wages, but remain statistically significant. Moreover, there are some suggestions of negative interaction effects on future employment, but they are relatively small and not consistently significant. In addition, the main effect of OSI on employment is typically positive. Finally, the OSI tends to be more negatively related to switches to outside, non-routine occupations for routine compared to non-routine workers. But this difference becomes small when controlling for the features used to construct the OSI, as discussed in relation to Figure 3.

Interacted controls. Table B1 reports estimates from models where subsets of the variables used to estimate the OSI are fully interacted with the routine indicator. The subsets are: second-order polynomials in all skill measures; region of residence FE; education level and field FE; second-order polynomials in age and industry-specific experience. This exercise reveals if the estimated effect of the OSI is mainly driven by any subset of $\boldsymbol{x}$. All models additionally include the group FE. Most coefficients are highly similar to the third model in Table 2. The exception is when controlling for interacted age and experience. The interaction effects on wage growth and earnings growth then around halves in absolute size and is no longer significant. Thus, age and experience appears to contribute disproportionately to this effect.

Table 2: Main regression results

	Occ. switch to N (1)	Log wage growth (2)	Rel. earnings growth (3)	Future employment (4)
(a) No controls				
OSI	-0.097	-0.014	0.005	0.012
	(0.002)	(0.001)	(0.002)	(0.001)
OSI × Routine	-0.011	-0.026	-0.048	-0.012
	(0.003)	(0.001)	(0.003)	(0.001)
N	640033	640033	914463	915678
R^2	0.0	0.0	0.0	0.0
(b) Feature controls				
OSI	-0.059	-0.003	0.001	0.002
	(0.002)	(0.001)	(0.003)	(0.001)
OSI × Routine	-0.002	-0.01	-0.013	-0.001
	(0.003)	(0.001)	(0.003)	(0.001)
N	640033	640033	914463	915678
R^2	0.0	0.0	0.0	0.0
(c) Group FE				
OSI	-0.124	-0.011	-0.011	0.005
	(0.004)	(0.001)	(0.004)	(0.001)
OSI × Routine	-0.073	-0.015	-0.024	-0.004
	(0.003)	(0.002)	(0.005)	(0.002)
N	640033	640033	914463	915678
R^2	0.0	0.0	0.0	0.0
(d) Occ. FE				
OSI	-0.081	-0.007	-0.013	0.0
	(0.004)	(0.002)	(0.005)	(0.002)
OSI × Routine	-0.061	-0.007	-0.02	-0.004
	(0.004)	(0.002)	(0.005)	(0.002)
N	640033	640033	914463	915678
R^2	0.0	0.0	0.0	0.0
(e) Wage controls				
OSI	-0.115	0.015	-0.002	0.002
	(0.004)	(0.001)	(0.004)	(0.001)
OSI × Routine	-0.071	-0.009	-0.021	-0.005
	(0.003)	(0.001)	(0.005)	(0.002)
N	640033	640033	914463	915678
R^2	0.0	0.0	0.0	0.0
(f) Incl. high-skilled occ.				
OSI	-0.181	-0.011	-0.012	0.004
	(0.003)	(0.001)	(0.002)	(0.001)
OSI × Routine	-0.103	-0.022	-0.031	-0.004
	(0.003)	(0.001)	(0.003)	(0.001)
N	1190555	1190555	1642884	1676070
R^2	0.0	0.0	0.0	0.0

Notes: The table reports the average coefficients and the standard deviation of the coefficients across the bootstrapped samples in parentheses from estimating equation (6) for different outcomes, sets of controls, and samples. Panel (a) reports results from regressions without additional controls. Panel (b) include controls for a second-order polynomial in the eight skill dimensions, age and 14 previous industry experience variables as well as education and field, and region fixed effects (FE). Models in (c) instead include FE for the groups from the decision tree classifiers. The models in (d) control for both group and occupation FE. The models in (e) include group FE and controls for a second-order polynomial in initial log wage fully interacted with the indicator for routine occupations. Model (f) is estimated using observations of workers in all, and not only low- to middle-skilled, occupations.

But the other characteristics still appear important. Moreover, experience is determined in part by the other characteristics. Thus, when controlling for experience, we're likely to also indirectly condition on, e.g., educational attainment and abilities.

Occupation-specific effects. To better understand what role specialization plays in each occupation, I estimate occupation-specific effects of the OSI on outcomes. These are reported in Figure B5. Without exception, the probability of switching to another non-routine occupation declines with specialization. Moreover, the OSI is typically negatively associated with both wage and earnings growth. This is especially true for routine occupations, in line with the previous results.

Trends in the returns to specialization. When in workers' careers does the effect of specialization on wage growth arise? To study this, I estimate equation (6) separately for log wage level in $t + \tau$ for all $\tau \in \{-5, \ldots, 12\}$ where t is the year when the OSI and occupation choice is measured. Figure B6 plots the main coefficients against τ. Routine workers enjoyed a high initial wage premium, which was increasing until t. This coefficient then decreases from 0.04 to 0.03 in $t + 12$. There is a positive relationship between specialization and wages: For all relative years, the common effect of the OSI is positive. Initially, the estimate is around 0.03 and increases marginally over time. The additional OSI premium of routine relative to non-routine workers is positive (around 0.01) at first, but declines to almost -0.01 in $t + 12$. Thus, being specialized in routine occupations used to yield a higher return. Over time, this effect turns into a relative penalty for workers initially observed in routine work.

5 Conclusions

In this paper, I create and estimate an index of worker occupation specialization (OSI) using detailed individual characteristics and machine learning techniques. The OSI is derived from a Roy (1951)-styled discrete choice model. Theoretically, it measures the expected difference between a worker's utility in his occupation and his best outside option. The index is simply a monotone transformation of the *ex ante* propensity for working in an outside occupation. This determines the worker's utility loss from a negative wage premium shock: Low-OSI workers with attractive non-shocked options are able to alleviate losses by moving. High-OSI specialists instead willingly remain and experience the full consequences of the shock.

The paper then analyzes to what extent the OSI can explain the consequences for incumbent employees of the falling employment in routine occupations, likely caused by shifting demand, during the period 1997–2013. I find that routine and non-routine generalists with low levels of OSI were highly mobile and did approximately equally well in terms of future earnings growth. Routine specialists instead by and large remained in routine work despite the overall employment decline in these occupations. They also experienced significantly lower earnings growth than both generalists and non-routine specialists.

These findings are broadly consistent with the predictions from the Roy (1951)-styled discrete choice model from which the specialization index is derived. Overall, the results indicate that the Roy (1951) model can characterize which workers lose from negative demand shifts. Moreover, the observable worker attributes at hand can be used to infer how dependent workers are on their current occupations.

Exploring the consequences of the historical decline in routine work is important in its own right. But the ability of the OSI to predict which workers experienced negative consequences from this shift also substantiates the general usefulness of the index for describing worker susceptibility. Currently, policy makers have very few tools at hand for forecasting individual consequences of future shifts in occupation demand. The OSI is solely based on current information. It could therefore be used to characterize workers employed in occupations today that we believe will experience negative demand shifts in the future.

References

Adermon, A., & Gustavsson, M. (2015). Job polarization and task-biased technological change: Evidence from sweden, 1975–2005. *The Scandinavian Journal of Economics*, *117*(3), 878–917.

Arntz, M., Gregory, T., & Zierahn, U. (2017). Revisiting the risk of automation. *Economics Letters*, *159*, 157–160.

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360.

Autor, D., & Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American Economic Review*, *103*(5), 1553–97.

Autor, D., Katz, L., & Kearney, M. (2008). Trends in u.s. wage inequality: Revising the revisionists. *The Review of Economics and Statistics*, *90*(2), 300–323.

Autor, D., Levy, F., & Murnane, R. (2003). The skill content of recent technological change: An empirical exploration*. *The Quarterly Journal of Economics*, *118*(4), 1279–1333.

Bachmann, R., Cim, M., & Green, C. (2019). Long-run patterns of labour market polarization: Evidence from german micro data. *British Journal of Industrial Relations*, *57*(2), 350–376.

Balassa, B. (1965). Trade liberalisation and "revealed" comparative advantage 1. *The manchester school*, *33*(2), 99–123.

Böhm, M. (2020). The price of polarization: Estimating task prices under routine-biased technical change. *Quantitative Economics*, *11*(2), 761–799.

Böhm, M., Gaudecker, H.-M. v., & Schran, F. (2023). Occupation growth, skill prices, and wage inequality. *Journal of Labor Economics*, forthcoming.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Caldwell, S., & Danieli, O. (2022). Outside options in the labor market. *Mimeo*.

Cavaglia, C., & Etheridge, B. (2020). Job polarization and the declining quality of knowledge workers: Evidence from the uk and germany. *Labour Economics*, *66*, 101884.

Coraggio, L., Pagano, M., Scognamiglio, A., & Tåg, J. (2025). Jaq of all trades: Job mismatch, firm productivity and managerial quality. *Journal of Financial Economics*, *164*, 103992.

Cortes, G. M. (2016). Where have the middle-wage workers gone? a study of polarization using panel data. *Journal of Labor Economics*, *34*(1), 63–105.

Cortes, G. M., & Gallipoli, G. (2018). The costs of occupational mobility: An aggregate analysis. *Journal of the European Economic Association*, *16*(2), 275–315.

Cortes, G. M., Jaimovich, N., Nekarda, C. J., & Siu, H. E. (2020). The dynamics of disappearing routine jobs: A flows approach. *Labour Economics*, *65*, 101823.

Cortes, G. M., Jaimovich, N., & Siu, H. E. (2017). Disappearing routine jobs: Who, how, and why? *Journal of Monetary Economics*, *91*, 69–87.

De Jong, G., Daly, A., Pieters, M., & Van der Hoorn, T. (2007). The logsum as an evaluation measure: Review of the literature and new results. *Transportation Research Part A: Policy and Practice*, *41*(9), 874–889.

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

Edin, P.-A., Evans, T., Graetz, G., Hernnäs, S., & Michaels, G. (2022a). Individual consequences of occupational decline. *Mimeo*.

Edin, P.-A., Fredriksson, P., Nybom, M., & Öckert, B. (2022b). The rising return to noncognitive skill. *American Economic Journal: Applied Economics*, *14*(2), 78–100.

Eggenberger, C., Janssen, S., & Backes-Gellner, U. (2022). The value of specific skills under shock: High risks and high returns. *Labour Economics*, 102187.

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

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280.

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

Goos, M., Arntz, M., Zierahn, U., Gregory, T., Gomez, S. C., Vázquez, I. G., & Jonkers, K. (2019). *The impact of technological innovation on the future of work* (tech. rep.). JRC Working Papers Series on Labour, Education and Technology.

Goos, M., & Manning, A. (2007). Lousy and lovely jobs: The rising polarization of work in britain. *The Review of Economics and Statistics*, *89*(1), 118–133.

Goos, M., Manning, A., & Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American Economic Review*, *104*(8), 2509–26.

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

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

Hanemann, W. M. (1984). Discrete/continuous models of consumer demand. *Econometrica: Journal of the Econometric Society*, 541–561.

Jaimovich, N., Saporta-Eksten, I., Siu, H., & Yedid-Levi, Y. (2021). The macroeconomics of automation: Data, theory, and policy analysis. *Journal of Monetary Economics*, *122*, 1–16.

Jovanovic, B., & Nyarko, Y. (1997). Stepping-stone mobility. *Carnegie-Rochester Conference Series on Public Policy*, *46*, 289–325.

Lindqvist, E., & Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *American Economic Journal: Applied Economics*, *3*(1), 101–28.

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

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

Mokyr, J., Vickers, C., & Ziebarth, N. L. (2015). The history of technological anxiety and the future of economic growth: Is this time different? *Journal of Economic Perspectives*, *29*(3), 31–50.

Mood, C., Jonsson, J. O., & Bihagen, E. (2012). Socioeconomic persistence across generations : Cognitive and noncognitive processes. *From parents to children: The intergenerational transmission of advantage* (pp. 53–84).

OECD. (2019). *Oecd employment outlook 2019: The future of work*. Organisation for Economic Co-operation; Development OECD.

Robinson, C. (2018). Occupational mobility, occupation distance, and specific human capital. *Journal of Human Resources*, *53*(2), 513–551.

Ross, S. L., & Ukil, P. (2021). Initial industry and long-term earnings growth. *AEA Papers and Proceedings*, *111*, 476–80.

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

Roys, N. A., & Taber, C. R. (2022). Skill prices, occupations, and changes in the wage structure for low skilled men. *Mimeo*.

Small, K. A., & Rosen, H. S. (1981). Applied welfare economics with discrete choice models. *Econometrica: Journal of the Econometric Society*, 105–130.

Traiberman, S. (2019). Occupations and import competition: Evidence from denmark. *American Economic Review*, *109*(12), 4260–4301.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Webb, M. (2019). The impact of artificial intelligence on the labor market. *Mimeo*.

A Derivations

This section shows the full derivation of the results discussed in Section 2.1 and 2.2. I begin by discussing some useful properties of the Gumbel distribution. I then characterize the distribution of the utility surplus in equation (2). Using these results, I proceed to derive the occupation specialization index. Finally, I derive a closed-form expression for the expected value of the utility loss following the wage premium in equation (2).

A1 Properties of the Gumbel distribution

All ε_{ik} are assumed to be IID standard extreme value type I (or Gumbel) distributed. The occupation choice probabilities then follow the multinomial logit formula:

$$p_k(\boldsymbol{x}) = \frac{e^{u_k(\boldsymbol{x})}}{\sum\limits_{l \in K} e^{u_l(\boldsymbol{x})}}. \tag{1}$$

Moreover, the maximum utility of a worker (see equation (1)) before conditioning on occupation choice is distributed as:

$$\max_{k \in K} u_{ik} \;\sim\; \text{Gumbel}\left(\ln\left(\sum_{k \in K} e^{u_k(\boldsymbol{x})}\right), 1\right). \tag{2}$$

The location parameter in this equation is commonly referred to as the log-sum. The expected utility takes the form $\mathbb{E}\Big[\max_{k \in K} u_{ik}\Big] = \ln\Big(\sum_{k \in K} e^{u_k(\boldsymbol{x})}\Big) + C$, where C is an unknown constant reflecting that absolute utility cannot be measured. See, e.g., Small and Rosen (1981), Train (2009) for a textbook treatment, and De Jong et al. (2007) for a literature review of applications of this expected value.

Using results from Hanemann (1984),[18] I show that maximum utility can be shown to be independent of the chosen occupation j. More specifically, Hanemann (1984) shows that the idiosyncratic term of the best choice j is distributed as:

$$\varepsilon_{ij} \mid \arg\max_{k \in K} u_{ik} = j \sim \text{Gumbel}\left(\ln\left(\sum_{k \in K} e^{u_k(\boldsymbol{x})}/e^{u_j(\boldsymbol{x})}\right), 1\right). \tag{3}$$

[18] His framework focuses on consumption and incorporates heterogeneous goods prices and a budget constraint.

Next, let $F(.)$ be the CDF of this distribution and s be some value. Then:

$$\begin{aligned}
\Pr(u_{ij} < s) &= F(s - u_j(\boldsymbol{x})) \\
&= \exp\left\{-\exp\left\{-(s - u_j(\boldsymbol{x})) + \ln\left(\sum_{k\in K} e^{u_k(\boldsymbol{x})}/e^{u_j(\boldsymbol{x})}\right)\right\}\right\} \\
&= \exp\left\{-\exp\left\{-s + \ln\left(\sum_{k\in K} e^{u_k(\boldsymbol{x})}\right)\right\}\right\}.
\end{aligned} \tag{4}$$

This is the CDF associated with the distribution in (2). Thus, the maximum utility conditional on any optimal choice j is distributed the same as the unconditional maximum in (2).

A final useful property of the Gumbel distribution is that the difference between two independent Gumbel distributed variables with location parameters a, b and common scale parameter c is known to follow a Logistic$(a-b,\ c)$ distribution.

A2 Characterizing the distribution of the utility surplus

Finding the distribution of the utility surplus in (2) is done in four steps.

Step 1: Rewriting the utility surplus. I begin by defining $P(s)$ as the probability that the utility surplus in (2) conditional on worker characteristics $\boldsymbol{x}$ and occupation choice j is smaller than the value s. This represents the CDF of the utility surplus for which I want to obtain an closed-form expression:

$$P(s) \equiv \Pr\left(u_{ij} - \max_{n\in N\setminus\{j\}} u_{in} \leq s \,\middle|\, \boldsymbol{x},\ \arg\max_{k\in K} u_{ik} = j\right). \tag{5}$$

Next, K can be partitioned into two subsets: One including R and a worker's chosen occupation j, and one with his non-routine outside options. The probability may then be rewritten in terms of the difference between the maximum utility in each set conditional on the first maximum being larger, and on the choice in the first set being j:

$$P(s) = \Pr\left(\max_{r\in R\cup\{j\}} u_{ir} - \max_{n\in N\setminus\{j\}} u_{in} \leq s \,\middle|\, \boldsymbol{x},\ \max_{r\in R\cup\{j\}} u_{ir} - \max_{n\in N\setminus\{j\}} u_{in} \geq 0,\ \arg\max_{r\in R\cup\{j\}} u_{ir} = j\right). \tag{6}$$

Thus, I recast the maximization problem as a two-step problem where the worker finds the best local options in the two subsets, and then compares them to each other. Notice here that j changes interpretation from the best option in K to the best option in the set $R\cup\{j\}$.

Step 2: Removing the argmax condition from $P(s)$. Both $\max_{r\in R\cup\{j\}} u_{ir}$ and $\max_{n\in N\setminus\{j\}} u_{in}$ are Gumbel distributed according to (2). According to (4), $\max_{r\in R\cup\{j\}} u_{ir}$ does not depend on arg $\max_{r\in R\cup\{j\}} u_{ir} = j$. Conditional on the value of $\max_{r\in R\cup\{j\}} u_{ir}$, $\max_{n\in N\setminus\{j\}} u_{in}$ must also be independent of the argmax condition. Hence, this condition can be removed from the set of conditions in (6):

$$P(s) = \Pr\left(\max_{r\in R\cup\{j\}} u_{ir} - \max_{n\in N\setminus\{j\}} u_{in} \,\middle|\, \boldsymbol{x},\ \max_{r\in R\cup\{j\}} u_{ir} - \max_{n\in N\setminus\{j\}} u_{in} \geq 0\right). \tag{7}$$

One may now think of j as a regular occupation rather than as the optimal choice, although it still represents the observed occupation of worker i. Finally, define v_{ij} as the difference between the two maxima in (7):

$$v_{ij} \equiv \max_{r\in R\cup\{j\}} u_{ir} - \max_{n\in N\setminus\{j\}} u_{in}. \tag{8}$$

One may then write:

$$P(s) = \Pr\left(v_{ij} \leq s \,\middle|\, \boldsymbol{x}, v_{ij} \geq 0\right). \tag{9}$$

Step 3: Determining the distribution of v_{ij}. By (2), $\max_{r\in R\cup\{j\}} u_{ir}$ and $\max_{n\in N\setminus\{j\}} u_{in}$ are both gumbel with scale one. Since by (7) one may now think of j as a regular occupation, they must also be independent. Hence, by the final property of the Gumbel distribution stated in Section A.1, v_{ij} is Logistic distributed with scale one.

From equation (2), one can also infer the location parameter μ of v_{ij}. It equals the difference between the location parameters of the two Gumbels. Using the multinomial logit formula from (1), μ can then readily be rewritten as a function of only choice probabilities. Finally, recall that equation (3) defines $\rho(\boldsymbol{x}, j, N)$ as the *ex ante* probability of working in a non-routine outside option, i.e.,

$$\rho(\boldsymbol{x}, j, N) \equiv \sum_{n\in N\setminus\{j\}} p_n(\boldsymbol{x}).$$

The location parameter of the distribution of v_{ij} is:

$$\begin{aligned}\mu = \mu(\boldsymbol{x}, j, N) &= \ln\left(\sum_{r\in R\cup\{j\}} e^{u_r(\boldsymbol{x})}\right) - \ln\left(\sum_{n\in N\setminus\{j\}} e^{u_n(\boldsymbol{x})}\right)\\ &= \ln\left(\sum_{r\in R\cup\{j\}} e^{u_r(\boldsymbol{x})} \Big/ \sum_{n\in N\setminus\{j\}} e^{u_n(\boldsymbol{x})}\right)\end{aligned}$$

$$
\begin{aligned}
&= \ln\left(\frac{\sum\limits_{r\in R\cup\{j\}} e^{u_r(\boldsymbol{x})}}{\sum\limits_{k\in K} e^{u_k(\boldsymbol{x})}} \Big/ \frac{\sum\limits_{n\in N\setminus\{j\}} e^{u_n(\boldsymbol{x})}}{\sum\limits_{k\in K} e^{u_k(\boldsymbol{x})}} \right) \\
&= \ln\left(\sum_{r\in R\cup\{j\}} p_r(\boldsymbol{x}) \Big/ \sum_{n\in N\setminus\{j\}} p_n(\boldsymbol{x}) \right) \\
&= \ln\left(\frac{1}{\sum\limits_{n\in N\setminus\{j\}} p_n(\boldsymbol{x})} - 1 \right) \\
&= \ln\left(\frac{1}{\rho(\boldsymbol{x}, j, N)} - 1 \right).
\end{aligned} \tag{10}
$$

Step 4: Finding the distribution function of the utility surplus. Next, I turn to finding an expression for the CDF and PDF of the utility surplus. First, denote the PDF and CDF of the Logistic$(\mu(\boldsymbol{x}, j, N), 1)$ distribution by $f(.)$ and $F(.)$, respectively. These are:

$$
\begin{aligned}
f(s \mid \boldsymbol{x}, j, N) &= \frac{e^{-s+\mu(\boldsymbol{x},j,N)}}{\left(1 + e^{-s+\mu(\boldsymbol{x},j,N)}\right)^2} \\
F(s \mid \boldsymbol{x}, j, N) &= \frac{1}{1 + e^{-s+\mu(\boldsymbol{x},j,N)}}
\end{aligned} \tag{11}
$$

Next, by properties of conditional probabilities, one can write the CDF of the utility surplus in terms of $F(.)$:

$$
\begin{aligned}
P(s) &= \Pr\left(v_{ij} \le s \,\middle|\, \boldsymbol{x}, v_{ij} \ge 0\right) \\
&= \frac{\Pr\left(0 \le v_{ij} \le s \,\middle|\, \boldsymbol{x}\right)}{1 - \Pr\left(v_{ij} \le 0 \,\middle|\, \boldsymbol{x}\right)} \\
&= \begin{cases} \frac{F(s \mid \boldsymbol{x},j,N) - F(0 \mid \boldsymbol{x},j,N)}{1 - F(0 \mid \boldsymbol{x},j,N)} \text{ for } s \ge 0 \\ 0 \text{ otherwise} \end{cases}
\end{aligned} \tag{12}
$$

Finally, to obtain the PDF of the utility surplus, differentiate (12) with respect to s:

$$
P'(s) = \begin{cases} \frac{f(s \mid \boldsymbol{x},j,N)}{1 - F(0 \mid \boldsymbol{x},j,N)} \text{ for } s \ge 0 \\ 0 \text{ otherwise} \end{cases} \tag{13}
$$

A3 Deriving the occupation specialization index

Let S represent the utility surplus. Using the PDF of the logistic distribution from (11) and its relationship with the utility surplus PDF in (13), one can obtain a closed-form solution to the expected value of the surplus. Finally, plugging in the location parameter from (10) and simplifying gives the OSI:

$$\begin{aligned}
\text{OSI}(\boldsymbol{x}, j, N) &\equiv \mathbb{E}\left[u_{ij} - \max_{n\in N\setminus\{j\}} u_{in} \;\middle|\; \boldsymbol{x},\ \arg\max_{k\in K} u_{ik} = j\right] \\
&= \left(1 - \frac{1}{1+e^{\mu(\boldsymbol{x},j,N)}}\right)^{-1} \int\limits_0^\infty S \frac{e^{-S+\mu(\boldsymbol{x},j,N)}}{\left(1+e^{-S+\mu(\boldsymbol{x},j,N)}\right)^2} dS \\
&= \frac{1+e^{\mu(\boldsymbol{x},j,N)}}{e^{\mu(\boldsymbol{x},j,N)}} \ln\left(1+e^{\mu(\boldsymbol{x},j,N)}\right) \\
&= -\frac{\ln\left(\rho(\boldsymbol{x}, j, N)\right)}{1-\rho(\boldsymbol{x}, j, N)} \qquad (14)
\end{aligned}$$

Both $\ln\left(\rho(\boldsymbol{x}, j, N)\right)$ and $1/(1-\rho(\boldsymbol{x}, j, N))$ are increasing in $\rho(\boldsymbol{x}, j, N)$. Therefore, the OSI is a monotonically decreasing function of $\rho(\boldsymbol{x}, j, N)$.

A4 Deriving the expected utility loss

The expected value of the utility change following the wage premium shock in (2) can be written as a probability-weighted average of d_j and a conditional expected value of S. A closed-form solution can then be derived in a similar way as in (14). Again, let S represents the utility surplus random variable. Then:

$$\begin{aligned}
\mathbb{E}\left[\Delta u_{ij} \;\middle|\; \boldsymbol{x},\ j\right] &= \mathbb{E}\left[\max\left\{-d_j, -\left(u_{ij} - \max_{n\in N\setminus\{j\}} u_{in}\right)\right\} \;\middle|\; \boldsymbol{x},\ \arg\max_{k\in K} u_{ik} = j\right] \\
&= -d_j - Pr(S \le d_j)\left(\mathbb{E}\Big[S \mid S \le d_j\Big] - d_j\right) \\
&= -\frac{\left(1 - \frac{1}{1+e^{-d_j+\mu(\boldsymbol{x},j,N)}}\right) d_j + \int\limits_0^{d_j} S \frac{e^{-S+\mu(\boldsymbol{x},j,N)}}{\left(1+e^{-S+\mu(\boldsymbol{x},j,N)}\right)^2} dV}{1 - \frac{1}{1+e^{\mu(\boldsymbol{x},j,N)}}} \\
&= -\frac{d_j - \ln\left(1 + \left(e^{d_j} - 1\right)\rho(\boldsymbol{x}, j, N)\right)}{1-\rho(\boldsymbol{x}, j, N)}. \qquad (15)
\end{aligned}$$

For any value of $d_j = \delta > 0$, (15) is monotonically increasing (i.e., the absolute loss becomes smaller) in $\rho(\boldsymbol{x}, j, N)$. To see this, differentiate (15) with respect to ρ which is used as shorthand for $\rho(\boldsymbol{x}, j, N)$:

$$\begin{aligned}
\frac{\partial \mathbb{E}\left[\Delta u_{ij} \mid \boldsymbol{x},\ j\right]}{\partial \rho} &= -\frac{d_j - \ln\left(1 + \left(e^{d_j} - 1\right)\rho\right)}{(1-p)^2} + \frac{\frac{e^{d_j}-1}{1+\left(e^{d_j}-1\right)\rho}}{1-\rho} \\
&= \frac{1}{(1-p)^2}\left[-\left(d_j - \ln\left(1 + \left(e^{d_j} - 1\right)\rho\right)\right) + \frac{(1-\rho)\left(e^{d_j}-1\right)}{1+\left(e^{d_j}-1\right)\rho}\right] \\
&= \frac{1}{(1-p)^2}\left[-\left(\ln\left(e^{d_j}\right) - \ln\left(1 + \left(e^{d_j} - 1\right)\rho\right)\right) + \frac{e^{d_j} - \left(1+\left(e^{d_j}-1\right)\rho\right)}{1+\left(e^{d_j}-1\right)\rho}\right]
\end{aligned}$$

$$= \frac{1}{(1-p)^2}\left[-\ln\left(\frac{e^{d_j}}{1+(e^{d_j}-1)\,\rho}\right)+\frac{e^{d_j}}{1+(e^{d_j}-1)\,\rho}-1\right] \tag{16}$$

This may be rewritten as:

$$\frac{\partial \mathbb{E}\left[\Delta u_{ij} \,\middle|\, \boldsymbol{x},\ j\right]}{\partial \rho} = a\Big[b-1-\ln(b)\Big],$$
$$\text{where}\ \ a=\frac{1}{(1-\rho)^2}\ \ \text{and}\ \ b=\frac{e^{d_j}}{(1+(e^{d_j}-1)\rho)}. \tag{17}$$

$a > 1$ for any $0 < \rho < 1$. For routine workers with $d_j = \delta > 0$, $e^{d_j} \geq 1 + (e^{d_j} - 1)\rho > 1$. This implies that $b > 1$. In turn, $b - 1 - \ln(b) > 0$. Thus, the derivative is positive, implying monotonicity.

A5 Simulating the model

To verify the the results derived above, I have also simulated the model in Section 2 for 10 routine and 10 non-routine occupations. I generate 10 million workers i with individual standard Gumbel draws for each occupation. All workers belong to one of 100 equally-sized groups $g \in \{1, \ldots, 100\}$. Each group draws a deterministic utility term u_{gk} for each occupation k from the standard normal distribution. The wage premium shock is set to $\delta = 1$. For each worker, I find $\max_{k\in K}\{u_{ik}\} - \max_{n\in N\setminus\{j\}}\{u_{in}\}$. Next, I calculate group-specific probabilities as the number of workers in g in k (N_{gk}) divided by the total size of g (N_g); $p_{gk} = N_{gk}/N_g$. Workers are classified as either routine or non-routine depending on whether $\arg\max_{k\in K}\{u_{ik}\} \in R$. To construct the OSI, I use the group-specific probabilities. Finally, I calculate $\max\Big\{-d_j, -\Big(\max_{k\in K}\{u_{ik}\} - \max_{n\in N\setminus\{j\}}\{u_{in}\}\Big)\Big\}$ and $\mathbb{1}\Big[\Big(\max_{k\in K}\{u_{ik}\} - \max_{n\in N\setminus\{j\}}\{u_{in}\}\Big) < d_j\Big]$.

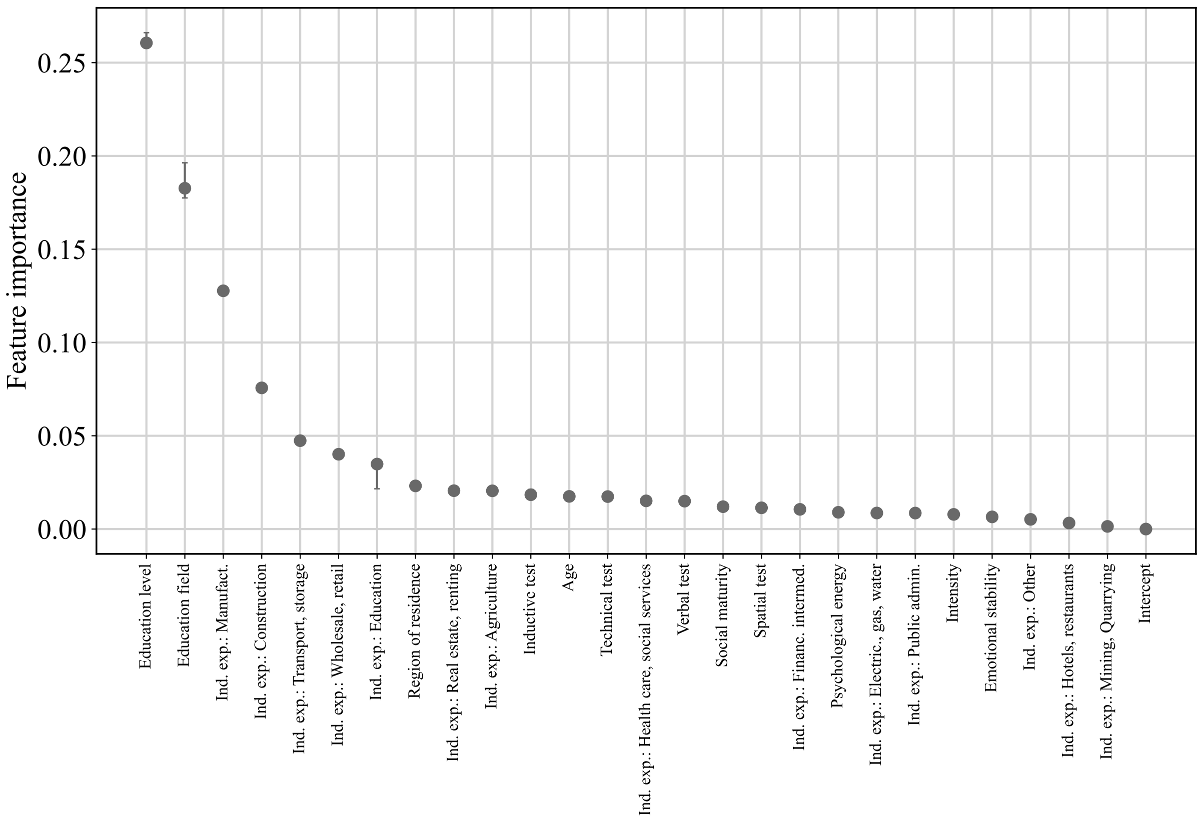
B Additional empirical results

Table B1: Regression results with interacted controls

	Occ. switch to N (1)	Log wage growth (2)	Rel. earnings growth (3)	Future employment (4)
(a) Skills				
OSI	-0.124	-0.011	-0.012	0.004
	(0.004)	(0.001)	(0.004)	(0.001)
OSI $\times$ Routine	-0.073	-0.015	-0.021	-0.003
	(0.004)	(0.002)	(0.005)	(0.002)
N	640033	640033	914463	915678
R^2	0.0	0.0	0.0	0.0
(b) Education				
OSI	-0.131	-0.013	-0.012	0.005
	(0.004)	(0.001)	(0.005)	(0.002)
OSI $\times$ Routine	-0.067	-0.017	-0.027	-0.004
	(0.004)	(0.002)	(0.005)	(0.002)
N	640033	640033	914463	915678
R^2	0.0	0.0	0.0	0.0
(c) Region				
OSI	-0.126	-0.011	-0.012	0.005
	(0.004)	(0.001)	(0.004)	(0.001)
OSI $\times$ Routine	-0.067	-0.015	-0.02	-0.003
	(0.004)	(0.002)	(0.005)	(0.002)
N	640033	640033	914463	915678
R^2	0.0	0.0	0.0	0.0
(d) Age and industry experience				
OSI	-0.131	-0.009	-0.006	0.005
	(0.004)	(0.002)	(0.005)	(0.002)
OSI $\times$ Routine	-0.033	-0.006	-0.009	0.008
	(0.008)	(0.004)	(0.01)	(0.004)
N	640033	640033	914463	915678
R^2	0.0	0.0	0.0	0.0

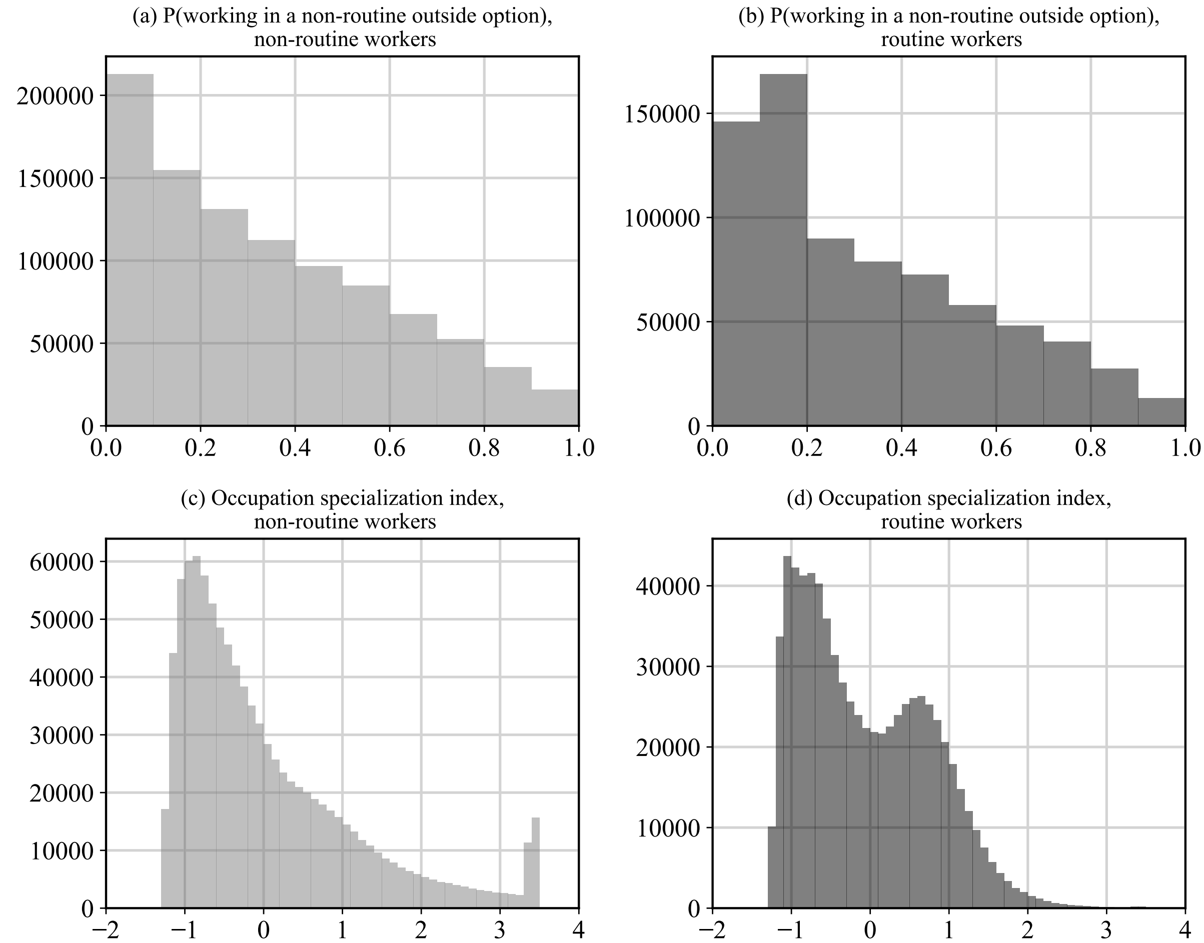
Notes: The table reports the average coefficients and the standard deviation of the coefficients across the bootstrapped samples in parentheses from estimating equation (6) for sets of controls. All models include fixed effects (FE) for the groups from the decision tree classifiers. Each panel reports results from models which interact the routine indicator with one of four sets of control variables: A second-order polynomial in all skill variables (panel a); education level and field FE (panel b); region of residence FE (panel c); and a second-order polynomial in the previous industry experience variables as well as age (panel d).

Figure B1: Feature importance in the decision tree classifiers



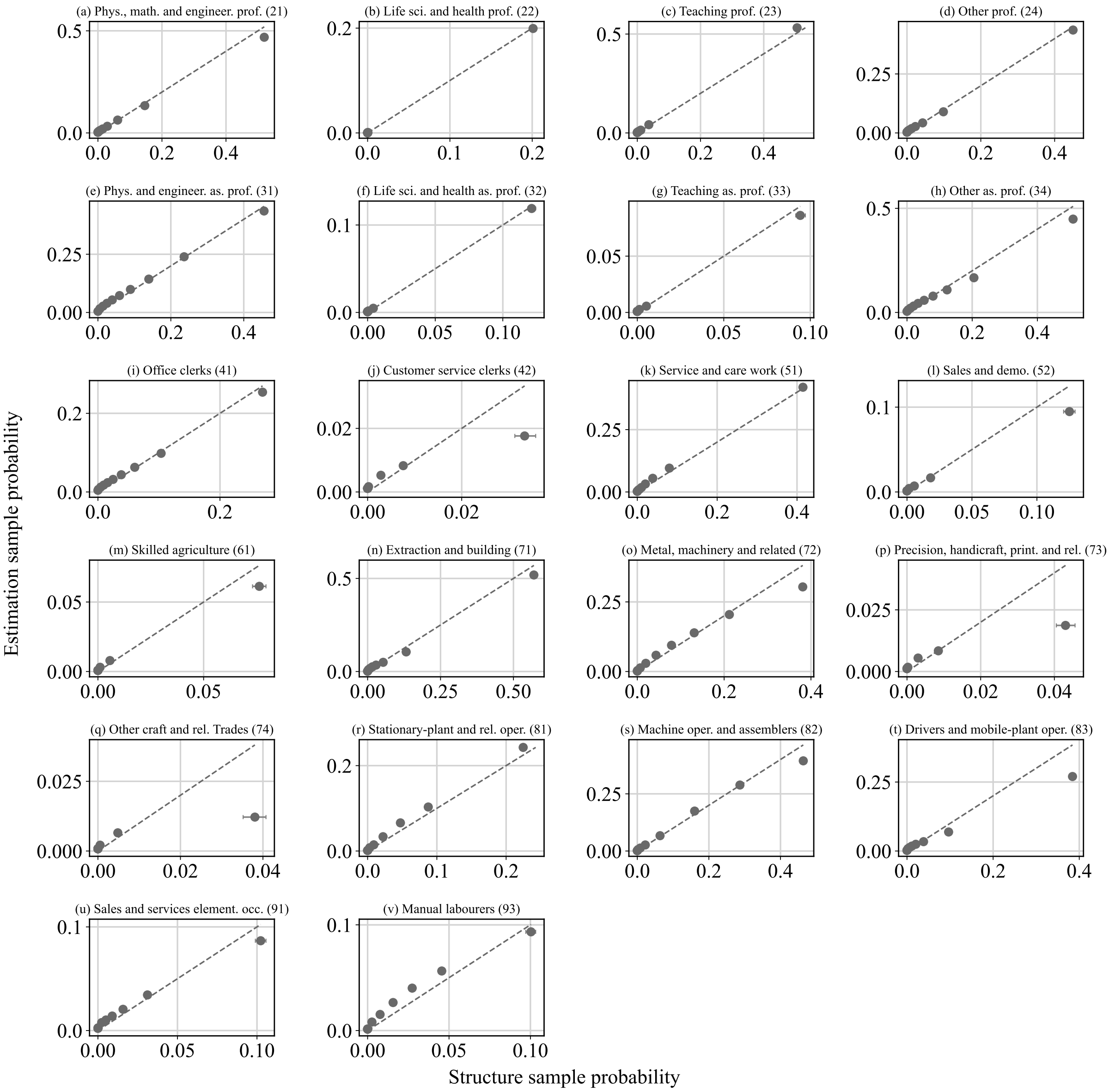
Notes: The importance of each feature/variable, or Gini importance, is defined as the normalized total reduction of the criterion due to splits on that particular variable. The markers report the mean while the error bars report the 5th and 95th percentile across all bootstrapped decision trees.

Figure B2: Histograms of the occupation specialization indices



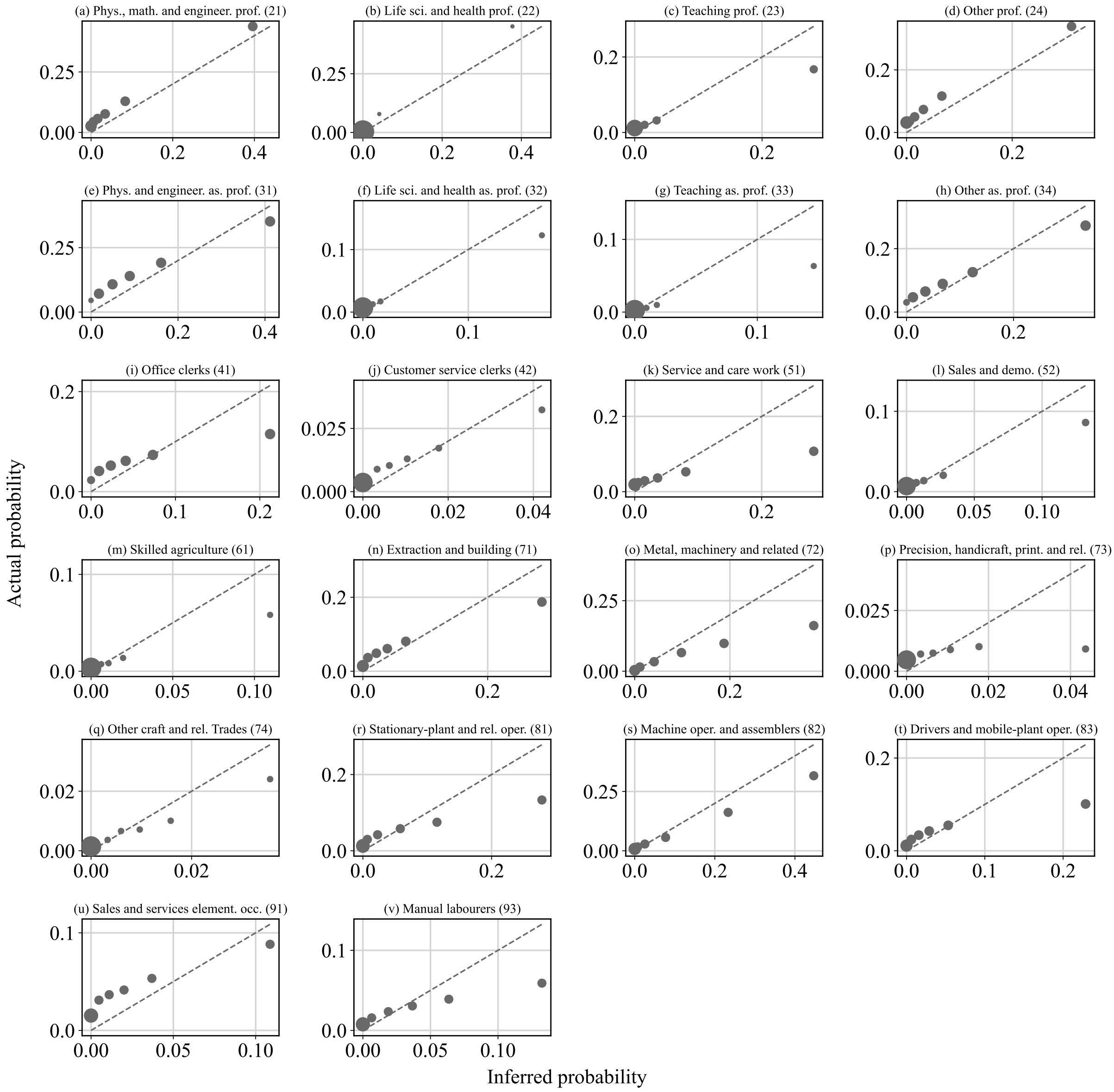
Notes: The figure reports histograms of the estimates of $\rho(\boldsymbol{x}, j, N)$, i.e., the *ex ante* probability of working in a non-routine outside option, and $\mathrm{OSI}(\boldsymbol{x}, j, N)$, i.e., the occupation specialization index, separately for routine and non-routine workers.

Figure B3: Relationship between occupation choice probabilities in structure and estimation sample



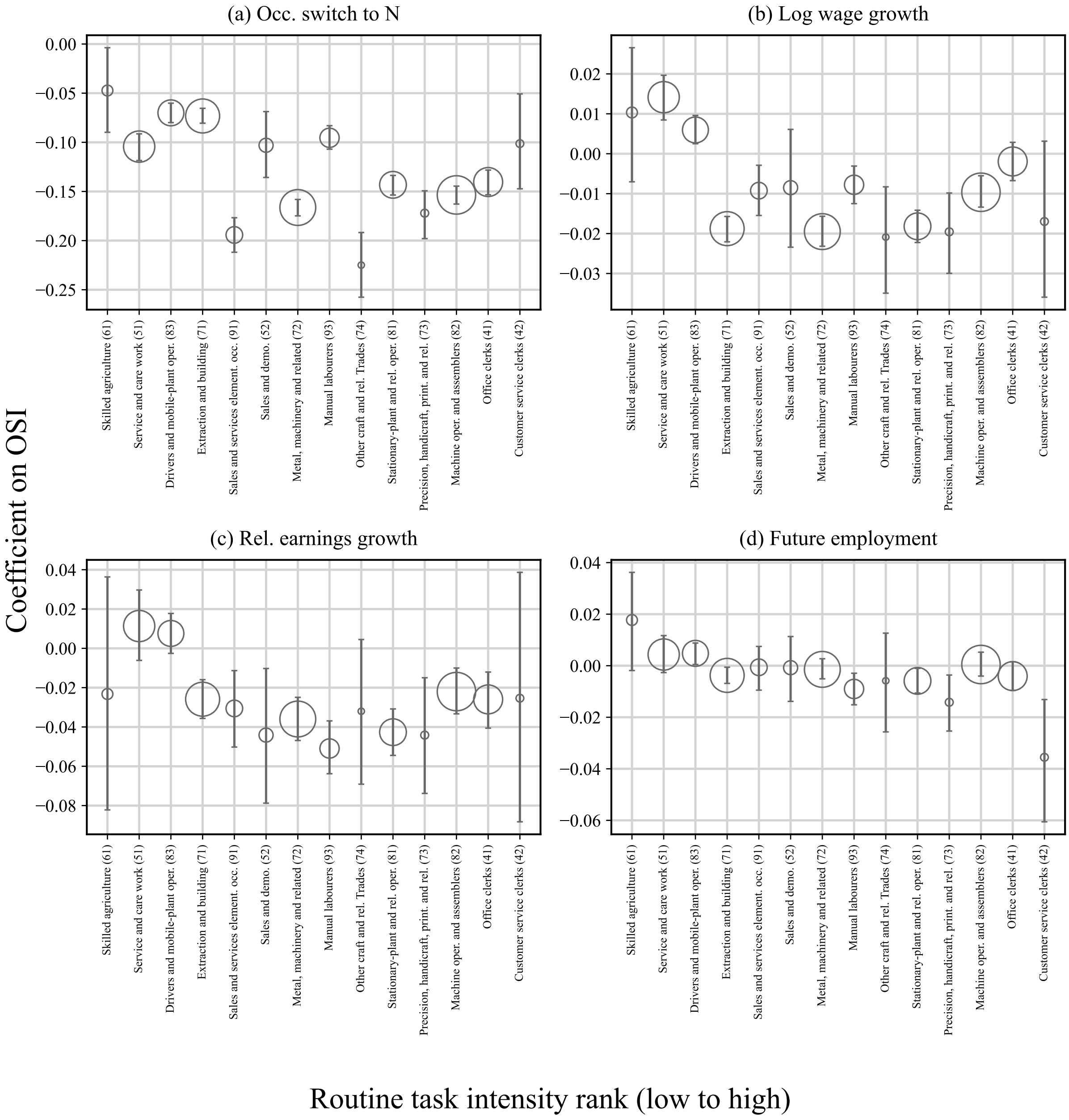
Notes: The figure plots the occupation choice probabilities according to the data in the sample used to determine the structure of each decision tree classifier against the probabilities according to the estimation sample which is not used in training. The observations are binned based on the decile groups of the structure sample probabilities.

Figure B4: Predicted and actual choice probabilities for switchers by destination occupation



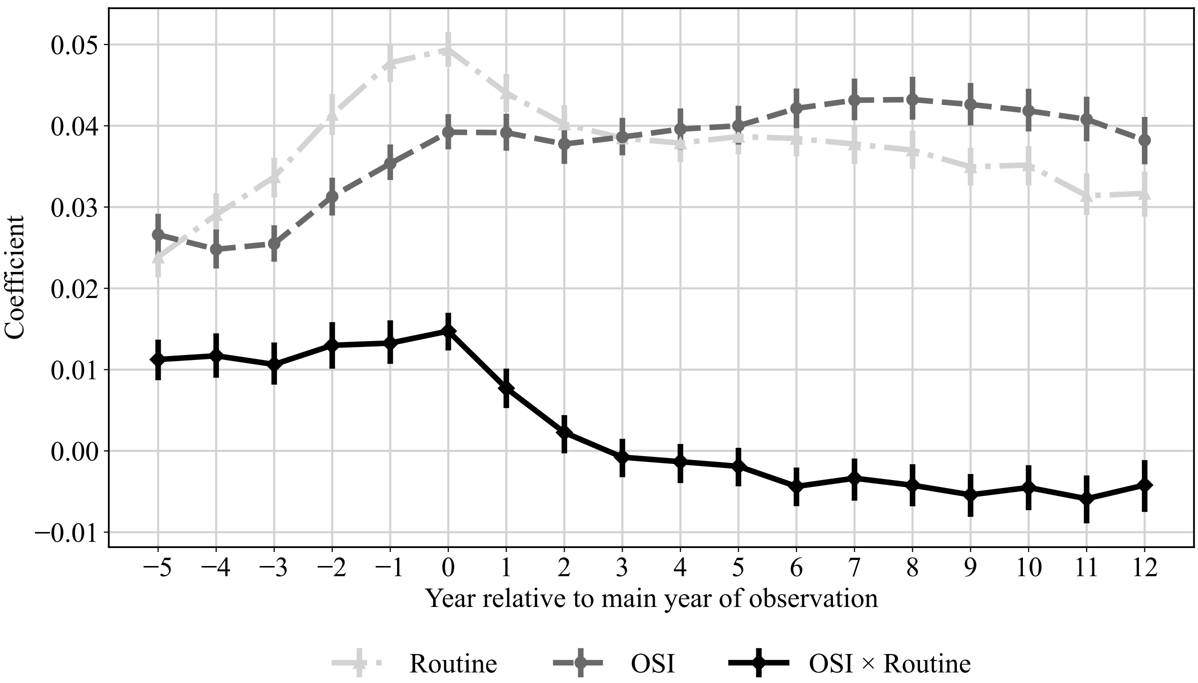
Notes: The figure is based on data for occupation switchers between t and $t + 10$ to $t + 12$. Separately by destination occupation, I produce binned scatterplots of the relationship between actual probability to switch to a certain occupation and the predicted choice probability conditional on not choosing the source occupation j. Only individuals that were not in a particular destination occupation to begin with are included. Thus, for instance, the figure for SSYK 21 reports the predicted and actual probabilities of individuals ending up in 21 for occupation switchers that did not work in 21 to begin with. Observations are grouped into six categories: one category is made up of observations with a inferred probability of zero. The other five refer to the quintile groups of observations with a positive probability. The figures are based on the first 10 bootstrap iterations.

Figure B5: Occupation-specific estimates of occupation specialization on labor market outcomes



Notes: The figure reports point estimates and 95-percent confidence intervals separately by occupation for different outcomes from a version of equation (6) which instead of the routine indicator interacts the OSI with all occupations to obtain occupation-specific OSI estimates. All regressions include fixed effects for the groups of the decision tree classifiers as well as for occupation. On the horizontal axis, the occupations are ranked by their routine task intensity (from low to high). Marker size is determined by the number of observations in each occupation.

Figure B6: Estimated effect of occupation specialization on log wage by routine and non-routine occupations and year relative to year of observation



Notes: The figure reports estimates and 95-percent confidence intervals for the routine indicator, the OSI, and the interaction between the two from equation (6) estimated on log wage in different years relative to the year of observation. I follow individuals five years back and 12 years forward in time and estimate a separate model for each time horizon. All regressions fixed effects for the groups of the decision tree classifiers.