Board of Governors of the Federal Reserve System

International Finance Discussion Papers

ISSN 1073-2500 (Print)
ISSN 2767-4509 (Online)

Number 1443

August 2026

Occupational Complexity, Capital-Skill Complementarity, and the Evolution of U.S. Wage Inequality: A Quantitative Analysis

Colin Caines, Florian Hoffmann, Gueorgui Kambourov

Please cite this paper as:
Caines, Colin, Florian Hoffmann, and Gueorgui Kambourov (2026). "Occupational Complexity, Capital-Skill Complementarity, and the Evolution of U.S. Wage Inequality: A Quantitative Analysis," International Finance Discussion Papers 1443. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/IFDP.2026.1443.

Occupational Complexity, Capital-Skill Complementarity, and the Evolution of U.S. Wage Inequality: A Quantitative Analysis*

Colin Caines† Federal Reserve Board

Florian Hoffmann‡ University of British Columbia

Gueorgui Kambourov§ University of Toronto

Monday 29th June, 2026

Abstract

We document a strong, positive relationship between occupational problem complexity, measured from US data on problem-solving requirements, and occupational wage growth since 1980. In contrast, employment shifts toward more complex occupations have been modest, suggesting a race between the demand for and supply of complex skills. We rationalize these findings by formulating and structurally estimating a quantitative general equilibrium model on the granular occupational level. In our model, workers have heterogeneous comparative advantages in solving complex problems and physical capital admits capital-skill complementarity in occupation space. The equilibrium features Positive Assortative Matching of worker skills to occupational problem complexity, and the model quantitatively explains the evolution of the occupational wage- and employment structure over the last four decades. The model estimates uncover two distinct periods of technological change. Until around 2000, rising complexity premia were driven by capital-skill complementarity and declining equipment capital prices. Post-2000 patterns reflect supply-side technological change whereby occupations became more efficient in utilizing worker skills for complex problem-solving. Our framework helps unify distinct approaches to studying task automation and task augmentation on the one hand and skill-biased technological change on the other.

Keywords: Occupational Task Content; Complex Tasks; Wage Polarization; Skills
JEL classifications: E24, J21, J23, J24, J31

*The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. The authors thank Maximiliano Dvorkin, Musa Orak, and participants at seminars and conferences. Hoffmann and Kambourov acknowledge funding from the Social Sciences and Humanities Research Council of Canada.

†Division of International Finance, Board of Governors of the Federal Reserve System, Washington, D.C. 20551 U.S.A. Email: colin.c.caines@frb.gov.

‡Department of Economics, University of British Columbia, 6000 Iona Drive, Vancouver, BC, V6T 1Z4, Canada. Email: Florian.Hoffmann@ubc.ca.

§Department of Economics, University of Toronto, 150 St. George St., Toronto, Ontario M5S 3G7, Canada. Email: g.kambourov@utoronto.ca.

1 Introduction

It is well-known that occupation-level technological changes are a key driver of the evolution of the United States (US) wage- and employment structure over the last five decades. In this paper, we show that a single-dimensional measure of occupational skill content, which we measure from data on the complexity of occupational tasks, has remarkably strong predictive power for this evolution. We document a positive and monotonic wage premium from working in more complex occupations, and we show that this skill premium has risen substantially since 1980. To rationalize the occupational skill premium and its changes over time, we formulate a quantitative general equilibrium model of the granular occupational wage- and employment structure in which worker skills, defined as comparative advantage in solving complex problems, and their complementarity with equipment capital are distinct concepts. We establish assumptions under which its parameters are identified, and we estimate the parameters on data from the US Census, the American Community Survey, the BEA Fixed Asset Tables, the US National Accounts, and O*Net.

Our quantitative model features two sources of an increasing occupational complexity premium: First, capital-skill complementarity in occupational space; Second, a technology that maps worker skills to occupational output. Skill-biased problem *automation* arises if the relative price of equipment capital declines and if labor in less complex occupations is more substitutable with capital. Occupational problem *augmentation* enters the model by way of technological progress that leads to a more efficient use of worker skills, holding constant occupational complexity.[1] Since we endogenize worker occupational choices, equilibrium quantities of labor inputs react to changes in aggregate occupational skill prices and may therefore decline in absolute levels - not only in ratios - in low-skill occupations over time. Our framework thus overcomes the criticism by Acemoglu and Restrepo (2018) of the canonical model of skill-biased technological change that it cannot feature equilibrium worker displacement. As a result, our paper builds a bridge between the literature on skill biased technological change and the rise of the college wage premium on the one hand, such as Krusell *et al.* (2000), Card and Lemieux (2001), Hubmer (2023) and Hassan *et al.* (2026), and the literature on task automation, task augmentation and the occupational wage- and employment structure, such as Autor and Dorn (2013) and Acemoglu and Restrepo (2018) and Althoff and Reichardt (2026),

[1] We use "problem" instead of "task" to emphasize the conceptual differences between our approach and the task-based approach to occupations.

on the other hand.

To construct our one-dimensional continuous measure of *occupational complexity* – defined as the degree to which an occupation involves complex problem solving – we build on work in Caines *et al.* (2017) and perform Principal Component Analysis (PCA) on O*Net occupational descriptors. In line with our structural model, which treats occupational skill content and occupational capital-labor substitutability as distinct concepts, our PCA does not incorporate any reference to whether a task can be automated or not. This is important because exposure to automation is a function of capital-labor substitutability and therefore a feature of the production technology, not of worker skills. Furthermore, complex occupations may involve manual, cognitive, or interactive problems, as long as they meet criteria for problem complexity we take from the Psychology of Problem Solving. We then document the two central descriptive findings of this paper: On the 3-digit occupational level, the wage premia on occupational complexity have increased strongly and monotonically since 1980, while the occupational employment distribution has remained comparatively stable. Hence, once one ranks occupations by their occupational complexity, the data are consistent with a race between technology and skills, similar to what has been found in the literature on Skill-Biased Technological change and the college wage premium.[2] This finding is robust to inclusion of standard occupational descriptors, such as average education and indicator variables for whether an occupation is routine or interactive, and to various choices of specification and measurement.

Characterizing occupations with an ordinal ranking of their problem complexity enables us to specify a quantitative general equilibrium model in which labor markets are segmented by skills and in which skill-specific labor-capital substitution elasticities impose testable restrictions on occupational labor-capital ratios and the corresponding relative prices. As a consequence, with data on equipment capital inputs it is not necessary to impose a-priori assumptions about how substitutable labor and capital are. This is an important distinction between our approach to modeling occupations and the task-based approach to occupations, such as Autor and Dorn (2013) and Acemoglu and Restrepo (2018). More specifically, on the demand-side of the labor market, we generalize the technology in Krusell *et al.* (2000), which features capital-skill complementarity for two types of labor, to the case with a large, but finite number of labor inputs. These inputs are distinguished by the complexity of the problems they are solving and by how complementary they are with equip-

[2] See Goldin and Katz (2008) for a book-length overview of this literature.

ment capital. In our quantitative analysis, these heterogeneous labor inputs are synonymous with occupations; and since occupational complexity ranks occupations by skill content, the model can admit capital-skill complementarity in occupational space.

On the labor supply side, we endogenize occupational choices of workers using a model of comparative advantage in solving complex tasks. We micro-found this model using a Roy-type structure, where individuals are endowed with a two-dimensional vector of skill levels, one for solving complex problems and one for solving a latent problem. A skill-aggregation technology maps this vector to a one-dimensional occupation-specific skill level. We derive conditions on this aggregator under which optimal worker choices generate positive assortative matching (PAM) in terms of a one-dimensional worker-level index of comparative advantage to occupations. This index is what we refer to as worker skills for the rest of the paper. We introduce supply-side technological change into our model by allowing the "factor productivity" of this index to vary over time. As a consequence, worker productivity in an occupation can change over time holding constant skill, capturing in a parsimonious way "task augmentation" for which Autor *et al.* (2024) provide strong evidence. This mechanism can explain why manual occupations we classify as complex, but which are labeled "routine" in the task-based approach of Autor *et al.* (2003), have experienced substantial wage growth over the last 5 decades.

A quantitative general equilibrium model such as ours, with endogenous capital-savings choices and heterogeneous workers making discrete choices, has a number of well-known technical challenges. For one, forward-looking individuals need to form expectations about the evolution of a large set of state variables, and their law of motion will be complex. In addition, the researcher needs to either assume that the economy is observed on a balanced growth path, or alternatively needs to characterize transition dynamics and where on the transition path the economy is located whenever it is observed in the data. Recent research, such as Bajari *et al.* (2007) in Industrial Organization and Kleinman *et al.* (2023) in Spatial Macroeconomics, has established that under relatively weak, though context-specific, assumptions, structural parameters can be inferred from the data without explicitly solving the entire sequence of equilibrium allocations. We adapt this approach to our context and establish sufficient conditions under which the model generates an invertible map from observable equilibrium quantities to the structural parameters of the model. One key assumption is that it is possible to estimate non-parametrically the change of the *aggregate* complexity premium relative to a base period using the Flat Spot Method from Heckman *et al.* (1998). This method is particularly useful for models such as ours that feature a distinction between the aggregate equilibrium price of skill per efficiency unit on the one hand and individual wages that are contaminated by selection on the other hand.[3] The crucial identifying assumption is that there are age groups of workers for which wage changes from one period to the next are entirely driven by changes in the aggregate price of efficiency units. Since the Flat Spot Method is conventionally used for the evolution of educational wage premia, we adapt it to the context of the occupational complexity premium, and we perform extensive validation exercises and robustness checks.

The second key assumption is that individuals make portfolio decisions in international markets, and firms rent their capital from individuals. In our model this has the implication that the user-cost formula for the prices of capital, as used for example in Krusell *et al.* (2000) and Hubmer (2023), is consistent with optimal portfolio choices and that it represents an observable version of Euler equations.[4] A prime advantage of our approach is that the map from data to structural parameters is independent of whether the economy is observed in steady state or not. On the other hand, for counterfactuals the aggregate equilibrium prices need to be treated as endogenous objects. Even so, we show that it is possible to conduct interesting counterfactual exercises without solving for the entire path of general equilibrium allocations.

Our structural estimation has four major findings. First, given external estimates of the changes in the aggregate price of skills and of physical capital, a tightly parameterized version of our model can fit the evolution of the *entire* granular occupational wage and employment structure over the last five decades exceptionally well. It matches both the within-year wage-complexity gradient across occupations and its evolution over time. We also derive a testable restriction on the cumulative distribution function of employment over occupational complexity generated by PAM and show that it is strongly supported by the data.

[3]See for example Bowlus and Robinson (2012) and Lagakos *et al.* (2018) for seminal studies using the Flat Spot Method to study the evolution of aggregate skill prices.

[4]Our assumption that asset markets are open and large relative to the economy is the same as in Autor *et al.* (2003). Other assumptions to modeling the supply of capital is the existence of a domestic technology that transfers output into capital, based on the framework in Greenwood *et al.* (1997) and often coupled with the assumption of full depreciation, as in Autor and Dorn (2013) and Caunedo *et al.* (2023), or the separation between individuals who supply labor and live "hand to mouth" on the one side and investors on the other side, as in Moll (2014) or Kleinman *et al.* (2023). In our context, our assumption is convenient because it endogenously delivers the user-cost of capital formula with depreciation as an observable equilibrium pricing schedule and because it facilitates the computation of counterfactuals. It is worthwhile to note that in Greenwood *et al.* (1997) the evolution of user-costs are fully exogenous as well.

Second, we find strong evidence for capital-skill complementarity. We estimate elasticities of substitution between equipment capital and labor of around 0.5 for the most complex occupations and around 1.6 for the least complex occupations, and they aggregate to an average elasticity of substitution between labor and capital of around 0.80 to 0.85. These estimates are broadly consistent with those found elsewhere in the literature, such as Krusell *et al.* (2000), Caunedo *et al.* (2023) and Oberfield and Raval (2021).[5] Capital-skill complementarity has two main implications for the effect of a drop in the user cost of equipment capital in our model. First, it generates a scale effect on production, which stimulates the demand for complex labor. Second, it induces firms to substitute towards equipment capital, and this substitution effect is relatively stronger for less complex occupations. This is the mechanism by which automation of low complexity occupations enters our model. Indeed, evaluated at the parameter estimates, our model predicts that occupations which are more substitutable with capital equipment were more likely to suffer declines in employment over the last five decades, and this effect is particularly pronounced at the bottom of the occupational complexity distribution.

Third, we document strong evidence for a structural break around 2000 in the time-series evolution of raw occupational skill premia and selection-corrected skill premia estimated from the Flat Spot Method. While the former grew at a substantially lower rate than the latter between 1980 and 2000, the opposite was the case thereafter. In fact, the aggregate skill premium began to level out in the post-2000 period while the slope of the wage-complexity relationship continued to rise at high rates. This result is consistent with recent findings in Beaudry *et al.* (2016), who document evidence in favor of a reversal in the demand for skills since 2000 using a task-based approach to occupations, and Hassan *et al.* (2026), who study a substantial slowdown in the growth of the college wage premium over the same time period.

Fourth, the structural break in 2000 can only be rationalized by our quantitative model with two distinct episodes of technological change. On the one hand, for the years between 1980 and 2000, the evolution of the raw occupational wage premium and the aggregate complexity premium is consistent with demand-side driven skill-biased technological change that arises from two sources. The first is labor factor productivity growth that is biased towards complex occupations, the second is the interaction between the declining price of equipment capital and capital-skill complementar-

[5] We provide a more detailed comparison in section 6.3.1.

ity. Cheaper equipment capital induced firms to increase their capital stock, driving up the relative demand for labor in occupations with high capital-labor complementarity and thereby the aggregate complexity skill premium.[6] Furthermore, PAM of worker choices led to a reallocation of workers with lower comparative advantages in solving complex problems towards more complex occupations, generating a lower growth of raw occupational wages compared to their selection corrected analogues.[7] This is a negative selection effect and is consistent with the differential evolution of the raw complexity wage premium and the estimates from the Flat Spot Method over the pre-2000 period. On the other hand, because this selection effect is hard-wired into our model, demand-side technological change cannot explain why raw occupational skill premia outgrew the aggregate price of skill in the post-2000 period. Therefore, a second type of technological change must have started to exert its force on the occupational wage structure after 2000, and this force must have come from the supply side of the labor market. In our quantitative model, this type of technological progress operates through the skill aggregation technology, and our estimates suggest that occupations have become more efficient in utilizing the comparative advantages of workers in performing complex tasks in the post-2000 period. While we do not endogenize the time-series of this efficiency process, we view it as consistent with recent findings of *labor augmentation* within 3-digit occupations in Autor *et al.* (2024) and their proposed model of directed technological change in the spirit of Acemoglu (1998) and Acemoglu and Restrepo (2018).[8] To provide some suggestive evidence for this interpretation, we correlate the various measures of task augmentation in Autor *et al.* (2024) with the change in the positive gap between raw occupational wages and the selection-corrected aggregate skill prices over the post-2000 period. As hypothesized, we find that occupations with larger exposure to task augmentation tended to be those in which raw occupational wages grew the fastest relative to the aggregate price of skill. This is a strong result because we do not use their measure at any point in our data construction of structural estimation, and because the finding that raw occupational wages have outgrown the aggregate complexity premium since 2000 and that it is informative about sources of technological change is to the best of our knowledge novel.

To quantify the relative contributions of technological progress, capital-skill complementarity,

[6]This is also the channel emphasized by Krusell *et al.* (2000) in their study of the college wage premium.

[7]See also Burstein *et al.* (2019) for a different model with comparative advantages to which the same logic applies.

[8]It also provides an explicit structural channel for the finding from French data in Berlingieri *et al.* (2024) that capital-skill complementarity is not sufficient to "generate the full increase in the relative demand for high-skilled workers observed in the data".

the decline in the price of equipment capital, worker selection due to PAM, and changes in the skill distribution over time, we use counterfactual analysis on the occupational wage- and employment structure. In contrast to the structural estimation of our model, this exercise cannot use pre-estimates of aggregate skill prices from the Flat Spot Method and instead needs to use fixed point iteration on equilibrium correspondences. This allows us to isolate the interaction between changes in the aggregate equilibrium price of skill and the distribution of occupational wages and employment. We find that without capital-skill complementarity, the ratio of occupational wages at the tails of the complexity distribution would have risen by half of its observed value, while without an increase in the efficiency of utilizing skills for solving complex problems, this increase would have been larger by a factor of almost 1.5. We also find that the skill distribution and the endogenous reallocation of labor across occupations have played quantitatively important roles in dampening these effects. As indicated by our structural estimation, there is not a single factor that can explain the structural break in the occupational wage structure after 2000. In particular, the post-2000 period is strongly impacted by supply side technological change, which slowed both the rise of the complexity premium and of employment in complex occupations.

Additional Literature Our paper contributes to a large literature on the evolution of the US cross-sectional wage distribution over time and its interaction with technological progress. Two views currently dominate this literature. On the one hand, the hypothesis of Skill-Biased Technological change (SBTC) postulates that gains from modern technological progress accrue mostly to highly educated workers. This hypothesis is supported empirically by the well-documented "race between skill and technology," whereby the returns to education and the college wage premium have risen dramatically since the late 1970s, exactly when the increase in the relative supply of skilled labor started to slow down. Key references are Katz and Murphy (1992), Card and Lemieux (2001) and the book-length treatment in Goldin and Katz (2008). Hoffmann *et al.* (2020) provide recent estimates of the contribution of rising education returns to trends in wage inequality in various advanced economies, and Autor (2014) highlight that the skill premium remains an important driving force of wage inequality. Hassan *et al.* (2026) document a slow-down in the growth of the college wage premium since the early 2000s and show that a model with technological creation and worker learning can quantitatively match its long-term trends.

On the other hand, proponents of the Routine-Task Replacing Technological Change (RRTC) hypothesis show that the evolution of the occupational employment and wage structure has featured polarization in terms of baseline occupational wages, usually measured in 1980, and interpret this as evidence that it cannot be fully accounted for by SBTC and by a monotonic rise of an occupational skill premium. The key to this literature is the task-based approach to occupations, innovated by Autor *et al.* (2003), which views occupations as bundles of tasks that workers must perform to produce output. Autor and Acemoglu (2011) and Autor and Dorn (2013) provide an in-depth analysis of the polarization of the occupational wage- and employment distribution since 1980 and argue that it cannot be explained entirely by SBTC. Goos *et al.* (2014) document similar job polarization in Europe, and Deming (2017) adds social skills as an important predictor of occupational wage growth.[9] An important generalization of the RRTC hypothesis is the theoretical model in Acemoglu and Restrepo (2018), in which technological progress does not only lead to the replacement of labor by capital (i.e. task automation), but also to the creation of new tasks solved by labor (i.e. task augmentation).

Much progress has been made recently in collecting more direct micro-level evidence on the effect of automating technology and of Artificial Intelligence (AI) on occupation- and industry-level labor market outcomes. Acemoglu and Restrepo (2022), Eloundou *et al.* (2023) and Acemoglu (2024) find that a substantial majority of the U.S. Labor force is exposed to task replacement by automation and large language models. On the other hand, Autor *et al.* (2024) and Hampole *et al.* (2025) find substantial evidence for changing task composition within 3-digit occupations, which can be interpreted as evidence of task augmentation. We use the measures developed in Autor *et al.* (2024) to argue that the type of supply-side technological progress we estimate for the post-2000 period likely captures task augmentation, at least to some extent. Althoff and Reichardt (2026) and Freund and Mann (2025) estimate quantitative life-cycle versions with skill heterogeneity of the model in Acemoglu and Restrepo (2018) using data from the NLSY.

Our paper takes a step towards unifying these two strands of literatures by formulating a model of the granular employment and wage distribution across occupations that features a well-defined skill premium and by showing that it can quantitatively match their observed changes since 1980.

[9] Gathmann and Schonberg (2010), Jaimovich *et al.* (2021), and Jaimovich and Siu (2020) further refine the categorization of occupational groups to study the role of occupations for various labor market outcomes.

Novelties to our approach are the explicit distinction between occupational skill content and capital-skill complementarity in measurement and modeling, and the incorporation of data on capital-labor ratios. An important finding from our structural estimation is that supply-side technological change that increases the within-occupational efficiency of using worker comparative skills in solving complex problems, and which captures task augmentation in a parsimonious way, is necessary to explain trends in the occupational wage- and employment structure since 2000. This confirms the importance of task augmentation in understanding the impact of AI on the occupational employment distribution emphasized by Autor *et al.* (2024) and Hampole *et al.* (2025).

Our quantitative exercise contributes to a long line of research in macroeconomics that formulates and calibrates macro models of the evolution of the cross-sectional wage distribution. The seminal study of skill premia is Krusell *et al.* (2000), who introduced the concept of capital-skill complementarity and its interaction with the decline of the relative price of capital equipment. More recent work has highlighted the importance of capital-skill complementarity in explaining the decline in labor shares, as in Karabarbounis and Neiman (2014) and Hubmer (2023). In light of the results from these works, methods for estimating capital-labor substitutability have attracted substantial attention recently. One strand of this literature uses micro data in order to handle inherent endogeneity issues when estimating substitution elasticities, a notable example being Oberfield and Raval (2021). In contrast, this paper recovers occupation-level capital-labor substitution elasticities through structural estimation of a macro model. While we find strong evidence of capital-skill complementarity, it cannot fully account for either the observed increase in complex skill premia or the redistribution of labor across occupations after 2000.

There is little quantitative work that incorporates physical capital into models of the occupational wage- and employment structures. Important exceptions are Burstein *et al.* (2019) and Caunedo *et al.* (2023), who combine the task-based approach to occupations with a technology in which each occupation produces an intermediate good using occupation-specific capital- and labor inputs. Neither of their frameworks features a concept of occupational skill premia and thus of capital-skill complementarity, which is a primary focus of our paper.[10] We therefore view our approach as complementary to theirs, with a substantially different focus.

[10] A detailed discussion of the difference between their technology and ours, which is a generalization of the production function in Krusell *et al.* (2000) and does not involve intermediate goods, is provided in section 4.3.

2 Data and Measurement

2.1 Overview

In this section we describe the construction of our data and the measurement of the variables at the center of our quantitative analysis. These are, principally, occupational complexity, occupational wages per efficiency units (referred to as *skill prices* in the following), and the user costs of capital.[11] We rely on five sources of data: the Decennial Census, the American Community Survey (ACS), the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS), the Bureau of Economic Analysis (BEA) Fixed Asset Tables, and the Occupational Information Network (O*Net).[12] We use occupational codes on the 3-digit level, and variables are measured at decennial time intervals, with the exception of skill prices, which we estimate on the annual level.

2.2 Employment and Wages: Census and ACS

We construct employment- and wage variables from the Census and the ACS following standard procedures in the literature. For the years 1980, 1990 and 2000 we use the 5 percent samples of the Census, and for 2010 and 2020 we use its successor, the ACS.[13] To have similar sample sizes we combine the 2008-2010 and the 2018-2020 ACS samples and treat them as a single year. To have a time-consistent definition of occupations we rely on the occupational classification created by Autor and Dorn (2013), which enables us to track employment and wages in 317 3-digit occupations from 1980 to 2010. We extend this classification to 2020, which we were able to do for 292 of these occupations. For each wave we restrict the sample to non-farm workers between the ages of 16 and 64 who are not institutionalized. We do not exclude any other demographic groups from our samples, which helps establishing the universality of our results.

Following seminal work in the literature on occupational employment- and wage structures, such as Autor and Acemoglu (2011) and Autor and Dorn (2013), we use the total number of hours worked as our measure of occupational labor supply. We also calculate hourly wages from reported

[11] Additional details about sample construction and measuring occupational complexity are given in the appendix and in our earlier paper Caines *et al.* (2017).

[12] In Caines *et al.* (2017) we also used the Dictionary of Occupational Titles (DOT) to measure occupational complexity. Since the results are qualitatively similar, but the O*Net is of substantially higher quality, we exclusively rely on the latter in this paper. See footnote (29) for further comments.

[13] Dollar-denominated variables are adjusted for inflation using the PCE price deflator.

hours of work and from reported labor income. Labor income data in the Census and the ACS are top-coded, and the imputation in the public-use files changes over time. Accordingly, we develop a time-consistent imputation approach based on a two-parameter Weibull distribution, which we describe in appendix C.1.[14]

2.3 Defining and Measuring occupational complexity

2.3.1 Defining occupational complexity

We construct a one-dimensional measure of occupational complexity by adapting standard concepts from the psychology of human problem solving to our context. This approach views an occupation as a set of problems a worker needs to solve, possibly with the support of machines. A **problem** in turn is defined as a tuple consisting of an initial set of state variables, a set of goal state variables, and a set of operations that moves the state from its initialization to the goal.[15] It is **well-defined** if there is full and unambiguous information about every element in this tuple. In this case the problem can be formulated algorithmically in a straightforward way.[16] A problem that is not well-defined is **ill-defined**, and an ill-defined problem is **complex** if some elements in the tuple change dynamically during problem solving and are intransparent. As a consequence, *complex problem solving implies the efficient interaction between a solver and the situational requirements of the task, and involves a solver's cognitive, personal, and social abilities and knowledge* [Wenke and Frensch (2003), p. 90].[17]

[14]It should be noted that our main findings are insensitive to the imputation approach. For example, Caines *et al.* (2017) do not impute top-coded earnings observations. Yet, their findings are qualitatively identical and quantitatively very similar to those we document in section 3.

[15]This definition of a problem is commonly referred to as the "Gap" approach in the Psychology of human problem solving.

[16]The algorithm may still be complex in the computational sense. The definitions we use here refer to human- rather than machine problem solving capabilities. An example is driving from point A to B, which is a well-defined problem which currently lies at the edge of computational capabilities of machines. As is clear from these definitions, a well-defined problem can be formulated in terms of Dynamic Programming or Reinforcement Learning.

[17]While the concepts of "complex problems" and "complex tasks" are ubiquitous in the psychology of problem solving, it is surprisingly difficult to find a precise and operational definition of "complexity" in the human context. Here we rely heavily on the summary paper on the topic, Wenke and Frensch (2003). On the other hand, the "gap" approach to defining problems as we do here, and the classification into well- and ill-defined problems, are standard in the field.

2.3.2 Measuring occupational complexity

The definition of complex problem solving above is precise enough to serve as a guide for measuring the complexity of an occupation from data that describe occupational skill- and task requirements. Let j be an index for occupations and o an index of occupation descriptors which characterize features of complex problem solving, including tasks or skills that meet the definition of "complexity" given above. Let $q_{o,j}$ be an ordinal measure of the intensity with which descriptor o is used in solving problems encountered when working in occupation j. We follow the approach in Yamaguchi (2012) and construct a measure $x_j \in [0, 1]$ that ranks occupations ordinally according to the complexity of their problem-solving requirements by performing a PCA on data for $q_{o,j}$. We draw these data from the O*Net.[18]

The O*Net groups its descriptors into eight subsections, such as "Abilities", "Interests" and "Work Values." Following earlier work in Caines *et al.* (2017), and analogous to Autor *et al.* (2003) and Yamaguchi (2012), we select a subset of these descriptors that map well into our classification of occupations, in our case "complexity" rather than "routineness." In particular we select descriptors which contain key words, or are related to key concepts, entering the definition of complexity introduced above. For example, "complex problem solving skills" is one descriptor in the subsection "Cross-Functional Skills." We add 34 more descriptors drawn from three subsections because (a) putting full weight on only one factor would heavily expose our measure of occupational complexity to measurement error and (b) the additional descriptors relate well to the concept of problem complexity[19]. Generally, these descriptors are related to various dimensions of "comprehension", "reasoning", "solving", "flexibility", and various higher-order social skills.

The O*Net assigns, for each of approximately 1000 occupations and for each occupational descriptor, a value between 0 and 7, which indicates the degree to which the descriptor captures a task that is performed in a given occupation. These are the empirical analogues of the measures $q_{o,j}$. We transform these values into a single-dimensional complexity score by computing the first principal

[18] The O*Net is operated and maintained by the U.S. Department of Labor. Two important differences between our approach and Yamaguchi (2012) are that we use the O*Net instead of the DOT and that we use a substantially broader set of descriptors.

[19] The descriptors are drawn from three subsections of the O*NET: "Abilities" (contained in "Worker Characteristics"), "Skills" (contained in "Worker Requirements"), and "Generalized Work Activities" (contained in "Occupational Requirements") – see Caines *et al.* (2017) for more details.

component across all 35 descriptors, which is then re-scaled to between 0 and 1.[20] We then match our complexity measure to the Census/ACS sample by mapping the O*Net-SOC occupational taxonomy into the occupational codes in Autor and Dorn (2013). Since our approach does not require any aggregation into occupational groups, the only degree of freedom in measurement is the choice of the descriptors for our PCA, which we view as a strength of our occupational complexity concept. The relevant descriptors and their factor loadings in the PCA are shown in table A.1.[21]

2.3.3 Routine or Simple?

Our complexity measure shares some conceptual similarity to "non-routine tasks", as discussed in work such as in Autor *et al.* (2003). The central difference is that we measure occupational complexity exclusively from the skill content of problems workers need to solve, without any reference to automation or capital-labor substitutability. This allows us to rank occupations by a one-dimensional measure of "skill" and to treat worker skill and capital-labor substitutability as distinct concepts. While our goal is not to attempt a horse race between these two approaches to occupational classification, it is helpful to demonstrate their conceptual differences with a more tangible example.

Consider tasks that require walking: As discussed for example in Pinker (1997), walking steadily and smoothly on two feet is a computationally difficult problem, so difficult in fact that no machine has been able to mimic it accurately. A similar example, used in Autor *et al.* (2003), is driving a car or a truck in heavy traffic. With routine tasks defined as those *"that can be accomplished by machines following explicit programmed rules"* [p. 1283 in Autor *et al.* (2003)], occupations that heavily feature walking or driving are "non-routine". However, most adult humans are adept at tasks that involve these problems and solve them repeatedly every day. Psychologists studying the human capabilities of problem solving therefore commonly refer to them as "everyday problems".[22] From a human problem solving perspective, they are therefore non-complex or **simple**.

This distinction matters for the predicted effect of automation on wages in such occupations. By

[20] For our benchmark analysis we use version 20.1 of the O*Net, published in 2015, though in our robustness checks we consider other waves of the O*NET.

[21] This table is reproduced from Caines *et al.* (2017).

[22] Ericsson (2003) defines everyday problems as those where *"after a limited period of training and experience [...] an acceptable standard of performance can be generated without much need for effortful attention. At this point, execution of the everyday activity has attained many characteristics of automated performance and requires only minimal effort."* [p. 62].

definition, labor in non-routine occupations cannot be substituted by machines, and it is therefore hypothesized that technological progress will not generate downward pressure on its wages. In contrast, simple occupations involve "every day tasks" and therefore present a feasible outside option for workers displaced by machines. Wages in such occupations are thus directly negatively affected by technological progress, *no matter if they can be automated or not, i.e. even if they are non-routine*. The routine-vs. non-routine classification segments the labor market by substitutability with machines, while the complexity-measure segments the labor market by skill. Caines *et al.* (2017) provide detailed evidence that the binary classification into simple vs. complex occupations is more consistent with empirical evidence on the distribution of occupational wage growth. We complement this evidence below using our continuous complexity-measure on a substantially longer time-span. We also formulate and estimate an equilibrium model with complex-task biased technological change that can fully account for the evolution of the occupational wage- and employment structure.

2.4 Capital and Capital Prices: BEA Fixed Asset Tables and U.S. National Accounts

Data on the stocks of capital equipment and capital structures are taken from the BEA's fixed asset tables. In each case we take current-cost values of the net capital stock and deflate by the consumption deflator as well as, respectively, the capital equipment and structures price deflator. This latter point is particularly important as the relative price of capital equipment has fallen dramatically since 1980, and so one would otherwise understate the dramatic increase in its input quality since 1980. This is shown in figure A.1, which plots the prices of equipment and structures relative to consumption (*i.e.*, the relevant capital deflator divided by the consumption deflator).[23] Depreciation rates are computed from the ratio of the current cost of capital equipment or structures depreciation, also taken from the BEA fixed asset tables, to the relevant capital stock. Appendix C.2 provides details on the relevant series.

[23] Apart from the longer time span in our data, this figure is identical to figure 5 in Hubmer (2023).

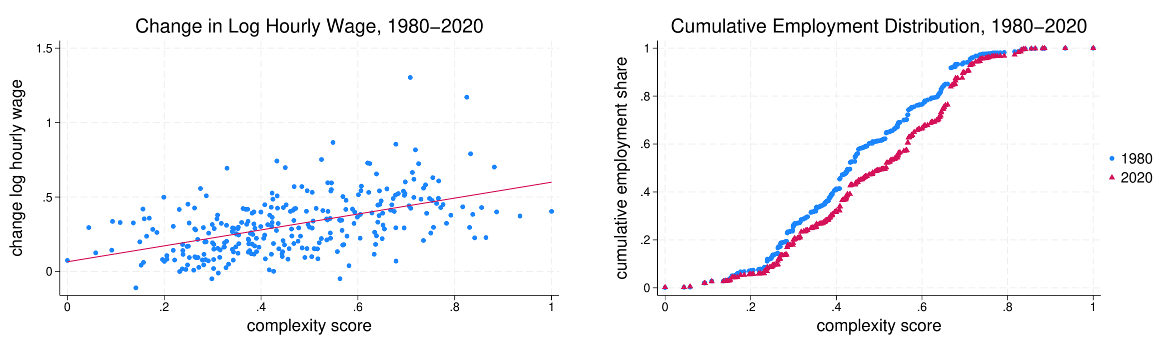
3 Descriptive Analysis

In this section, we present the main descriptive results of this paper, which motivate the equilibrium model we will formulate and structurally estimate below. All figures and tables in the main text use version 20.1 of the O*Net for constructing the complexity measure. Results from robustness exercises that use different waves of the O*Net can be found in the appendix.[24]

3.1 Aggregate Trends in Occupational Wages and Employment

The two panels of 1 show the empirical relationship between occupational complexity on the one hand and changes in log hourly wages (left panel) and in occupational shares of raw hours worked (right panel) between 1980 and 2020 on the other hand. They are the simplest visualization of the motivating facts of this paper: Changes in log hourly wages have a strong positive correlation with occupational complexity, but the distribution of raw hours worked has only modestly shifted towards more complex occupations.[25]

Figure 1: Change in Occupational Wage and Employment Share by occupational complexity, 1980-2020.



We estimate these effects net of other skill-relevant factors using linear regressions at the 3-digit occupational level with hourly wages and employment shares as outcomes and the following

[24]This section is an updated version of the descriptive results in Caines *et al.* (2017). The main differences to our earlier empirical analysis are: (i) The final sample year is 2020 instead of 2005; (ii) We exclusively use a continuous measure of task-complexity; (iii) We present results using occupational complexity measured in four, rather than one wave of the O*Net, but we do not use the DOT anymore; (iv) Top-coded earnings in years before 2000 are imputed using the procedure described in section 2; (v) Alaska and Hawaii are included in the sample; and (vi) women are included in the sample.

[25]The panel plots changes in raw log hourly wages, together with a line of best fit, for the 292 3-digit occupations that are consistently available across the years in our data

variables as controls: means of the age of a worker and of the number of children they have, and the shares of workers who are female, who are non-white, who are married, and who have at least a high school degree. Results are shown in columns (1) and (4) of table 1. For hourly wages we find a highly significant and quantitatively large estimate of 0.61, implying that the most- and least complex occupations experienced a difference in wage growth of over 60 log points. On the other hand, the estimate for the changes in employment shares, while significant on the 5 percent level, is rather small, with a point estimate of 0.005. While this coefficient is difficult to interpret, an issue we address further below, the important take away is that complexity has modest predictive power for long-run employment changes, at least relative to long-run changes in hourly wages.[26]

Table 1: Occupation-level Relationship between Complexity and Hourly Wage/Employment

	Δ log hourly wage 1980-2020			Δ log hours 1980-2020		
	(1)	(2)	(3)	(4)	(5)	(6)
Complexity	0.607***	0.611***	0.499***	0.00535**	0.00422*	0.000663
	(5.76)	(5.70)	(3.84)	(2.39)	(1.89)	(0.25)
RTI		0.0130	0.0293		-0.00381***	-0.00329**
		(0.21)	(0.47)		(-2.97)	(-2.55)
Social Skill			0.133			0.00423**
			(1.52)			(2.33)
N	292	292	292	292	292	292

Note: Regressions include demographic controls: occupation-level shares of whether workers have a college/high-school degree and whether they are female/non-white/married, and occupation-level averages of age and the numbers of children. Standard errors are clustered on the level of occupations. t-statistics are in parenthesis.
Significance levels are: *** 1%, ** 5%, * 10%.

As a robustness check we estimate regression models on more disaggregated data, which allows us to control more flexibly for other factors that may affect changes in occupational wages- and employment. To this end, we calculate average changes in log hourly wages and employment on the level of cells defined by occupation, sex, educational attainment, race and age. We then estimate complexity regressions that include fixed effects for each of the categories defining cells. The baseline results are shown in columns (1) and (4) of table A.2 and can be directly compared to the estimates

[26] One quantitative interpretation of this coefficient is as follows. The area underneath the regression line $\beta_0 + \beta_1 * x$ from $x = a$ to $x = 1$ is the approximate change of the employment share among occupations with a complexity above a. For $a = .5$ for example, this area is bounded above by $.5 * \beta_1$. Since we estimated $\beta_1 = .005$, this is a rather small value.

from the aggregate regressions in table 1. The point estimates for changes in log wages are very close: 0.61 for the aggregate specification and 0.55 for the disaggregated specification. Results for employment changes are more sensitive as the estimate from the disaggregated regression is smaller by a factor of 13. This is to be expected since whereas the relationship between complexity and changes in log-wages is quite precise, the same is not true for employment shares, as shown earlier.

3.2 Occupational Wages and Employment: Decadal Analysis

Next, we estimate the relationship between log-wages and employment *in levels* separately for each year included in our sample. Results are shown in the five panels of figure A.3 in the appendix. All figures use the same scales on both axes so that slopes of the lines of best fit are directly comparable.[27] The central finding from these figures is that the empirical relationship between log-wages and occupational complexity is remarkably precise in all 5 decades and is well captured by a simple linear regression. Importantly, there is virtually no evidence for non-monotonicity in the complexity wage premium. Stated differently, for each decade there is a well-defined complexity wage premium, and this premium is simply the slope of the fitted line. At the same time, the rate at which this rate has increased is not constant across decades. On average it increased by approximately 0.15 log points per decade, but it rose substantially faster in the 1990s and not at all after 2010, a point we return to in the next section.

Figure A.4 repeats the same exercise, but using the share of total hours worked as a function of occupational complexity as dependent variable. Because the relationship in shares is noisy, we plot the empirical cdf's instead. In comparison to hourly wages, their changes are relatively small, but not zero. For example, in 1980 the median occupational complexity was around 0.43 and increased to about 0.52 in 2020. The corresponding numbers are 0.28 and 0.3 for the 20th percentile and 0.64 and 0.66 for the 80th percentile. Hence, while the employment changes at the tails of the complexity distribution were muted, there was substantial reallocation near its median. This can also be seen in the left panel of figure 1, where some particularly large employment declines are observed for a level of occupational complexity between 0.4 and 0.5, whereas the largest employment increases took place

[27] One may be concerned that the slope of the estimated regression line is driven by a few large occupations. We therefore have also estimated the fitted linear model when observations are weighted by employment. This has only a negligible impact on estimates. This is as expected, given the limited amount of noise in the scatter plots. We therefore show only results for the unweighted regressions here.

in occupations with occupational complexity above these complexity scores. Hence, employment reallocation took place mainly in the center of the occupational complexity distribution and towards occupations that are relatively close in occupational complexity space. This is broadly consistent with findings of direct worker job transitions in Gathmann and Schonberg (2010) and Fan (2025), but quite different from the RRTC-hypothesis, according to which employment is reallocated towards the tails of the occupational skill distribution as measured by its wage levels in 1980.

We perform two robustness checks for this graphical analysis. First, one may worry about the fact that we measure occupational complexity at one point in time in the figures above. While this continues to be the standard approach in the literature, it is an important concern because it may lead to confounding changes in returns to complexity across occupations with tasks becoming more complex within occupations over time.[28] In appendix figure A.5 we thus replicate our analysis from figure A.3, but using three other waves of the O*Net instead. We choose version 15.1, version 20.1 and version 27.3 because this covers as large a time span as possible while allowing for a time-consistent measure of occupational complexity.[29] Since a simple regression line captures the relationship between average log-wages and occupational complexity very well, we show, in the same plot, the fitted models for the four different complexity measures without the underlying scatter.

Our second robustness check addresses the concern that the scale of our continuous complexity measure, though standardized to be between 0 and 1, is hard to interpret. We therefore reproduce figure A.5 using the complexity measure in terms of percentiles instead, which is shown in the five panels of figure A.6.

Overall we find that our main results are extremely robust across all of these specifications.

3.3 Routine- vs. Non-Complex Occupations

In this section we discuss the effect of adding the routine-task index (RTI) from Autor and Dorn (2013) and the social skill measure from Deming (2017) to our baseline regressions. As explained in section 2.3.3, under the hypothesis of complexity biased technological change labor markets are

[28] The majority of work that studies long-run trends in the occupational wage- and employment structures indeed relies on a single wave of task-usage data. A recent example is Autor *et al.* (2024) which uses the DOT for the entire post-1980 period.

[29] We use version 15.1 as the earliest wave because the BLS released updated SOC codes starting with this wave. This allows us for a time consistent match between occupational codes in the different data we use that relies on the crosswalk published by David Dorn and first used in Autor and Dorn (2013).

segmented by the comparative advantage of solving complex problems rather than by whether occupational tasks can be automated or not. As a consequence, wages across occupations that are close in complexity space are equalized, but those which start to be automated experience employment losses. Columns (2) and (5) of table 1 show this to be the case. Conditional on occupational complexity, RTI has no impact on the changes of occupational wages between 1980 and 2020, but it is strongly negatively related to changes in employment. The same is true when adding social skills as well in columns (3) and (6) of the table, but with the opposite sign for employment. We find qualitatively identical results when estimating the models on the disaggregated data instead, as shown in table A.2. Overall these results support the findings in Autor and Dorn (2013) and Deming (2017) for occupational employment. The primary difference is how we interpret the evolution of the occupational wage structure, and this informs our choices for the quantitative model formulated and estimated below.

3.4 Is there a Race between the Demand for and Supply of Complex Skills?

The primary evidence for a so-called "race between education and technology", summarized at length in Goldin and Katz (2008), is the finding that since at least 1970 the college premium has increased more rapidly during periods of relatively slow growth in the supply of educated workers. In order to test if there exists an analogous race between the supply of and demand for labor specialized in solving complex problems, we re-estimate our regressions decade by decade. Figure 2 plots the coefficients from regressions of hourly wage and employment changes on occupational complexity for each of the four decades covered in our sample. We perform this exercise using our richest specification, which includes fixed effects for all categorical demographic variables as well as both the RTI-index and the Social-Skill index, and which is performed using disaggregated data.[30]

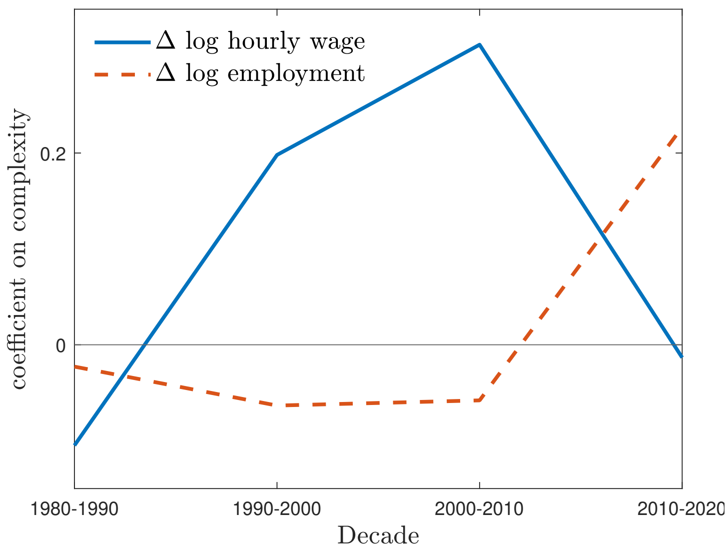
The results are quite striking. Conditional on covariates, decades that saw growth in the complexity wage premium correspond to those that saw relatively little increase in the supply of complex labor. Indeed, the strongest growth in the complexity premium was observed between 2000 and 2010, when complex employment actually decreased.[31] In contrast, the following decade was the only period that saw increased complex employment, and this was also the only decade with no ob-

[30]Point estimates and their standard errors, together with some additional specifications, are shown in tables A.5 and A.6.

[31]Note, this relationship is significant at the 1 percent level, see table A.6.

served increase in the complexity wage premium. We conclude that once occupational skill content is characterized by their problem complexity, the data are consistent with a race between skills and technology, akin to what has been documented in the context of the college wage premium.

Figure 2: Regression of Wage & Employment Changes, by Decade, on Complexity.



4 The Model

4.1 Overview

We consider a general equilibrium model of an economy which, at each date $t \in \mathbb{N}_0$, is organized into a unit measure of individuals and a continuous measure of firms. A firm produces y_t units of output using labor in J occupations, written $l_{j,t}$, and two types of capital inputs, equipment capital k_t^e and capital structures k_t^s. Labor inputs $l_{j,t}$ are measured in efficiency units. An individual derives utility from goods, and it has earning from wages on its labor, which it supplies inelastically, and and from the returns on its accumulated savings. Compared to a neoclassical multi-sector model, our framework integrates two novel features. First, we formulate a production function that combines capital with labor in occupations that are differentiated by task-complexity x, where each type of labor input has its own complementarity with equipment capital. Second, we endogenize the selection of workers into occupations, with a resultant mapping from worker skills to occupational complexity that is non-standard.

All markets are competitive, with the following market-level prices: p_t denotes the price of the final good, $\overline{w}_{j,t}$ denotes wages per efficiency unit of labor in occupation $j \in \{1, 2, ..., J\}$, and (r_t^e, r_t^s) denotes the rental rates of the two types of capital. There is no uncertainty. For notational convenience, define vectors $\overline{w}_t = (\overline{w}_{1,t}, ..., \overline{w}_{J,t})$, $\mathbf{r}_t = (r_t^e, r_t^s)$, and $\mathbf{k}_t = (k_t^e, k_t^s)$. Also note that an occupation j is uniquely characterized by its task complexity $x_j \in [0, 1]$, so that we can treat occupational choices and the choices of occupational complexity interchangeably. We define the aggregate price schedule of skill by $\pi_t(x_j) = \overline{w}_{j,t}$ and will refer to $\pi_t(x)$ as the skill price function in the following. The **aggregate skill premium** as a function of x is given by $\Phi(x) = \frac{\pi'(x)}{\pi(x)}$. This will correspond to the selection-corrected wage premium paid to complex labor, hereafter referred to as the **complexity premium**.

4.2 Individuals

4.2.1 Decisions

Individuals have preferences $U = \sum_{t=0}^{\infty} \beta^t \cdot U(C_t)$ over sequences of consumption bundles $\{C_t : t \in \mathbb{N}_0\}$, where $\beta \in (0, 1)$ is the discount factor and where flow utility over consumption satisfies $U'(C_t) > 0$, $U''(C_t) < 0$ and $U'(C_t) \to \infty$ as $C_t \to 0$.[32] Individuals receive income from three sources. First, at each date t it owns a portfolio $\mathbf{s}_t = (s_t^e, s_t^s)$ of claims on capital equipment and capital structures, which it rents to firms at the rates $\mathbf{r}_t = (r_t^e, r_t^s)$. Capital is purchased on international markets at prices $\mathbf{q}_t = (q_t^e, q_t^s)$ and depreciates at rate $\mathbf{d} = (d^e, d^s)$. Second, it holds one-period bonds b_t with a return of $\bar{r}_t$, which it purchases on an international bond market. Third, it is endowed with a unit of labor it supplies inelastically and with a two-dimensional skill endowment $(a, d) > 0$ that is a random draw from a population joint cdf $G_t(a, d)$. Given this endowment it makes occupational choices, and since there is no preference heterogeneity and no skill dynamics, this choice is made to maximize income. Occupational income in turn is the product of the aggregate skill price and individual efficiency units. The latter is determined by a technology $h_t(x, a, d)$ that maps individual skill endowments to efficiency units of labor in an occupation characterized by task complexity x. Hence, a worker with skill endowment (a, d) receives an income

[32] Neither our calibration nor our simulation of counterfactuals rely directly on any assumptions about preferences. However, as explained below, we recover user-costs of capital from Euler equations, which implicitly imposes the assumption that consumption paths are along interior solutions. The assumptions on flow utility stated here guarantee that this is satisfied.

of $w_{j,t}(a,d) = \pi_t(x_j) \cdot h_t(x_j, a, d)$ when choosing occupation j. Notice that a two-dimensional skill structure is required to ensure that the equilibrium distribution of occupational choices is non-degenerate, otherwise workers would always make corner choices (where $x = 0$ or $x = 1$). In the next section we will derive sufficient conditions under which occupational choices are fully characterized by the ratio of the two skills, $\frac{a}{d}$, which we will interpret as an index of comparative advantage.

Given the separation of occupational choices from consumption- and savings choices, the following timing of the decision process is convenient. At the beginning of a period t, an individual observes the state $(a, d, \mathbf{s}_{t-1}, b_{t-1})$, the Walrasian Auctioneer announces prices $(\pi_t(x), \mathbf{p}_t, \mathbf{r}_t)$, and individuals decide which occupation to work in. We write $M_t(a,d)$ for the income from optimal occupational choices, and we let $j_t^*(a,d)$ be the corresponding policy function. Next, the individual chooses consumption C_t and bond and capital holdings, b_t and $\mathbf{s}_t$, the latter of which are rented to firms. Formally, the decision at the first stage is described by

$$\begin{aligned} j_t^*(a,d) &= \arg\max\{w_{j,t}(a,d), j = 1, ..., J\} \\ M_t(a,d) &= \max\{w_{j,t}(a,d), j = 1, ..., J\} = w_{j_t^*(a,d)}, \end{aligned} \tag{1}$$

and at the second stage the consumption-savings decision solves the perfect foresight problem

$$\begin{aligned} V(M_t, \mathbf{s}_{t-1}) &= \max\{U(C_t) + \beta \cdot V(M_{t+1}, \mathbf{s}_t)\} \qquad (2) \\ p_t \cdot c_t + \mathbf{q}_t \cdot (\mathbf{s}_t - (1 - \mathbf{d}) \circ \mathbf{s}_{t-1})' + b_t &= M_t + \mathbf{r}_{t-1} \cdot \mathbf{s}'_{t-1} + (1 + \bar{r}_{t-1}) \cdot b_{t-1} \end{aligned}$$

The first-order conditions for portfolio choices, combined with the Envelope conditions and the assumptions on flow utility deliver the two Euler equations

$$r_t^k = q_t^k \cdot \left(1 + \bar{r}_t - (1 - d_t^k)\frac{q_{t+1}^k}{q_t^k}\right). \tag{3}$$

These equations are the standard user-cost-of-capital formulae. At the margin, individuals are indifferent between holding their wealth in capital (structures or equipment) or bonds. The equations in (3) can therefore be alternatively derived by simply equalizing the equilibrium returns on each type of asset.[33]

[33] As will be discussed below, our estimation approach avoids solving for General Equilibrium, and in particular

4.2.2 Skill Aggregation

Occupational choices described in equations (1) are driven by the aggregate skill price function $\pi_t(x)$, individual-level skill endowments (a, d) and the aggregator $h_t(.)$ that maps two-dimensional skills into occupation-specific efficiency units. We now develop conditions under which these equations deliver a model with **Positive Assortative Matching (PAM)** in which workers sort monotonically into occupations based on their comparative advantage in solving complex problems. We focus on equilibria with PAM because otherwise a model with multidimensional skills such as ours does not have a natural concept of a "skill premium" and of "capital-skill complementarity".[34]

To start with, assume that $h_t(.)$ is constant-returns to scale in (a, d) so that $h_t(x, a, d) = d \cdot h_t(x, \tilde{a})$, where $\tilde{a} \equiv \frac{a}{d}$. Then d plays a role akin total factor productivity, and occupational choices of individuals depend only on $\tilde{a}$, the comparative advantage in solving complex skills. How these choices depend on $\tilde{a}$ is a non-trivial question. To address this formally, it is convenient to approximate discrete occupational choices as a continuous choice over complexity x. This is generally not an innocuous assumption because the production function we will formulate below features inherently discrete occupation groups. However, given our focus on equilibria with PAM, and given that we consider the case with a large number of occupations J, the approximation bias can be bounded and will generally be small. More precisely, in section 5.4 we will show that for discrete occupations and for equilibria with the PAM property, the policy function $j_t^*(a, d)$ in equation (1) is a monotonic cutoff rule in $\tilde{a}$, that the corresponding wage function is a monotonically increasing step function that converges pointwise to the equilibrium wage function as $J \to \infty$ and that distribution-free bounds of the error can be computed from the data.[35] For the continuous approximation, where $x \in [0, 1]$, any solution of the optimization problem in (1) satisfies, for each

for sequences of consumption- and portfolio choices. As a consequence, if one is only interested in estimates of the model, specification of these decisions as we do here is not required. Instead, one could declare the user-cost of capital formulae as an exogenous input, as for example in Hubmer (2023). We specify the full equilibrium structure instead for two main reasons. First, it clarifies which assumptions we need to impose to use an inversion-type estimation approach as outlined below that is consistent with a general equilibrium model. For example, given our assumptions, the user-cost-of capital formulae are observational versions of the Euler-equations for portfolio choices and therefore an (observable) equilibrium outcome of our model. Second, it specifies the structure of the model that allows for computation of *equilibrium* counterfactuals after the structural model has been estimated.

[34] Lindenlaub (2017) studies and structurally estimates a quite general model of occupational matching between workers and firms with multidimensional heterogeneity on both sides of the labor market. Since her model is non-neoclassical, does not feature physical capital, and is not formulated to explain the time-series evolution of occupational skill premia, its setup is quite different from ours.

[35] See Proposition 5 below. Yamaguchi (2012), who studies occupational choices over the life-cycle in complexity space, uses the continuous approximation throughout his analysis.

period t, the following first-order condition:

$$\frac{\partial}{\partial x}\left(\ln h\left(x^*(\tilde{a}),\tilde{a}\right)\right) = -\Phi(x^*(\tilde{a})) \qquad \Leftrightarrow \qquad \frac{h_x(x^*(\tilde{a}),\tilde{a})}{h(x^*(\tilde{a}),\tilde{a})} = -\Phi(x^*(\tilde{a})), \tag{4}$$

where $x^*(\tilde{a})$ is the policy function.[36]

We will now derive a functional form for $h(\cdot)$ that generates a policy function $x^*(\tilde{a})$ with the following three desirable properties. First, it needs to be a maximizer. Second, it needs to have full support in the sense that $x^*(0) = 0$ and $x^*(\tilde{a}) \to 1$ as $\tilde{a} \to \infty$. Third, since we will focus on equilibria with the PAM property, and since $\tilde{a}$ is the comparative advantage in performing complex problems, $x^*(\tilde{a})$ should be strictly monotone and increasing. Whether these three conditions are satisfied will depend on the properties of the skill aggregator $h_t(.)$ and the complexity premium $\Phi_t(x)$, as can be seen from equation (4). Because $h_t(.)$ plays a role similar to match-specific output in models of bilateral matching with two-sided heterogeneity, and because various articles in this literature, such as Shimer and Smith (2000), Costinot (2009) and Lindenlaub (2017), have established log-supermodularity of the output function to be crucial for PAM, one might conjecture that the same applies to our context.[37] Indeed, the slope of $x^*(\tilde{a})$ will be characterized by the total differential of equation (4), which depends on the cross-partial $\frac{\partial}{\partial x \partial a}\ln h_t\left(.\right)$, and this cross-partial is positive if $h_t(.)$ is log-supermodular.[38] However, $\frac{\partial^2}{\partial x^2}\ln h_t\left(.\right)$ and $\Phi^{'}(x)$ also enter the total differential of equation (4), so that log-supermodularity is not sufficient for PAM in our model. We therefore need to impose additional assumptions.

We proceed as follows. First, we restrict $h_t(.)$ to be in the class of CES-aggregators:

$$h_t(x,a,d) = \left[\gamma_t(x)\cdot(a)^{\eta_t} + \delta_t(x)\cdot(d)^{\eta_t}\right]^{\frac{1}{\eta_t}}, \tag{5}$$

with $\eta_t \leq 1$. This is a standard specification in the literature on the production of human capital in the presence of multidimensional skills or investments, such as in Cunha *et al.* (2010) and Caucutt and Lochner (2020).[39] The functions $\gamma_t(x)$ and $\delta_t(x)$ are occupation-specific factor loadings and

[36]Here and in the following we will drop the index t whenever there is no loss of information.

[37]Supermodularity (rather than log-supermodularity) plays a key role in the literature on monotone comparative statics in optimization problems, such as Milgrom and Shannon (1994)). Athey (2002) establishes the importance of log-supermodularity for monotone comparative statics in models with uncertainty.

[38]This implication holds for the twice-differentiable case considered here.

[39]It is also ubiquitous in quantitative models of comparative advantage. For example, Gathmann and Schonberg

are assumed to be non-negative. If in addition $\gamma_t^{'}(x) > 0$ and $\delta_t^{'}(x) < 0$, conditions which we will show below to be necessary, then the higher the complexity of an occupation, the more weight it places on the a-skill. We can therefore think of a as an individual's innate ability for performing the higher-order problem solving associated with complex occupations. Similarly, d can be interpreted as the innate ability for performing the relatively simple problems and tasks associated with low complexity occupations.

Equation (4) also implies that the functions $\gamma_t(x)$ and $\delta_t(x)$ govern the marginal benefits and marginal costs that are relevant for optimal occupational choices. Consider a worker with skill endowment (a, d) who moves from occupation x to occupation $x + \varepsilon$, with ε arbitrarily small. This move has a market-wide rate of return of $\Phi(x^*(\tilde{a}))$, holding constant efficiency units, but it also carries with it an individual-level change in efficiency units of $\frac{\partial}{\partial x}\left(\ln h\left(x^*(\tilde{a}), \tilde{a}\right)\right)$. The latter comes from a marginal change in the weights the occupation with complexity $x + \varepsilon$ puts on the two skills. As we will show below, for PAM to arise in equilibrium, these two functions need to behave like benefit- and cost-functions. The following assumptions make this explicit:

Assumption 1. $\gamma_t(x)$ *is weakly concave and positive:* $\infty > \gamma_t(x) \geq 0,\ \ \gamma_t^{'}(x) > 0,\ \ \gamma_t^{''}(x) \leq 0$ *for all* $x \in [0, 1]$.

Assumption 2. $-\delta_t(x)$ *is strictly convex and negative:* $\infty > \delta_t(x) \geq 0,\ \ \delta_t^{'}(x) < 0,\ \ \delta_t^{''}(x) < 0$ *for all* $x \in [0, 1]$.

An interesting consequence of these two assumptions is the inequality $\frac{\gamma_t^{'}(x)}{\gamma_t(x)} > \frac{\delta_t^{'}(x)}{\delta_t(x)}$, which turns out to be a sufficient condition for $h_t(x, a, d)$ to be log-supermodular.[40]

Assumptions 1 and 2 are restrictions on the structural skill aggregation technology, but as indicated by equation (4), PAM also requires a restriction on equilibria that are realized:

Assumption 3. *The realized equilibrium skill price function features positive, but weakly decreasing complexity premia:* $\Phi(x) > 0$ *and* $\Phi^{'}(x) \leq 0$.

(2010) choose a formulation with $\eta = 1$ to study occupational mobility and wages, using the task-based approach to occupations. They measure $\gamma_t(x) \in [0, 1]$ as the share of time a worker spends on the task utilizing the first skill and consequently set $\delta_t(x) = 1 - \gamma_t(x)$. Biasi *et al.* (2025) also use a technology with $\eta = 1$ to model the comparative advantage of teachers in teaching high- and low-achieving students, and they use constant weights.

[40] See the proof of Proposition 1 in Appendix B.

This restriction comes with little loss of generality in our context since we use our model as a quantitative framework for understanding the positive complexity premium documented in section 3, and since we rely on an inversion-type method for identification and estimation. Rather than solving for the endogenous skill-price function, we will recover it from the data, and assumption (3) can therefore be interpreted as an identifying assumption. More specifically, suppose one indexes sequences of equilibria by some initial condition and all structural parameters of the model. Conditional on the indexing, such equilibria will be unique because our model is neoclassical without any source of non-convexities and without uncertainty. Assumption 3, combined with assumptions 1-2 and equation (4), states that we only consider the set of indices that deliver equilibria with the PAM-property. Clearly, there will be initial conditions and structural parameters for which this will not be the case, but these are not the ones that have been realized. The assumption is therefore a restriction on the map from data to parameters, and it is crucial for identification if one wants to avoid a full-solution method. It is also testable, and in section 6.2 we show that it is strongly supported by the empirical evidence.

This leads to our first proposition:

Proposition 1. *Under assumptions 1 and 2, the skill aggregator $h_t(x, a, d)$ is log-supermodular, and $h_t(x, a, d) \geq 0$ on its support. If in addition assumption 3 and the full-support condition $x_t^*(0) = 0$ are satisfied, then the following is true:*

i. (a, d) cannot be too complimentary: $\eta \in (0, 1]$.

ii. Occupational choices satisfy PAM: $\frac{\partial x_t^*(\tilde{a})}{\partial \tilde{a}} > 0$ for all $\tilde{a} \geq 0$.

Proof. See the appendix. □

A converse is that an aggregator with linear weight functions, as for example in Gathmann and Schonberg (2010), will not generate PAM in general. Mathematically the reason is that $x^*(\tilde{a})$ maximizes $\pi_t(x) \cdot h_t(x, \tilde{a})$, and this objective function needs to be sufficiently concave for the solution to be unique.[41] Assumptions 1 to 3 are sufficient for this to be the case.

One particularly important implication of Proposition 1 is that task complexity is a perfect proxy for the comparative advantage in solving complex problems because it implies that one can

[41] Here and in the following it is important to note that conditional on $\tilde{a}$, the $d-$skill is irrelevant for optimal choices.

invert optimal observed x^* to the corresponding unobserved $\tilde{a}$. We will write this inverse policy function as $\tilde{a}_t^*(x)$ in the following. Solving equation (4) and using our assumption on $h_t(x, a, d)$ yields

$$\tilde{a}_t^*(x) = \left[\frac{-\left(\delta_t^{'}(x) + \eta \cdot \Phi_t(x) \cdot \delta_t(x)\right)}{\gamma_t^{'}(x) + \eta \cdot \Phi_t(x) \cdot \gamma_t(x)} \right]^{\frac{1}{\eta}}, \tag{6}$$

and the corresponding efficiency units are given by

$$h_t\left(x, \tilde{a}_t^*(x)\right) = \left[\frac{\left(\delta_t(x) \cdot \gamma_t^{'}(x) - \delta_t^{'}(x) \cdot \gamma_t(x)\right)}{\gamma_t^{'}(x) + \eta \cdot \Phi_t(x) \cdot \gamma_t(x)} \right]^{\frac{1}{\eta}}. \tag{7}$$

Notice in particular that, holding everything else constant, a uniform increase in the complexity premium $\Phi_t(x)$ decreases the efficiency units of labor that are optimally assigned to complexity x. In the presence of PAM, an increase in the return to choosing a higher level of complexity lowers the comparative advantage that optimally chooses that level. This property will be crucial for the identification of the skill aggregation technology, as will be shown below.

Equation (6) is a closed-form policy function for occupational choices that, for each t, is a one-to-one relationship between worker skills and occupational complexity. The equilibrium allocation of comparative advantages in solving complex problems to occupations is therefore deterministic as our model does not feature non-monetary preference components or informational- or search- and matching frictions.[42] We do not build these components into our model for two main reasons. First, we study the long-run *evolution* of the cross-sectional distributions of occupational wages and employment, for which labor market frictions are unlikely to be a quantitatively important mechanism. Second, we will show in section 6.4.2 that our model can match very well both of these distributions in all years, implying that adding non-monetary preference components would only marginally improve the explanatory power of our model.[43]

[42] Costrell and Loury (2004) formulate an assignment model in which workers are matched to jobs that are ranked by their sensitivity to worker ability. In the equilibrium of their model, PAM arises as well, and the mapping from worker skills to jobs is one-to-one, as in our model.

[43] It is worth noting that introducing occupational taste shocks is relatively straightforward. Suppose one assumes that preferences feature an additively separably individual-level taste shock, as for example in Diamond (2016). One can solve the individual's choice problem using multi-stage budgeting, as before. If the taste component is Gumbel distributed, then occupational choices will be of the logit form, conditional on $\tilde{a}$. We do not consider this extension here because we can match the employment distribution near-exactly. See section 6.4.2.

4.3 Firms

Firms combine two types of capital, structures and equipment, with efficiency units of labor in J distinct occupations to produce output y_t. They choose the cost-minimizing input mix that generates target revenue R_t as follows:[44]

$$\min \sum_{j=1}^{J} \overline{w}_{j,t} \cdot l_{j,t} + r_t^e \cdot k_t^e + r_t^s \cdot k_t^s$$

subject to

$$\begin{aligned} p_t \cdot y_t &= R_t \\ y_t &= A_t \cdot F_t \left(l_{1,t}, ... l_{J,t}, k_t^e, k_t^s\right). \end{aligned} \tag{8}$$

As before, $\overline{w}_{j,t} = \pi_t(x_j)$ by definition. The production function is defined recursively as follows:

$$\begin{aligned} F\left(l_{1,t}, ... l_{J,t}, k_t^e, k_t^s\right) &= \left(k_t^s\right)^{\alpha_t} \cdot \left\{ \left[\lambda_{1,t} \cdot \left(l_{1,t}\right)^{\sigma_1} + \left(\widetilde{l}_{2,t}\right)^{\frac{\sigma_1}{\sigma_2}} \right]^{\frac{1}{\sigma_1}} \right\}^{1-\alpha_t}. \\ &\text{with} \\ \widetilde{l}_{j,t} &= \begin{cases} \lambda_{j,t} \cdot \left(l_{j,t}\right)^{\sigma_j} + \left(\widetilde{l}_{j+1,t}\right)^{\frac{\sigma_j}{\sigma_{j+1}}} & \text{for } j \in \{2, ..., J-1\} \\ \lambda_{j,t} \cdot \left(l_{j,t}\right)^{\sigma_j} + \lambda_{k,t} \cdot \left(k_t^e\right)^{\sigma_j} & \text{for } j = J. \end{cases} \end{aligned} \tag{9}$$

where we assume that the structural parameters governing capital-labor substitutions, σ_j for $j = 1, ..., J$, are constant over time. This production function is a generalization of the specification in the seminal paper on capital-skill complementarity in Krusell *et al.* (2000) to a setting with J discrete and distinct labor inputs, where $J \ll \infty$ can be large.[45] In particular, output is produced using a Cobb-Douglas production function with two inputs, capital structures k_t^s and a CES-aggregator of the remaining inputs. For the case with arbitrary J the CES-aggregator in turn needs to be written recursively, with capital equipment in the last step of the iteration. This non-standard feature of the production function is required to allow for occupation-specific capital-labor substitutability. On

[44] As will be shown in section 5.6 the formulation in terms of cost-minimization instead of profit maximization is important for parameter identification, as long as one has a measure of R_t, such as value added, in the data.

[45] In Krusell *et al.* (2000), the $\lambda_{j,t}$'s enter as weights and take values in $[0, 1]$. Applied to our production function this would mean that $\lambda_{j,t}$ multiplies $(l_{j,t})^{\sigma_j}$ and $(1 - \lambda_{j,t})$ multiplies $\left(\widetilde{l}_{j+1,t}\right)^{\frac{\sigma_j}{\sigma_{j+1}}}$. We do not choose this specification because the recursive nature of the production function means that the actual weight for a labor input approaches zero as j grows because it is not only composed of $\lambda_{j,t}$ itself, but all the $\lambda_{m,t}$ for $m < j$. Instead, as we will show below, identification requires that one of the λ's is normalized. We will set $\lambda_{k,t} = 1$ so that the $\lambda_{j,t}$'s have the interpretation of factor productivities relative to equipment capital.

the other hand, common features of neoclassical production functions are preserved. First, as can be checked easily, it has constant-returns-to-scale in inputs $(l_{1,t}, ...l_{J,t}, k_t^e, k_t^s)$. Second, there is no need to introduce occupation-specific capital, facilitating measurement. Instead, as in standard production theory, the aggregate capital stock combines with heterogeneous labor to produce aggregate output. The novelty of our specification is that it allows for many types of labor inputs and for full heterogeneity in capital-labor substitutability. Third, the elasticity of substitution between labor employed in task j and capital k_t^e is $\frac{1}{1-\sigma_j}$ and is invariant to the order in which they appear in the production function. This result is an implication of the following Proposition, where $F_{j,t}(.) = \frac{\partial F(.)}{\partial l_{j,t}}$ is the marginal product of type-j labor in period t, and where $F_{k^e,t}(.)$ is the marginal product of capital equipment.:

Proposition 2. *Let $j = 2, 3, ..., J$, with $J \ll \infty$, and define the list $\omega_{j,t} = \left\langle \left\{ \lambda_{m,t}, \left(\frac{k_t^e}{l_{m,t}}\right) \right\}_{m=j+1}^{J}, \{\sigma_l\}_{l=j}^{J} \right\rangle \in \Omega_j$, where $\Omega_j = \left(\mathbb{R}_0^+\right)^{2(J-j)} \times (-\inf, 1]^{J-j+1}$ is of dimension $3\cdot(J-j)+1$. Let $MRTS(l_{j,t}, k_t^e) = \frac{F_{j,t}(.)}{F_{k^e,t}(.)}$ be the marginal rate of technical substitution between type-j labor and equipment capital in period t,* ***holding constant all other factor inputs****. We then have:*

$$MRTS(l_{j,t}, k_t^e) \quad = \quad \lambda_{j,t} \cdot \left(\frac{k_t^e}{l_{j,t}}\right)^{1-\sigma_j} \cdot G_j(\omega_{j,t}) \tag{10}$$

for a list of functions $\{G_m\}_{m=j}^{J}$.

Proof. See the appendix. □

This proposition implies that for any j the MRTS between type-j labor in equipment capital is a function of the structural parameters and the list of capital-labor ratios $\left\{\frac{k_t^e}{l_{m,t}}\right\}_{m=j}^{J}$. That is, none of the production inputs appear in any other form than capital-labor ratios. As a consequence, the elasticity of substitution between type-j labor and equipment capital is defined as an elasticity that holds constant the capital-labor ratios for all types m of labor with $m > j$. Using proposition 2 it is easily checked that this elasticity is given by $\frac{1}{1-\sigma_j}$, as claimed ealier. Furthermore, as long as one holds constant the correct set of capital-labor ratios, this elasticity is neutral to reordering labor types in the production function. The only factor that always needs to have the same position in the nested production function is k_t^e, and it always needs to appear in the final nest.

At cost-minimizing input choices, where factor prices are equal to marginal products, the proposition delivers the following reduced-form regression:

$$\log\left(\frac{\overline{w}_{j,t}}{r_t^e}\right) = \Lambda_{j,t} + (1-\sigma_j)\cdot\log\left(\frac{k_t^e}{l_{j,t}}\right) \tag{11}$$

The coefficient on the capital-labor ratio is structural and is the inverse of the elasticity of substitution between type-j labor and capital equipment, but the intercept $\Lambda_{j,t}$ is a reduced form of all capital-labor ratios for labor indexed by $m > j$ and the structural parameters. There are various reasons for the importance of this equation. First, it establishes a tight connection between our framework and the literature estimating the extent of capital-labor substitution. Indeed, regressions of this form are ubiquitous in the literature, examples of which are Karabarbounis and Neiman (2014), Oberfield and Raval (2021) and Hubmer (2023). The difference to these papers is that our equation applies occupation-by-occupation and requires careful conditioning on other factor ratios. Second, the equation has direct implications for identification of the substitution elasticities. Most importantly, it is subjected to the same well-known fundamental identification problem as established in Diamond *et al.* (1978), because the fixed effect $\Lambda_{j,t}$ varies on the same order as the explanatory variable $\log\left(\frac{k_t^e}{l_{m,t}}\right)$. There are two structural sources of this non-identification result. One is that the reduced-form object $\Lambda_{j,t}$ contains the unrestricted structural occupation-specific time trends of factor productivities, given by $\lambda_{j,t}$. This is also the source of non-identification emphasized in Diamond *et al.* (1978). The other source comes from the fact that any change in the explanatory variable $\log\left(\frac{k_t^e}{l_{m,t}}\right)$ leads to an endogenous adjustment of the other capital-labor ratios appearing in $\Lambda_{j,t}$. We will adopt the standard solution to this problem and will impose parametric restrictions on the non-linear time-trend $\lambda_{j,t}$ and will estimate the parameters structurally, an issue we will discuss further below.

Further Discussion of Production Function Two recent papers, by Burstein *et al.* (2019) and Caunedo *et al.* (2023), formulate quantitative macro models with occupations and physical capital, but use a structure in which each occupation produces an intermediate good using occupation-specific capital- and labor inputs according to a technology $y_{j,t} = \left[\left(k_{j,t}^e\right)^{\sigma_j} + \lambda_{j,t}\cdot(l_{j,t})^{\sigma_j}\right]^{\frac{1}{\sigma_j}}$. Intermediate inputs aggregate to final output by way of $y_t = F_t\left(k_t^s, y_{1,t}, ..., y_{J,t}\right)$. This technology also features occupation-specific capital-labor substitutability. The major difference to our production function, which to the best of our knowledge is novel, is how equipment capital enters production. In our specification, all occupations share the same capital stock, while efficiency units of labor are occupation specific. In contrast, the technology with occupation-level intermediate output assumes that there is occupation-specific capital equipment.

To clarify the conceptual differences, consider the example of a car repair shop employing workers either of two occupations, unskilled labor and car mechanics (both of which are Census occupational categories). Both, our problem-based approach to occupations and the task-based approach applied by Burstein *et al.* (2019) and Caunedo *et al.* (2023) views the two types of labor inputs as fundamentally different. Unskilled labor in a car repair shop typically performs basic maintenance like oil changes, tire rotations, and shop cleaning, while car mechanics carry out skill intensive diagnostics and engine repairs. Our PCA on task measures assigns the latter a substantially higher complexity score, while the task-based approach gives it a higher "non-routine" score. However, the two production functions differ substantially in the way in which the two types of labor inputs combine with physical capital to produce shop-level output. Our technology assumes that the two groups produce the final good offered by the car repair shop by working on problems of different complexities using the same stock of capital equipment, such as machines and tools. In contrast, the specification in Burstein *et al.* (2019) and Caunedo *et al.* (2023) assumes that each group of workers is assigned to different production processes that produce a differentiated good using completely separate stocks of physical capital. The two differentiated goods then aggregate into the final output offered by the car repair shop, and the capital stocks in the two production processes add up to the shop-level capital stock. Real-world production technologies arguably have features of both specifications, and we see no a priori reason to prefer one assumption over the other. We choose our approach for two reasons: First, given our focus on skill premia, our production function is a direct extension to the quantitative literature on capital-skill complementarity which relies on the technology in Krusell *et al.* (2000). Second, it avoids several measurement issues, such as computing occupation-specific capital inputs and the levels- and relative prices of intermediates.

4.4 Equilibrium

Define $\mathbf{l}_t = (l_{1,t}, ...l_{J,t})$ and $\mathbf{k}_t = (k_t^e, k_t^s)$, and let $\tilde{G}_t(\tilde{a})$ be the marginal distribution of $\tilde{a}$ generated by $G_t(a, d)$. An equilibrium of the model economy is a sequence of prices, $\{(\pi_t(x), p_t, \mathbf{r}_t)\}_{t=0}^{\infty}$ with $\overline{w}_{j,t} = \pi_t(x_j)$, factor demand levels $\{(\mathbf{l}_t, \mathbf{k}_t)\}_{t=0}^{\infty}$, and individual policy functions $\{(j_t^* (\widetilde{a}), \mathbf{s}_t (a, d))\}_{t=0}^{\infty}$ and a list of aggregate outcomes $\{(C_t, y_t)\}_{t=0}^{\infty}$ such that for each t:

1. $\{(j_t^* (\widetilde{a}), \mathbf{s}_t (a, d))\}_{t=0}^{\infty}$ solves the individual maximization problem.
2. $\{(\mathbf{l}_t, \mathbf{k}_t)\}_{t=0}^{\infty}$ are the cost-minimizing levels of factor demand, given $\{(\overline{w}_t, p_t, \mathbf{r}_t)\}_{t=0}^{\infty}$ and $\{(C_t, y_t)\}_{t=0}^{\infty}$.
3. Markets clear:
   (a) **(Goods)** $p_t \cdot y_t = p_t \cdot c_t + \mathbf{q}_t \cdot (\mathbf{k}_t - (1 - \mathbf{d} \circ \mathbf{k}_{t-1}))' + (b_t - \overline{r}_t \cdot b_{t-1})\,.$
   (b) **(Labor)** $l_{\chi,t} = \int_0^{\infty} h_t (x_\chi, a, d) \cdot dG_t (a, d : j_t^* (\widetilde{a}) = \chi)$ for all $\chi = 1, ..., J$.
   (c) **(Capital)** $s_t^n = k_t^n$ for $n = \{e, s\}$

5 Identification and Estimation

5.1 Preliminaries

We derive identification results for and estimators of the model's structural parameters under the assumption that, in each period covered by the data, the economy is observed in equilibrium and that the corresponding equilibrium prices $(\pi_t(x), p_t, \mathbf{r}_t)$ can be treated as "observables". This avoids solving for the recursive general equilibrium and allows us to adopt similar ideas as in e.g. Bajari *et al.* (2007) for IO and in Kleinman *et al.* (2023) for spatial macroeconomics to our context. This approach is unproblematic for asset markets since with data on asset prices and depreciation rates, the Euler equations in (3) invert to the rental rates $\mathbf{r}_t$. It is more problematic for labor markets since the equilibrium price schedule is in terms of efficiency units while observed wages are contaminated by selection. We therefore estimate skill price *growth* using the "Flat Spot Method" of Heckman *et al.* (1998) modified to the case with many occupations. This is the only model object that we "pre-estimate" outside the model.

In a second step, we estimate a parsimonious parameterization of the remaining model objects using data on occupational labor shares, capital inputs, hourly wages, and rental rates of capital. The unknown parameters are the structural parameters of the skill distribution $G_t(a,d)$, the skill aggregator $h_t(a,d,x)$, and the production function (all of which are structural), together with the initial condition of the skill-price function. Since we estimate equilibrium prices from the data, identification can be achieved without knowing individuals' preferences. Moreover, we do not need to make any assumptions about whether the economy is in steady state or not. However, the order in which one estimates parameters from the data is important. In particular, one needs to know the skill price function $\pi_t(x)$ before estimating the parameters of $h_t(a,d,x)$, and one needs to know $h_t(a,d,x)$ and the rental rates before recovering the parameters of the production function in turn. We will discuss identification of the structural parameters and their calibration in the order in which they need to be performed. Throughout we will assume that we have asymptotically valid, repeated cross-sectional data indexed by j for occupations and $t \in \{0,1,...,T\}$ for calendar time on the observables of the model, namely aggregated factor inputs or aggregate prices. By "asymptotically valid data" we mean that the empirical analogues of population aggregates are consistent estimators and that individual-level variation is averaged out. Hence, we derive identification results from deterministic aggregate relationships, with the exception of the Flat Spot Method, which we implement on individual-level data.

5.2 Measuring the Price of Complexity: The Flat Spot Method

Realized wages of an individual with skill endowment (a,d) who chooses occupation j in our model are equal to the highest income it can achieve. By definition, this is the value function of the optimization problem in equation (1):

$$M_t(a,d) \equiv M_t\left(\tilde{a}_t^*(x_j), d\right) = \pi_t(x_j) \cdot h_t\left(x_j, \tilde{a}_t^*(x_j)\right) \cdot d, \tag{12}$$

where $\tilde{a}_t^*(x)$ is the inverse policy function.[46] Let i be the index for an individual worker, and let $w_{i,j,t}$ be its observed wage in t, given that it has chosen occupation j. The assumption that data

[46]Proposition 1 implies that this inverse exists on $x \in [0,1]$.

are observed in equilibrium implies that

$$w_{i,j,t} = M_t\left(\tilde{a}_t^*(x_{i,j}), d_i\right). \tag{13}$$

Observed wages are therefore the product of the aggregate skill price and a selection term that is equal to efficiency units in the optimally chosen occupation. As a consequence, the skill-price function $\pi_t(x)$ is not identified from occupational wage data without additional restrictions.[47] One approach to overcome this identification problem is imposing strong exogeneity restrictions on labor supply choices, as for example in Card and Lemieux (2001) and

To describe this method, let τ be a time-index that varies on the annual level, and suppose one finds a group of workers such that $h_\tau(x,a,d) = h_{\tau+1}(x,a,d)$.[48] Then the difference in observed log-wages $(\ln w_{\tau+1} - \ln w_\tau)$ for these workers identifies the evolution of $\pi_\tau(x)$ *up to its initial level in some base period.* The flat-spot method relies on the theoretical prediction from human-capital theory that there is a point in the life cycle of a worker during which the increase in efficiency units $h_\tau(.)$ from human capital investments exactly balances the decrease in efficiency units from skill depreciation. At that point of the life cycle, the causal experience-earnings profile net of aggregate skill price changes becomes flat, and changes in earnings from one period to the next are driven, on average, by the change of the price of these efficiency units $\pi_\tau(x)$. It is important to note that the flat-spot assumption concerns the *life-cycle evolution* of the individual skill component from one year to the next for a particular age group of workers and is therefore not contradictory with the possibility that the skill aggregation technology changes over time in the *cross-section* of all workers. At the same time, it is possible that the flat-spot region shifts over time, a possibility we address below when estimating aggregate skill prices.

Conditional on a group of workers being in the flat spot of their experience-earnings profile, and conditional on observed occupational complexity x, the time-series of skill prices is non-

[47] Generally, the non-identification does not depend on whether one uses an estimation approach that avoids equilibrium computation, as we do, or whether one solves for the entire equilibrium and matches predicted wages to their observed counterparts.

[48] We use the index τ rather than t here because the latter will be reserved for the analysis on decennial data. In contrast, we implement the flat-spot method on annual data.

parametrically identified from the equation

$$\ln \pi_\tau(x_j) - \ln \pi_{\tau-1}(x_j) = E\left[\ln(w_{i,\tau,x}) - \ln(w_{i,\tau-1,x}) | \, x, (\tau-1), i \in flatspot\right], \qquad (14)$$

which follows from equations (12) and (13) and the flat-spot assumption

$$E\left[\ln(d_{i,\tau} \cdot h_\tau\left(x_{i,j}, \tilde{a}^*_\tau(x_{i,j})\right)) - \ln(d_{i,\tau-1} \cdot h_{i,\tau-1}\left(x_{i,j}, \tilde{a}^*_{\tau-1}(x_{i,j})\right) | x, (\tau-1), i \in flatspot\right] = 0. \quad (15)$$

We index random variables by (x, i) to emphasize that these equations condition on occupational complexity and that we will estimate them on individual-level wage data. Furthermore, the population of workers in the data needs to be in the flat-spot for identification, which we indicate using the notation "$i \in flatspot$". If for each calendar year there is exactly one age group in the flat spot, then this conditioning can be expressed as a cohort-level quasi-panel regression in first differences, a result we utilize in the implementation of the method. Bowlus and Robinson (2012) provide ample evidence that the flat spot is reached after about 30 years of labor market experience.

In its ideal implementation, the flat-spot method is a non-parametric quasi-panel regression, where panels are defined by cohorts and each cohort is followed for exactly two years. It delivers biased estimates of the evolution of $\pi_\tau(x)$ if, within cohorts, the individual efficiency term $d \cdot h_\tau\left(x_j, \tilde{a}^*_\tau(x_j)\right)$ does not remain constant during the two years defining the flat spot. Possible reasons are a change of the composition of workers between these two years, a poor choice of the age at which the flat spot is reached in the sense that skill appreciation- and depreciation are not in balance, imperfect substitution between age groups in the production function as in Card and Lemieux (2001), or changes of the age-experience profile over time. While we follow as much as possible the seminal study on this method by Bowlus and Robinson (2012), which addresses some of these points, we carry out various robustness checks in section 6.2 to address these issues.

Implementation Equation (14) can be estimated by replacing population conditional moments by their sample analogues. However, in our context this is too flexible for at least two reasons. First, any estimator of $(\ln \pi_\tau(x_j) - \ln \pi_{\tau-1}(x_j))$ will likely be affected by short-to medium run fluctuations in aggregate conditions and by sampling noise. Since the corresponding estimates will be an input into our structural estimation below, this variability will be passed on to the estimates of our other parameters. Second, in contrast to usual applications of the flat-spot method to educational premia, where x indexes few groups only, in our context the group variable is occupational complexity and thus continuous. We find that the following parameterized version of equation (14), formulated in levels, captures the most salient features of the observed evolution of occupational wages well:

$$log(w_{i,\tau,x}) = \mu_x + \sum_{\iota=0}^{T} \left[\mathbf{1}(\iota = \tau) \cdot (\beta_{\iota,0} + \beta_{\iota,1} \cdot x)\right] + \epsilon_{i,\tau,x}. \tag{16}$$

The index ι on the summation operator ranges through all time periods observed in the data. Hence, the coefficients in $\{\beta_{\iota,0}\}_{\iota=0}^{T}$ are time fixed effects, and the parameters in $\{\beta_{\iota,1}\}_{\iota=0}^{T}$ are the coefficients on a linear term in complexity x which are allowed to vary freely over time. Since we use log-wages as dependent variable the latter have the interpretation of aggregate returns to complexity and therefore correspond to the equilibrium object $\Phi_\tau(x)$. Equation (16) imposes the restriction that the $\Phi_\tau(x)$ are constant in x.[49] We chose this specification after extensive pre-testing. Most importantly, our analysis in section 3 did not uncover any evidence for non-linearities in the complexity premium as a function of x, and we find the same for skill-prices estimated using the flat-spot method. We will present various validation exercises when presenting results in section 6.2 that support this assumption. We will maintain it for the rest of the paper and state it here:

Assumption 4. *The complexity premium in period t does not depend on x: $\Phi_t(x) = \Phi_t$ for all $x \in [0,1]$ and for all t. It is allowed to vary freely over time.*

Regression equation (16) is a parametric version of the Flat-Spot regression model in equation (14), but formulated in log-wage levels rather than changes. Inclusion of occupation fixed effects μ_x guarantees that the remaining parameters continue to be identified from within-occupation variation only. As a consequence, just like the non-parametric model, equation (16) pins down only the evolution of the complexity premium over time, but not its levels. More specifically, the complexity premium in the baseline period, say $\beta_{0,1}$, is not identified, and for any subsequent periods, corresponding to $\iota > 0$, the $\beta_{\iota,1}$ are the complexity premium relative to this baseline period. Since $x \in [0,1]$ they are also the difference-in-difference (DiD) in the aggregate skill price $\pi_\tau(x)$ between the most- and the least complex occupations and between period τ and the baseline period. The

[49]The assumption that the complexity premium does not depend on the level of complexity is analogous to the assumption that returns to education do not depend on the level of education.

$\beta_{\iota,0}$ will need to be identified differently, as discussed below.

5.3 The Rental Rates of Capital

A second set of equilibrium prices are the prices of equipment capital and capital structures, $\mathbf{q}_t$, and the corresponding rental rates of capital, $\mathbf{r}_t$. Our assumption that capital and bonds are traded in competitive international markets (and that $\mathbf{q}_t$ and the risk-free rate are therefore exogenous) implies that the $\mathbf{r}_t$ can be identified from the optimality condition in equation (3), given data on capital prices, depreciation rates, and the risk free rate. This condition coincides with the standard user cost of capital formulae, and we follow Hubmer (2023) in measuring user costs from the data, the details of which are provided in Appendix C.2. We also plot our measured user cost of equipment capital in figure A.2, which visualizes the substantial reduction in the user cost of equipment capital between 1980 and 2000 and the much smaller reduction in user costs thereafter.

5.4 Skill Aggregation Technology and the Reduced-Form Wage Function

In this section we parameterize the functions $(\gamma_t(x), \delta_t(x))$ that enter the skill-aggregation technology $h_t(x, a, d)$, establish identification from the reduced-form wage equation generated by the model, and describe parameter estimation.

Parameterization The skill aggregator is one of the two nonstandard features of our model, and therefore there is little guidance in the literature on how to parameterize it, apart from the restriction on the CES-family and the conditions stated in assumptions 1 and 2. We choose a parsimonious specification and restrict our search to polynomials and exponentials. This leads us to the following set of parametric assumptions:

Assumption 5. *The factor loadings of the skill aggregator $h_t(x, a, d)$ are given by $\gamma_t(x) = (x + \gamma_{0,t})$ and $\delta_t(x) = \delta_{1,t} - (x + \delta_{0,t})^{c_t}$, with $(\gamma_{0,t}, \delta_{0,t}, c_t, \eta_t) > 0$ and $c_t > 1$. Furthermore, $c_t = c$ and $\eta_t = \eta$ for all t, and $\delta_{1,t} = (\delta_{0,t})^c \cdot \left(1 + \frac{c}{\delta_{0,t} \cdot \eta \cdot \Phi_t}\right)$.*

This specification implies that $h_t(x, a, d)$ has 5 structural parameters $(\gamma_{0,t}, \delta_{0,t}, \delta_{1,t}, c_t, \eta_t)$ for each t. Furthermore, the sequence of parameters $\delta_{1,t}$ is not free because of the full-support condition $\tilde{a}_t^*(0) = 0$, which is equivalent to the final equation in assumption 5.[50]

[50]This equation is derived from evaluating equation (6) at $x = 0$. The full-support condition is important because

The parameters $(\gamma_{0,t}, \delta_{0,t})$ are required to control the surjectivity of $x^*(\tilde{a})$ on $[0, 1]$. To illustrate this, if $\gamma_{0,t} = 0$ then the marginal benefit of choosing an occupation with complexity close to zero will be close to zero as well, and such occupations will never be optimal for any level of $\tilde{a}$. Similarly, as $\delta_{0,t} \to 0$ the marginal cost of choosing a complexity near zero will become arbitrarily large, also implying that such occupations will never be chosen. Finally, the parameter family c_t governs the curvature of the marginal cost function and, as we will show, is crucial for the PAM property to hold so that $x^*(\tilde{a})$ is injective.[51]

Assumption 5 forces some of the parameters to be constant over time. We choose these restrictions to reduce the size of the parameter space. For example, we view $\eta_t = \eta$ and $c_t = c$ as "deep" technology parameters that are unlikely to be affected by aggregate technological change. On the other hand, recent research on task usage within 3-digit occupations, such as Autor *et al.* (2024), suggests that the mapping from skills to occupational efficiency units may have changed over time. We allow for this possibility by treating $\gamma_{0,t}$ and $\delta_{0,t}$ as free parameters. Conditional on x, their evolution over time represents technological progress in the utilization of worker-level comparative advantage in solving complex tasks, $\tilde{a}$. As we will show below, it is crucial for explaining why skill prices grew at a slower rate than raw wages.

Our discussion of the role of the parameters is reflected in the following Proposition:

Proposition 3. *Let $h_t(x, \tilde{a}) = [\gamma_t(x) \cdot (\tilde{a})^\eta + \delta_t(x)]^{\frac{1}{\eta}}$, with $x \in [0, 1]$, $\eta \in (0, 1]$, and $\tilde{a} \geq 0$. Also assume that assumptions 4 and 5 are satisfied. A sufficient condition for occupational choices to have the PAM property is $\gamma_{0,t} > \delta_{0,t}$. If in addition $c > 2$, then the value function $M_t(\tilde{a}_t^*(x), d)$ is strictly increasing in x.*

Proof. The PAM property can be established directly by checking that the functional forms for $\gamma_t(x)$ and $\delta_t(x)$ imply the conditions in assumptions 1 and 2. Since assumption 3 is implied by 4, PAM follows directly from proposition 1. On the other hand, the implication of $c > 2$ stated in proposition 3 is harder to establish and is proven in the appendix. $\square$

The statement in the proposition that equilibrium wages, which are equal to $M_t(\tilde{a}_t^*(x), d)$, are

we observe in the data positive employment shares over the entire support of occupational complexity in all periods, and the model needs to be able to replicate this fact.

[51] Generally, for any of the occupations to be chosen in equilibrium, either the marginal benefit function $\gamma_t(x)$ needs to be concave or the marginal cost function $-\delta_t(x)$ needs to be convex. The simplest parameterization to achieve this is the one chosen in the assumption above, with $c > 1$.

an increasing function of x only for sufficiently large c is motivated by the finding of a positive correlation between task complexity and raw wages. Interestingly, this is not necessarily the case in theory. The reason is that moving up marginally in complexity space creates a benefit, since more weight is put on the first skill, but also a marginal cost, since the second skill is used less. Wages may decline if the second effect is too strong, which will indeed be the case if $c \in (1, 2]$.

Identification We now explore the identification of the parameters of the skill aggregator, $\left(\{\gamma_{0,t}, \delta_{1,t}\}_{t=0}^{T}, \delta_0, c, \eta\right)$, and of the complexity premium in the base period, Φ_0, from wages. This identification must come from the reduced-form wage equation (13), taking as given observed wages and the parameters of the skill price function that are identified from the flat-spot method. To start with, notice that the equation implies that

$$E_t[\ln w_t\left(x, \tilde{a}_t^*(x)\right)|\tilde{a}, x] = E_t[\ln(d)|\tilde{a}] + \ln \pi_t(x) + \ln h_t\left(x, \tilde{a}_t^*(x)\right), \tag{17}$$

where expectations E_t are taken over the distribution of d given $\tilde{a}$. Conditional on d, there is no source of randomness left, and the conditional expectation function becomes an exact relationship. It is immediately clear from equation (17) that $E_t[\ln(d)|\tilde{a}]$ and $\ln h_t\left(x, \tilde{a}\right)$ are not separately identified unless one either has a direct measure of the skill components $(d, \tilde{a})$, or one treats d as an error component and matches higher-order moments of the wage-complexity distribution as well. We do not pursue either avenue here because (a) our focus is on the *average* complexity premium rather than its full distribution and (b) a central motivation of our paper is to demonstrate that a model with one skill, the comparative advantage of solving complex tasks, can explain *quantitatively* the observed changes in the occupational wage and employment distribution.[52] We therefore impose the following identifying assumption:

Assumption 6. *The model features* ***Pure Comparative Advantage*** *in the sense that skill heterogeneity is fully described by $\tilde{a}$. In particular, $d = 1$.*

The empirical relationship between observed occupational wages and the model prediction, which

[52] It is important to note that the underlying micro-founded model delivering this model of comparative advantage needs to have two skills. Otherwise one cannot generate a fully non-degenerate distribution of equilibrium choices over occupations.

is the **reduced form wage equation**, is thus given by

$$\ln w_{j,t} \quad = \quad \ln \pi_t(x_j) + \ln h_t\left(x_j, \tilde{a}_t^*(x_j)\right), \tag{18}$$

where $h_t\left(x_j, \tilde{a}_t^*(x_j)\right)$ is given by equation (7). Using Assumption (4) we can write, for each t,

$$\ln \pi_t(x_j) = \beta_{t,0} + \Phi_t \cdot x_j. \tag{19}$$

As discussed in section 5.2, the Flat Spot Method identifies the complexity premium Φ_t relative to the base period indexed by $t = 0$ and an aggregate time trend $\beta_{t,0}$ up to normalization. We therefore can write

$$\Phi_t = \Phi_0 + \beta_{t,1}, \tag{20}$$

normalize $\beta_{0,0} = 1$ and treat the $(\beta_{t,0}, \beta_{t,1})$ as knowns.[53] Combining equation (7) with the functional forms in Assumption (5) yields the following reduced-form parametric wage equation for all t:

$$\ln w_{j,t} - \beta_{t,0} - \beta_{t,1} \cdot x_j = \Phi_0 \cdot x_j + \frac{1}{\eta} \cdot \ln\left[\frac{\delta_{1,t} + (\delta_{0,t} + x_j)^{c-1} \cdot \left[c \cdot (\gamma_{0,t} + x_j) - (\delta_{0,t} + x_j)\right]}{1 + \eta \cdot (\Phi_0 + \beta_{t,1}) \cdot (\gamma_{0,t} + x_j)}\right]. \tag{21}$$

The left-hand side of this equation consists of data and parameters that are identified from the Flat Spot Method, while the right-hand side involves all unknown parameters and data. Notice in particular that the complexity premium in the base period, Φ_0, needs to be treated as an unknown parameter. The next Proposition states that all structural parameters are locally identified from wage data:

Proposition 4. *Let $h_t(x, \tilde{a})$ satisfy the assumptions of proposition 3 and assume that our model features pure comparative advantage in the sense of assumption 6. Take as given the parameters identified from the Flat Spot Method. Also assume that the data satisfy the following two conditions: (a) the data are generated by equilibria satisfying assumptions 4 to 6 and (b) there exist lower bounds $\underline{\eta}$ and $\underline{\Phi}$ such that $\eta \geq \underline{\eta}$ and $\Phi_0 \geq \underline{\Phi}$, where $0 < \underline{\eta} < 1$. Then there is a locally invertible map from data to the structural parameters of the skill aggregator, $\left(\{\gamma_{0,t}, \delta_{0,t}, \delta_{1,t}\}_{t=0}^{T}, c, \eta\right)$ and the complexity*

[53] Since the $\beta_{t,1}$ are the differences between Φ_t and Φ_0, we have $\beta_{0,1} = 0$ by definition.

premium in the baseline period, Φ_0.

Proof. Because of the non-linearities of the wage equation (21), the proof of the proposition is quite involved. We first construct a mapping of the data and the structural parameters and use Brouwers Fixed Point Theorem to show that it has at least one solution. We then combine Sard's Theorem and the Inverse Function Theorem to show that the map is locally invertible. The detailed proof is given in the appendix. □

A few points regarding this identification result warrant further discussion. First, as common for non-linear models, without further assumptions on the data we can only proof local, but not global identification. Yet, this is sufficient for standard consistency results of M-estimators, such as the one we use below, which also rely on local invertibility of the map from parameters to data.

Second, it is tempting to interpret proposition 4 as stating that the parameters of the skill aggregator are identified from raw wages under strong functional form assumptions. This however is not correct. The identifying variation of the skill aggregator comes entirely from the *difference* between raw wages $w_t(x)$ and the skill price function $\pi_t(x)$, the latter of which is identified up to levels in the baseline period from the Flat Spot Method. Hence, an essentially model-free selection correction is already implicit in equation (17) because $\pi_t(x)$ enters as data.[54]

Finally, the proposition also implies that identification and estimation of the parameters entering the skill aggregator and those describing the distribution of heterogeneity in unobserved comparative advantage can be treated separately. Indeed, the distribution $\tilde{G}_t(\tilde{a})$ does not enter the reduced-form wage equation (17). There are two reasons for this separation result. First, PAM implies that there is an invertible map between unobserved heterogeneity and observed occupational choices. Second, we treat occupational choices as a continuous problem over occupational complexity x rather than a discrete problem over occupations indexed by $j = 1, .., J$. As a consequence, each occupational wage corresponds to a single level of $\tilde{a}$ and is therefore independent of the skill distribution.

This raises once more the question of how large the bias is from approximating wages in discrete occupations by the solution of a continuous choice problem, which we now address. For notational convenience, we drop time subscripts in the following. We keep the following notation: $w_j = M(\tilde{a}^*(x_j))$ is the continuous approximation of the wage in occupation j, $x(j) \in [0, 1]$ is the task

[54]Notice that whether a "pre-estimated" skill price function comes from the Flat Spot Method or any other semiparametric approach does not matter as long as it is "cleaned" from variation in worker-level efficiency units.

complexity of occupation j, and $\tilde{a}^*(x)$ is the inverse policy function and provides the skill level (in terms of the comparative advantage in solving complex problems) optimally assigned to an occupation with task complexity x. Notice that $M\left(\tilde{a}^*(x_j)\right)$ is the value function of the worker's occupational choice problem and is the upper envelope of the family of wage functions $w\left(x, \tilde{a}\right)$ parameterized by continuous $\tilde{a}$. In the continuous approximation, each worker's wage will lie on this envelope, but in the discrete case, this is true only for a measure zero of individuals for whom skills are exactly equal to one of the values in the list $\{\tilde{a}^*(x_j) : j = 1, ..., J\}$. The remaining workers will not attain this envelope, and this makes deriving bounds on the approximation error non-trivial. Most importantly, skill levels greater than $\tilde{a}^*(x_j)$ will not necessarily choose more complex occupations even in the log-supermodular case. At the same time, with a generalized definition of monotone matching as in Legros and Newman (2002), sorting will still feature the PAM property.

To make this precise, first note that our parameterization implies that equation (6) evaluated at $x = 1$ solves for a finite $\tilde{a}^*(1)$, which we write as $\overline{\tilde{a}}$. Hence, PAM needs to be defined as an interval partition rule: There is a partition of $[0, \overline{\tilde{a}}]$, written $\{\tilde{a}_j : j = 1, ..., J\}$ with $0 < \tilde{a}_1 < ...\tilde{a}_{J-1} < \overline{\tilde{a}}$, such that $x^*(\tilde{a}) = x_0$ if $\tilde{a} < \tilde{a}_1$ and $x^*(\tilde{a}) = x_j$ if $\tilde{a} \in (\tilde{a}_{j-1}, \tilde{a}_j]$ for $J \geq 2$. Also, $\tilde{a}^*(x_j) \in (\tilde{a}_{j-1}, \tilde{a}_j]$ because selecting occupation j maximizes income for this skill level. Our characterization of the continuous choice problem above directly implies that *any equilibrium of our model needs to have the interval partition property*. Indeed, we know that $\tilde{a}^*(x_{j-1}) < \tilde{a}^*(x_j)$ and that log-supermodularity of the objective function implies that higher levels of $\tilde{a}$ will never choose smaller values of x. Hence, there must be a cutoff level $\tilde{a}_{j-1} \in [\tilde{a}^*(x_{j-1}) < \tilde{a}^*(x_j)]$ such that workers with $\tilde{a} < \tilde{a}_{j-1}$ will choose occupation $j-1$ and workers with $\tilde{a} \geq \tilde{a}_{j-1}$ will choose occupation j. The "true" wage for occupation j is therefore given by $E\left[w\left(x_j, \tilde{a}\right) | \tilde{a} \in (\tilde{a}_{j-1}, \tilde{a}_j]\right]$, and the approximation bias at point x_j is $\varepsilon(x_j) \equiv \left|M\left(\tilde{a}^*(x_j)\right) - E\left[w\left(x_j, \tilde{a}\right) | \tilde{a} \in (\tilde{a}_{j-1}, \tilde{a}_j]\right]\right|$. We characterize the properties of this bias in the following Proposition:[55]

Proposition 5. *Let assumptions 1 to 6 be satisfied, and assume that realized equilibria feature wages that are an increasing function of x (for which $c > 2$ is a sufficient condition). Then $\varepsilon(x_j) \to 0$ as $J \to \infty$. A distribution-free and computable bound of $\varepsilon(x_j)$ for fixed J is the difference between the true wages of occupations $j+1$ and $j-1$.*

[55]We consider a limit process in which an increasingly fine-grained occupational classification is used, but in which the value of $x \in [0, 1]$ for occupation j remains the same. Task complexity is a fundamental property of an occupation, and along the entire limit process the probability that two occupations have exactly the same value is zero.

Proof. See the appendix. □

With "distribution-free and computable" we mean that we can calculate a bound on the approximation error for occupation j from the data, without imposing any distributional assumptions on $\tilde{a}$. This bound is simply the difference in hourly wages observed for occupations $j+1$ and $j-1$. Since the complexity-wage relationship in the data is nearly log-linear in all years, the bound in logs is approximately proportional to $2 \cdot \frac{\ln w(1)^{obs} - \ln w(0)^{obs}}{J}$, where $w(1)^{obs}$ and $w(0)^{obs}$ are observed wages for the occupations with the highest and lowest complexity values respectively. Since J is 317 in our sample, this bias is small. The cost of using the continuous approximation on the labor supply side can therefore be expected to be small in our empirical implementation.

Implementation We prove proposition 4 by deriving as many equations as there are structural parameters. In principle, one can construct consistent estimators based on the same equations, but this is inefficient. In the estimation we instead target the entire set of occupational wage data, stacked over all periods. This is operationalized by estimating the parametric reduced-form wage equation (21), which is on the level of occupations and time, by non-linear least squares (an M-estimator), thereby minimizing the empirical distance between the two sides of the equation, evaluated at $(T+1) \times J$ data points.

5.5 The Skill Distribution and the Labor Market Equilibrium Condition

The Labor Market Equilibrium Condition Proposition 4 implies that the structural parameters of $h_t(x, \tilde{a})$ can be identified from wage-complexity profiles without knowledge of the distribution of the comparative advantage in solving complex problems, $\tilde{G}_t(\tilde{a})$. On the other hand, identification of the parameters of $\tilde{G}_t(\tilde{a})$ requires knowledge of $h_t(x, \tilde{a})$, as we now show. Furthermore, since observed wages are used for inferring $h_t(x, \tilde{a})$, identification of $\tilde{G}_t(\tilde{a})$ needs to rely on the observed choice distribution and its evolution over time. This distribution is an equilibrium object and needs to be consistent with the labor market equilibrium condition in section 4.4. PAM simplifies this condition substantially and, coupled with a parametric assumption on $\tilde{G}_t(\tilde{a})$, delivers testable implications.

To see this, consider an occupation that has a complexity index of, say, x_t^q. PAM has two important implications for the supply side of the labor market. First, the policy function $\tilde{a}_t^*(x)$ is bijective, so there is a unique skill level $\tilde{a}^*_t(x^q_t)$ that is optimally assigned to this occupation. Second, the population share of workers who optimally sort into occupations with $x \leq x^q_t$ is $\tilde{G}_t(\tilde{a}^*_t(x^q_t))$. On the demand side of the labor market, let $S_t(x)$ be the cdf of occupational employment optimally chosen by firms in terms of raw labor rather than efficiency units. Choose the x^q_t to be the q-th quantile of $S_t(x)$, and let $\tilde{a}^q_t$ is the $q-$th quantile of $\tilde{G}_t(\tilde{a})$. Then the labor market equilibrium condition with PAM becomes

$$\tilde{a}^*_t(x^q_t) = \tilde{a}^q_t. \tag{22}$$

In equilibrium, the cumulative share of workers choosing occupations with $x \leq x^q_t$ must be equal to the cumulative labor demand $\frac{q}{100}$, and this is exactly the case if equation (22) is satisfied. If for example $\tilde{a}^*_t(x^q_t) < \tilde{a}^q_t$, too many workers go to occupations with complexity above x^q_t as firms would like to assign a share of $\frac{q}{100}$ of workers to occupations with $x \leq (x^q_t)$, but the supply of workers who is willing to work in these occupations is $\tilde{G}_t(\tilde{a}^*_t(x^q_t))$. Since PAM implies that $\tilde{a}^*_t(x)$ is monotonically increasing, we know that $\tilde{G}_t(\tilde{a}^*_t(x^q_t)) < \frac{q}{100}$, and therefore there is an excess demand. As a consequence, the equilibrium complexity premium Φ_t needs to *decrease*, inducing more workers to go to lower-complexity occupations and increasing the demand for higher-complexity occupations.

Parameterization and Identification We parameterize the population distribution of the worker-level comparative advantage in solving complex tasks as follows:

Assumption 7. *Comparative advantage $\tilde{a}$ is Weibull-distributed: $\tilde{G}_t(\tilde{a}) = 1 - e^{-\left(\frac{\tilde{a}}{\rho_t}\right)^{\kappa_t}}$, with $\tilde{a} \geq 0$ and $(\rho_t, \kappa_t) \in \mathbf{R}^2_+$.*

This assumption is motivated by two considerations. First, the Weibull distribution is a parsimonious yet flexible distribution for non-negative random variables. Second, under this assumption the labor market equilibrium condition outlined in section 4.4 becomes strikingly simple, and it imposes testable restrictions on the data. More specifically, the q-th quantile of the 2-parameter Weibull distribution is given by $\tilde{a}^q_t = \rho_t \cdot \left(-\ln\left(1 - \frac{q}{100}\right)\right)^{\frac{1}{\kappa_t}}$, and by defining $Q(q) \equiv -\ln\left(1 - \frac{q}{100}\right)$, the equation (22) turns into the following log-linear relationship:

$$\ln \tilde{a}^*_t(x^q_t) = \ln \rho_t + \left(\frac{1}{\kappa_t}\right) \cdot \ln Q(q). \tag{23}$$

Identification of the Weibull parameters $\{\rho_t, \kappa_t\}_t$ can now be established almost immediately. Take some arbitrary value $q_1 \in (0, 100]$. Let $Q_1 = Q(q_1)$, which is, by definition of $Q(q)$, independent of the structural Weibull parameters. The population cdf $S_t(x)$ over occupational employment is an observable equilibrium object, and $x_t^{q_1}$ can therefore be treated as observed. Proposition (4) implies that we know all parameters entering the policy function $\tilde{a}_t^*(x)$, so that $\tilde{a}_t^*(x_t^{q_1})$ is known as well. For period t, this delivers a point $(\ln Q_1,\ \ln \tilde{a}_t^*(x_t^{q_1}))$ on the linear equation (23). We can now repeat this exercise for many values of q and for all periods covered by the data, and PAM implies that the policy function will be non-degenerate. There will therefore be variation in $\tilde{a}_t^*(x_t^q)$ as the Econometrician varies q, and both the intercept and the slope of the line in equation (23) will be well-defined for all t. These will be exactly $\ln \rho_t$ and $\left(\frac{1}{\kappa_t}\right)$, respectively. This gives the following proposition:

Proposition 6. *Suppose the assumptions of proposition 4 are satisfied, and assume that $\tilde{G}_t(\tilde{a})$ is the Weibull CDF as stated in assumption 7. Then the parameters $\{\rho_t, \kappa_t\}_t$ are identified.*

Implementation We estimate the Weibull parameters exactly as described in the previous paragraph. A particularly attractive feature of this approach is that the labor market equilibrium condition in equation (23) can be evaluated using simple scatter plots. Under the model's assumption, this scatter plot should be a line. We will perform this exercise for each year in our data in the results section.

5.6 Production Function

Identification We now explore identification of the production function parameters in equation (9) from data on the occupational wage- and employment structure, input prices, and value added. The cost-minimization problem in (8) imposes the following restriction on the demand-side variables and parameters:

$$\frac{\partial\left(p_t \cdot y_t\right)}{\partial l_{j,t}} = \overline{w}_{j,t}, \text{for all } j,t \tag{24}$$

$$\frac{\partial\left(p_t \cdot y_t\right)}{\partial k_t^e} = r_t^e, \quad \text{for all } t \tag{25}$$

$$\frac{\partial\left(p_t \cdot y_t\right)}{\partial k_t^s} = r_t^s, \quad \text{for all } t \tag{26}$$

$$p_t \cdot y_t = R_t, \quad \text{for all } t \tag{27}$$

For each t, Euler's Formula implies that the last equation holds whenever the first three sets of equations are satisfied. It is therefore not possible to separately identify p_t and A_t from these equations. Instead we define $\widetilde{A}_t \equiv p_t \cdot A_t$ and treat $\tilde{A}_t$ as parameters. Furthermore, our production function features constant returns to scale, and we therefore can impose, without loss of generality, the normalization $\lambda_{k,t} = 1$ for all j,t.

There are two primary challenges to establishing identification. First, firms do not hire raw inputs, such as observed values of the capital stock and total hours of work, but their quality-adjusted counterparts, and their prices need to be quality-adjusted as well. For capital inputs and their user costs we follow standard procedure in the literature, as described in sections (2.4), (C.2) and (5.3). For labor inputs and their wages we can use the result in proposition 4 to map observed raw quantities observed in the data, say $(s_{j,t}, w_{j,t})$, into efficiency units and their marginal costs. This is possible because $l_{j,t} = s_{j,t} \cdot h_t\left(x_{j,t}, \tilde{a}_t^*(x_{j,t})\right)$ and $w_{j,t} = \overline{w}_{j,t} \cdot h_t\left(x_{j,t}, \tilde{a}_t^*(x_{j,t})\right)$, and $h_t\left(x_{j,t}, \tilde{a}_t^*(x_{j,t}\right)$ can be treated as known.[56]

The second challenge is that without further parametric restrictions, the order condition for point identification is violated. Indeed, the system of equations (27) provides $(J+2) \times (T+1)$ restrictions on the data, but there are $(J \times (T+1) + J + T)$ structural parameters in the list $\{\lambda_{j,t}, \sigma_j, \widetilde{A}_t\}_t$, and with $T << J$ the former is smaller than the latter. We therefore impose the following assumption:

Assumption 8. $\ln\left(\lambda_{j,t}\right) = \widetilde{\lambda}_j + \theta_j \cdot t$.

This is a rather flexible specification as it allows for occupation-specific time trends. This type of parametric assumption is standard and is strictly required for identification as a consequence of a

[56] This is the reason for why the structural parameters of the supply side need to be uncovered from the data first.

result in Diamond *et al.* (1978).

Identification can now be established using the results in lemma 1 in the appendix and proposition 2. It is straightforward to show that the ratio of the first order conditions for $l_{J,t}$ and k_t can be written as

$$\frac{\overline{w}_{J,t}}{r_t} = \lambda_{J,t} \cdot \left(\frac{k_t}{l_{J,t}}\right)^{1-\sigma_J}, \tag{28}$$

or, in logs and substituting in assumption 8, as:

$$\ln\left(\frac{\overline{w}_{J,t}}{r_t}\right) = \widetilde{\lambda}_J + \theta_J \cdot t + (1-\sigma_J) \cdot \ln\left(\frac{k_t}{l_{J,t}}\right). \tag{29}$$

This is a regression of the log difference in user costs between type-J labor and capital equipment on occupation fixed effects, a linear occupation-specific time trend and the log-difference between the two inputs. All parameters of this equation are identified from variation across time. Importantly, $(1-\sigma_J)$ is equal to the ratio of $\Delta^2 \ln\left(\frac{\overline{w}_{J,t}}{r_t}\right)$ and $\Delta^2 \ln\left(\frac{k_t}{l_{J,t}}\right)$, where Δ^2 is the Difference-in-Differences operator. We can use this result as the initialization for an induction that iterates backwards from $j = J$ to $j = 1$. At each step, we work with an equation of the form

$$\ln\left(\frac{\overline{w}_{j,t}}{r_t}\right) = \widetilde{\lambda}_j + \theta_j \cdot t + (1-\sigma_j) \cdot \ln\left(\frac{k_t}{l_{j,t}}\right) + \ln G_j(\omega_{j,t}), \tag{30}$$

where $\omega_{j,t} = \left\langle \left\{\lambda_{m,t}, \left(\frac{k_t^e}{l_{m,t}}\right)\right\}_{m=j+1}^{J}, \{\sigma_l\}_{l=j}^{J} \right\rangle$ (see Proposition 2). This is a standard regression equation that is used for estimating capital-labor substitutability, plus a bias term $\ln\left(G_j(\omega_{j,t})\right)$ that depends on all capital-labor ratios and parameters for occupations indexed by $m \geq j$. Induction provides a characterization of this term, and in the appendix we show that it does not cause a failure of identification.

The induction only uses equations involving quality-adjusted wages and input quantities. Data on value added immediately delivers identification of the α_t from equation (26) :

$$\alpha_t = \frac{r_t^s \cdot k_t^s}{R_t} \tag{31}$$

Indeed, because the technology is Cobb-Douglas in capital structures and the composite labor input, the α_t coincides with its capital-structures share of value added. This is the first example where using data on R_t is crucial for achieving identification.

The arguments above leave unused either the first-order condition for $l_{1,t}$ *in levels* or the constraint $R_t = p_t * y_t$. Since $y_t = \frac{F\left(l_{1,t},\ldots l_{J,t},k_t^e,k_t^s\right)}{A_t}$ and we have established identification of all parameters in $F(.)$, any of these two sets of equations identifies the $\tilde{A}_t$'s.[57]

We summarize these results with the following Proposition:

Proposition 7. *Suppose the assumptions of proposition 4 are satisfied. If $\lambda_{j,t}$ satisfy assumption 8, the first-order conditions (24) to (27) invert to the set of structural parameters $\{\lambda_{j,t}, \widetilde{A}_t, \sigma_{j,t}\}_{j,t}$. Hence, the production function parameters are point identified.*

Proof. See the appendix. □

Implementation We have explored identification of the production function parameters for a rather flexible specification. Without imposing additional restrictions, this can generate overfitting in the estimation. After all, each occupation is associated with a fixed effect $\widetilde{\lambda}_j$, with a time trend θ_j, and capital-labor substitution parameter σ_j. While they are theoretically identified, as stated in proposition 7, in practice we are working with three-digit occupations in which there can be idiosyncratic variation due to small cell sizes from decade to decade. The model is flexible enough that the parameters will capture this noise at least to some extent. This is particularly problematic in our context because the production function is defined recursively, and thus any small-sample bias in a parameter will potentially have large effects on others. In the estimation we therefore impose the additional restrictions that

$$\sigma_j = \tilde{\sigma_0} + \tilde{\sigma_1} \cdot x \tag{32}$$

and

$$\ln(\lambda_{j,t}) = \widetilde{\lambda}_j + \widetilde{\lambda}_t + \theta_1 \cdot x + \theta_2 \cdot x \cdot t. \tag{33}$$

We then estimate the production function using Non-linear Least Squares on the system of equations (30), stacked over all years and all occupations. Since $\ln G_j(\omega_{j,t})$ does not have a closed

[57] As noted above, Euler's Formula implies that only one of these two remaining equations have identification power.

form solution, the right-hand side needs to be evaluated numerically.[58] Although the baseline complexity premium Φ_0 is theoretically identified, in practice we use as our estimate the coefficient on occupational complexity in a cross-sectional wage regression for the years 1976 to 1980 that includes fixed effects for age, education, race, sex and 3-digit industry codes.

6 Results

6.1 Overview

In this section, we present the results from the structural estimation of our parameterized model. We emphasize two main findings. First, capital-skill complementarity in occupational complexity space and its interaction with technological progress in the aggregate production function can generate quantitatively the complexity premium's rise and slowdown over the last five decades. Second, the slow-down in the growth of the aggregate complexity premium, in conjunction with a steady rise of the complexity premium in raw wages, suggests that technological change in the skill aggregator is quantitatively important as well. We refer to this channel as supply-side technological change.

It is important to keep in mind that the estimation must be carried out in the order presented in section 5. As discussed there, it is the availability of external estimates of the skill-price function that allows us to estimate separately the parameters of the supply-side and the demand-side. We therefore start with discussing our estimates of this function, for which we rely on the Flat-Spot Method. This method has a number of well-known drawbacks, which we address with extensive validation- and robustness exercises. Next, we use the discrepancy between aggregate skill prices and occupation-level raw wages to estimate the parameters of the skill aggregator, and we use data on the occupational employment choices in combination with the restrictions from the labor market equilibrium condition to recover the skill distribution. Given these estimates, we have all the ingredients to map raw occupational wages and hours of work to efficiency units, and this allows us to pin down the aggregate production function. Even though we implement our estimation

[58] Since equation (9) is highly non-linear in parameters, naive estimation with this many parameters is computationally problematic. Instead we use an EM-like algorithm that exploits the fact that conditional on one set of parameters, the rest of the parameters can be estimated by OLS. Thus, the joint estimation of the parameters can be performed by breaking up each iteration into two nested steps that have better computational properties. In particular, for given values of $\{\sigma_j\}_{j=1}^{J}$ we can estimate the implied set of $\{\lambda_{jt}\}_{j=1}^{J}$ by iteratively performing the regression in equation (30). We therefore use a nested algorithm that loops over σ_j, recovering the resulting λ_{jt} in each step, to best fit the first order conditions (24) - (25). This algorithm turns out to be numerically very robust.

exactly in this order, we find it more natural for expositional clarity to present production function estimates before the supply-side estimates.

6.2 The Price of Complexity

Sample We estimate the regression in equation (16) on data from the March files of the CPS, referred to as Annual Social and Economic Supplement (ASEC), and construct variables to ensure maximal comparability with our decennial data. This includes imputation of top-coded earnings observations. Our wage variable relies on the series on annual labor earnings and annual hours of work that are consistently available since 1976. Following Bowlus and Robinson (2012) we assume that workers reach their flat-spot at the age of 55, and to increase sample size we widen the age window to be between 53 and 57 years.[59]

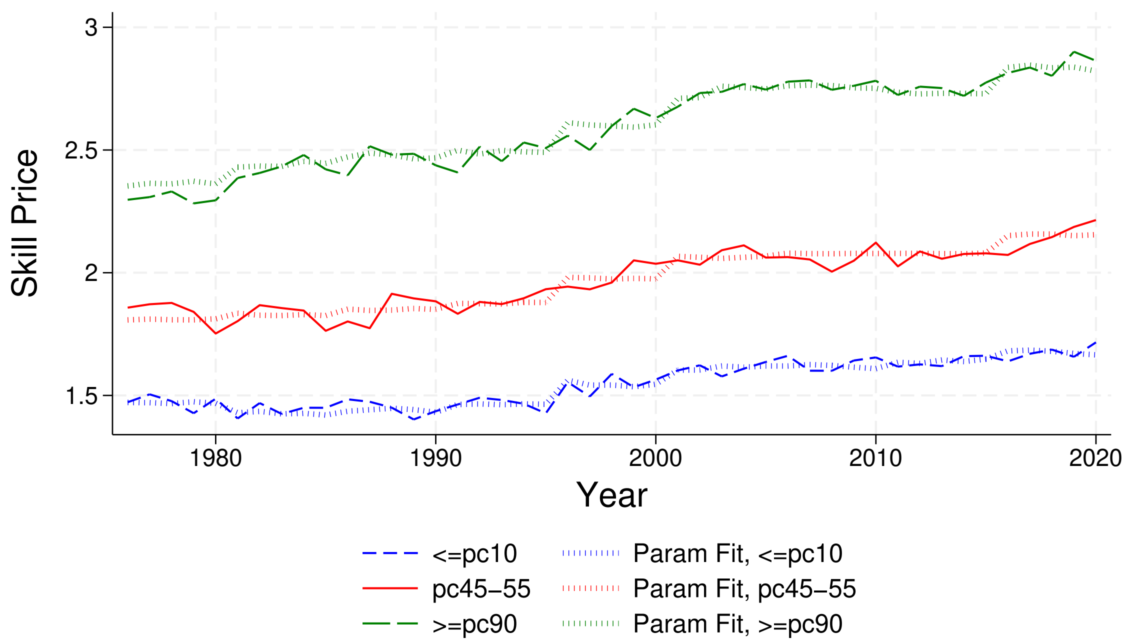
Validating the Parametric Flat Spot Regression Proposition 1 states that assumption 3 of non-increasing *returns* to complexity, defined as $\Phi_t(x) = \frac{\pi_t'(x)}{\pi_t(x)}$, is important for PAM to arise in equilibrium. The parametric Flat Spot Regression in equation (16) imposes the stronger assumption 4 that $\Phi_t(x)$ is constant in x, which we stated in section 5.2 to be strongly supported by empirical evidence. In this section we present this evidence.

As a first step we implement the Flat Spot Method non-parametrically by estimating equation (14), but we are forced to aggregate the data due to many small cells on the occupation-time level. For comparability with Bowlus and Robinson (2012), who estimate skill prices for three different education groups, we aggregate up to three approximately equally sized groups of 3-digit occupations and plot the time-series of the $\pi_\tau(x)$ starting in 1976, using average group-specific wages in 1976 as baseline. We then overlay these time-series with the prediction for the same occupation groups from the parametric specification in (16). We construct these three groups as follows. First, we order 3-digit occupations by their occupational complexity and compute their cumulative employment shares over the entire sample period, which we denoted by $S(x)$ in section 5.5. We then keep only 3-digit occupations who fall into one of three categories, all of which employ approximately 10 percent of all workers: low complexity occupations with $S(x) \leq 0.1$, medium-complexity occupations with

[59] Bowlus and Robinson (2012) provide strong evidence in favor of this assumption. We deviate slightly from their implementation by not allowing the flat-spot to vary between education groups because years of education have no counterpart in our structural model. On the other hand, using the same flat spot for all education groups allows us to use a narrower age window as in their paper.

$S(x) \in [0.45; 0.55]$, and high-complexity occupations with $S(x) \geq .9$. Results from this exercise are shown in Figure 3.

Figure 3: Estimated Equilibrium Skill Prices for Three Occupation Groups.



We draw three main conclusion from this exercise. First, the parameterized model in equation (16) provides an exceptionally good fit as it captures the trends for all three groups near perfectly. In light of the small number of parameters, this result is quite remarkable. Second, as expected the growth rate is larger for more complex occupations so that the dispersion of skill prices between the three occupation groups increases over the sample period. In particular, the difference between skill prices for the most- and least complex group of occupations was 81 log points in 1980. In 2020 this same difference was 115 log points. Finally, this dispersion in skill prices between the most- and least complex occupations largely takes place prior to 2000. Indeed, between 2000 and 2020 the premium paid to the most complex occupations relative to the least edges up just 8 log points. As we will discuss below, the dispersion in raw wages by occupation complexity displays a quite different pattern.

One concern with this test of our parametric specification is that it drops all occupations not contained in one of the three groups and that it is performed on a highly aggregated sample. As a second test we produce scatter plots of average raw wages and their predicted values from the parametric regression on the 3-digit occupational level, separately for each of the 5 decades covered by our data. These plots, overlaid with a 45-degree line, are shown in figure A.7 in the appendix. They confirm our findings from figure 3 as the scatter plots are tightly concentrated around the 45-degree line in each decade. Hence, our parametric specification that imposes assumption 4 captures the entire cross-sectional distribution of occupational wages and its change over time exceptionally well. There is virtually no evidence for non-linearities in the complexity premium, and the assumption is therefore corroborated by the empirical evidence, as claimed earlier.

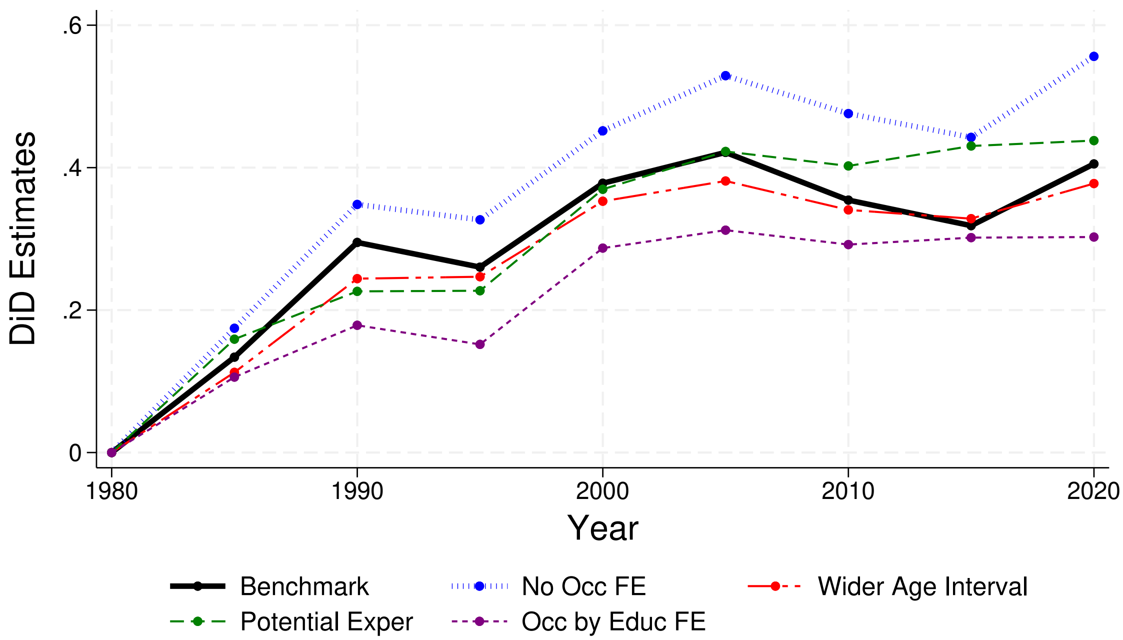
Results Estimates of the parameters in equation (16) together with their standard errors are shown in table A.7.[60] We also display the predicted growth of the skill price gap between occupations at the extremes of the complexity distribution (with $x = 1$ and $x = 0$, respectively), which is the sample analogue of the difference-in-difference $\beta_{2020,1} - \beta_{1980,1}$ (*ie.* the complexity premium). The benchmark estimates, which are computed as described above, are reported in the first column. The interactions of complexity with the time fixed effects are statistically significant at the 1 percent level and the R^2 of the model is 27 percent. Given that we estimate the model on approximately 282,000 observations on individual wages, we consider this to be a very good fit.

Under our benchmark specification, the estimated overall growth of the complexity premium is 40.5 log points. Importantly, our estimates also reveal a substantial slowdown in the skill premium over the course of our sample, consistent with the evidence in figure 3. To see this more clearly, figure 4 plots the evolution of the complexity premium implied by our estimates (black line). While the estimated skill premium grew by around 30 log points in the 1980s, it only grew by about 8 log points between 1990 and 2000, and overall growth in the skill premium was near-zero after 2000.

To explore robustness of these results, we estimate a variety of alternative specifications. The estimated complexity premia from these specifications are plotted in figure 4, and the estimation results are listed in table A.7. We first re-estimate the model without occupation fixed effects, but with a linear control for occupational complexity, so that cross-sectional variation across occupations contributes to the identification of the parametric skill price function. As shown in the blue dotted

[60] Since we estimate the model on worker-level data, and since there may be unobserved serial correlation of wages within occupation groups, we cluster standard errors on the occupation level.

Figure 4: Evolution of Complexity Premium relative to 1980: Robustness.



line in figure 4 (and column (2) in table A.7) this delivers a much larger value of 55.6 log points for growth in the predicted skill price gap between high- and low-complexity occupations. Failure to restrict identifying variation to within-occupation wage changes therefore seems to lead to severe upward biases in estimates of occupational wage gaps, which is why researchers rely on the Flat Spot Method in the first place. That said, the estimates still point to a slowdown in complexity premium. It is also noteworthy that at 20 percent the R^2 remains relatively high for a repeated-cross-sectional regression on individual wage data, especially bearing in mind that the model in column (2) has substantially fewer parameters. It is not the 317 occupation fixed effects that explain most of the model fit in our baseline.

Next, we bring back occupation fixed effects, but widen the sample to workers between 50 and 59 years of age, the same age range as used for highly educated workers in Bowlus and Robinson (2012). This raises the sample size substantially, but has a near-negligible effect on benchmark parameter estimates (red dash-dotted line in figure 4, column (3) in A.7). One concern with the Flat-Spot method is that a change of the age-earnings profiles across cohorts will contaminate estimates of aggregate skill prices. For example, a steepening of these profiles over time will lead to overestimates of the growth of the skill price function. Our finding that our estimates are robust to widening the Flat Spot suggests that this is not an empirically relevant issue. Key to this argument is that we focus on a sufficiently old group of workers, for whom the net returns to skill accumulation are close to zero no matter the time period.

Bowlus and Robinson (2012) define the flat-spot by years of potential experience, not by age, and we next experiment with this approach. First, we replace restrictions on age by the restriction that workers have between 28 and 32 years of potential experience. We find similar growth in the occupational skill price gap in this case to our baseline estimates (green dashed line in figure 4, column (4) in A.7). A potential issue with this approach is that for each occupation we are comparing year-to-year wage growth between lowly educated workers who are relatively young with highly educated workers who are relatively old. To restrict identifying variation to comparisons within education groups we therefore replace occupation fixed effects with occupation-by-education fixed effects in the last column of the table. This brings the predicted growth in the skill price gap down to about 30 log points (purple short-dashed line in figure 4, column (5) in A.7), though again the estimates suggest a slowdown in the complexity premium, particularly after 2000.

Overall, the results suggest that a parametric flat-spot method captures the relationship between wage growth and occupational complexity over the last five decades quite well. We find no evidence that our estimates are sensitive to the definition of the Flat Spot in terms of age or potential experience, as long as we focus on a reasonably old group of workers.

6.3 Structural Estimates: The Demand Side

Suppose that in addition to the equilibrium skill price function $\pi_t(x)$ we also know the skill aggregator $h_t\left(x, \tilde{a}\right)$. Then we can map raw hours of work for each occupation into efficiency units and, since $\pi_t(x)$ is the equilibrium price in efficiency units, we can implement the production function estimation described in section 5.6. We now present the results from this step, and we will discuss the supply-side estimates afterwards.

There are two main channels through which our production function is consistent with the observed large increase in the complexity premium and a moderate increase in realized quantities of occupation-specific labor inputs (in efficiency units). The first is capital-skill complementarity in combination with the dramatic decline in the user cost of equipment capital.[61] If the labor-capital complementarity is an increasing function of the complexity of occupational labor, and if the supply of labor to these occupations is sufficiently inelastic, then this mechanism can potentially generate the rise in the complexity premium we have estimated from the data. One interpretation of this mechanism is automation, whereby increasingly cheap robots are more likely to replace labor in low-complexity occupations. The second channel is an increase of the factor productivity of high complexity occupations relative to both, low complexity occupations and equipment capital.

We now present the point estimates that capture either of these channels. We will quantify their relative importance using counterfactual simulations in section 7.

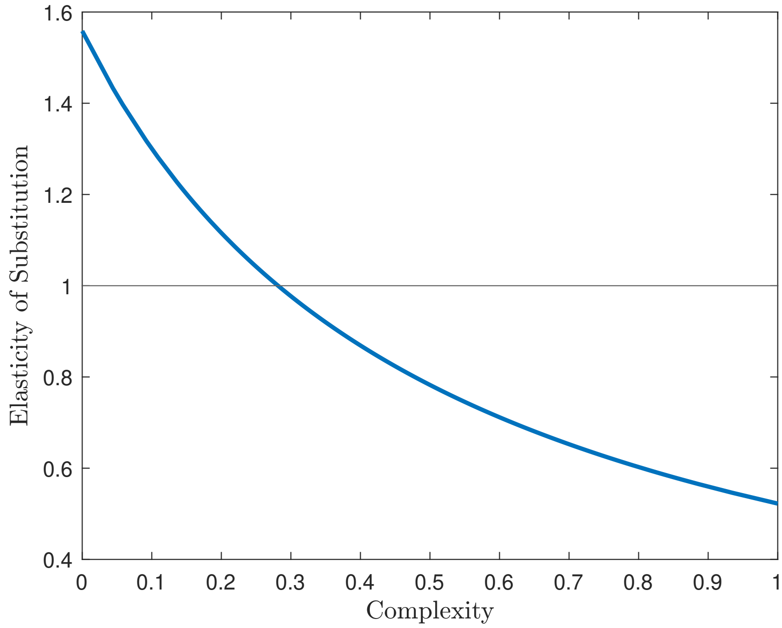
6.3.1 Capital Skill Complementarity and Automation

Estimates of the elasticity of substitution as a function of occupational complexity are shown in Figure 5. They are given by $\rho(x) = \frac{1}{1-\sigma(x)}$, where we use the estimates of equation (32) for $\sigma(x)$. We find strong evidence for capital-skill complementarity. The estimated relationship decreases from a value of about 1.6 for $x = 0$ to about 0.55 for $x = 0$. The mean elasticity of substitution is 1.58 for the 25 least complex occupations and 0.77 for the 25 most complex occupations. These estimates are comparable with those found elsewhere in the literature that relies on different skill measures. For example, Krusell *et al.* (2000) find, in a model with just two types of labor, an elasticity of substitution of 1.67 and 0.67 between capital equipment and unskilled- and skilled labor respectively. Similarly, Caunedo *et al.* (2023) estimate occupation-level elasticities of substitution between capital and labor ranging from 0.65 to 2.18.

For further comparison with the existing literature, it is helpful to aggregate our occupation-specific elasticities. Since this aggregation needs to use employment as weights, we perform this exercise for each year in our data separately. This yields average elasticities of substitution between labor and capital of 0.86 in 1980 and of 0.82 in 2020 (not shown in the table). These estimates are directly comparable to those reported in the existing literature. For example, Oberfield and Raval (2021) find below-unity aggregate elasticities using micro-data from the US manufacturing sector, while the occupation-level estimates in Caunedo *et al.* (2023) also imply below-unit elasticities. Our occupation-level estimates therefore aggregate to a capital-labor elasticity that is broadly consistent

[61] In the context of the college premium, this is the channel emphasized by Krusell *et al.* (2000).

Figure 5: Estimated Capital-Labor Elasticity of Substitution.



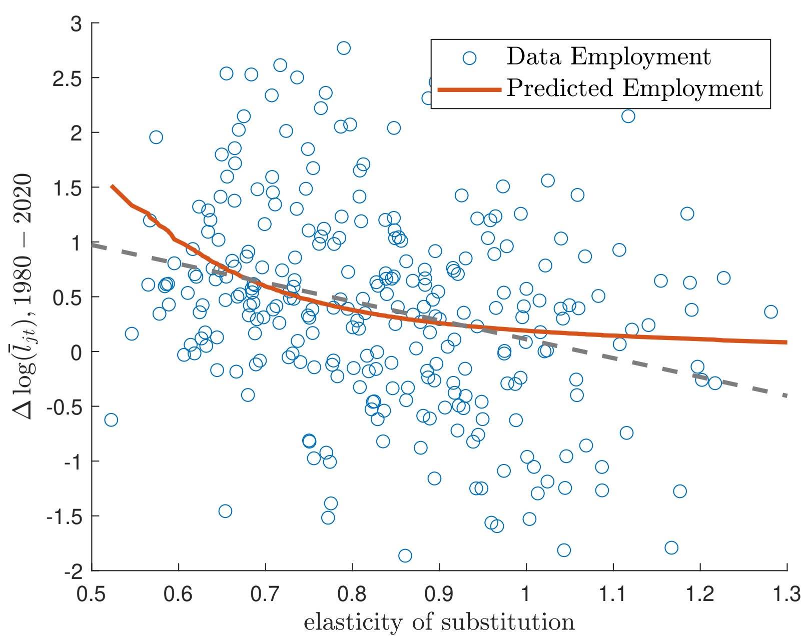
with the literature.[62]

The canonical model of SBTC cannot generate a decline in the absolute demand for some groups of labor in response to automation, as argued for example in Autor and Acemoglu (2011) and Acemoglu and Restrepo (2018). What does our model say about automation? Advances in computer technology and robotics enter our model through the decrease in the relative price of capital equipment shown in figure A.1. If, on average, labor in low complexity occupations is highly substitutable with these technologies, while it is complementary with them in high complexity occupations, then firms in our model have the incentive to automate the former. This leads to a decrease in the demand for low skill labor and an increase of the aggregate price of complexity. To quantify the importance of this mechanism, we plot in figure 6, for each of the 317 occupations, our estimates of the elasticity of substitution between labor and equipment capital against the model's predicted *changes* in occupational employment in efficiency units from 1980-2020. We also show a raw scatter plot of the underlying data together with the line of best fit. The model indeed predicts that occupations that are more substitutable with capital equipment were more likely to

[62]The question of whether the aggregate capital-labor substitution elasticity is above or below one remains contested however. For example, Hubmer (2023) documents above-unity estimates in more aggregated data, and Berlingieri *et al.* (2024) obtain qualitatively similar results in French micro data. León-Ledesma *et al.* (2010) also argue that declining labor shares in the aggregate requires labor and capital inputs to be gross substitutes.

suffer sizable declines in employment. This effect is particularly pronounced for occupations at the bottom of the complexity distribution.[63] On the other hand, the unexplained variation in the scatter plot remains quite large, so that automation is only one of several mechanisms that drive trends in the occupational employment distribution.[64] As shown by Autor *et al.* (2024), even within 3-digit occupations there is substantial changes of tasks over time which seem to partially shield its employment from the effect of automation. This effect will be captured by technological change of our skill aggregator, as we will discuss in some detail below. We conclude that our model does feature automation. It therefore provides a framework that builds a bridge between the literature on SBTC on the one hand and the literature on automation and routinization on the other hand.

Figure 6: Estimated Capital-Labor Elasticity of Substitution.



6.3.2 Factor Productivities

The second set of parameters of the aggregated production function consists of the λ_{jt} in equation (9), which governs the evolution of the factor productivities of type-j labor relative to equipment capital. They too can generate a rise in the aggregate price of complexity if their growth rate over time is a sufficiently increasing function in task complexity. In contrast to capital-skill complemen-

[63]We find a similarly negative relationship when using raw hourly employment instead of efficiency units.

[64]The regression line of best fit, shown in gray, has a slope of -1.83, which is significant on the 1 percent level.

tarity, this mechanism is independent of the price of equipment capital, which enables us to separate the two mechanisms.

We show the estimated relationship between the growth-rate of λ_{jt} and occupational complexity in Figure 7. In anticipation of our results from the supply side, which reveal distinct periods that characterize technological progress between 1980 and 2020, we produce this plot separately for each decade in our sample. The blue line indicates that during the 1980s factor productivity growth favored complex occupations, with λ_{jt} growing over 90 log points more in the most complex occupations relative to the least. This pattern reversed thereafter, with factor productivity growth substantially *slower* in more complex occupations, but not to the extent that the relative productivity gains of labor in complex occupations during the 1980s were eliminated all together.

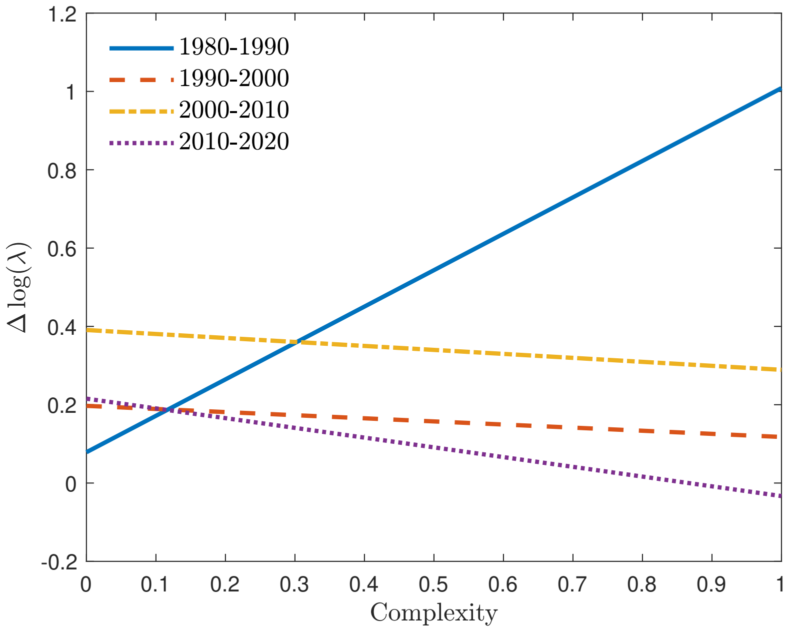
Statistically, this reversal is inferred from two patterns in the data. First, as discussed above, the growth of equilibrium skill prices slowed substantially after 1990 and was essentially zero after 2000. Second, the price of equipment capital continued to drop over the same time period, albeit at a slower rate in later years, so that capital-skill complementarity continued to exert upward pressure on skill prices. As a result, skill-prices would have risen at counterfactually high rates had growth in λ_{jt} evolved as in the 1980s. Economically, the results in the figure may seem puzzling. However, they are consistent with recent research documenting a slowdown in the demand for skilled labor, variously defined, in recent decades. Most prominently, Beaudry *et al.* (2016) document evidence for a reversal in the demand for cognitive tasks around the year 2000. They posit that the IT revolution took the form of a general purpose technology that reached maturity around the turn of the millennium. It is plausible to conjecture that the patterns of our λ_{jt}-estimates partially capture this reversal.

6.4 Structural Estimates: The Supply Side

6.4.1 The Skill Aggregator and Occupational Wages

The estimation of the production function is based on data expressed in efficiency units. This requires estimates of the skill aggregation technology $h_t(x, \tilde{a})$, which we now present together with estimates of the skill distribution $\tilde{G}_t(\tilde{a})$. We also discuss the evidence for and the importance of supply-side technological change.

Figure 7: Growth of Factor Productivities



Model Fit and Parameter Estimates We estimate the parameters of $h_t(x, \tilde{a})$ using Non-Linear Least Squares on the reduced-form wage equation (21), as described in section 5.4. The theoretical prediction on the right hand side of the equation involves twelve free parameters – the two time constant parameters $\{\eta, c\}$ together with the time-varying $\gamma_{0,t}$ and $\delta_{0,t}$ – which we estimate from 1,553 wage observations on the occupation-year level.[65] In the following it is important to keep in mind that the identifying variation of this regression comes from the difference between occupation-level raw average wages and the aggregate skill prices.

Figure A.8 documents the model fit by showing, for each of the five decades in our data, a scatter plot of occupational complexity and raw wages (in levels), overlaid with the model prediction. The model captures very well the salient features of the wage-complexity relationship, in particular the positive complexity-wage gradient (in all years), its convexity when using wage-levels (rather than log-wages), and its evolution over time. This is an important result because it is direct evidence that our model can match quantitatively the evolution of the entire occupational wage distribution since 1980.

The point estimates of the structural parameters from which figure A.8 is computed are listed in Table 2. Controlling for a rich set of demographic- and skill variables delivers an estimate of the

[65] Given any value of $\gamma_{0,t}$ and $\delta_{0,t}$, the $\delta_{1,t}$ are pinned down by the full-support condition and are therefore not free parameters.

1980 baseline complexity premium of 0.55, substantially lower than the difference in raw average log wages between the most and least complex occupations in the same year (approximately 1, see figure A.3). The estimated value of η hits its upper bound of 1, suggesting that the two skills a and d are perfectly substitutable within occupations. This result may be surprising at first, but is consistent with the lack of substantial non-linearities in the complexity-wage relationship *within* each year. That is, imperfect substitutability has quantitative implications for which we do not find evidence in the data.[66]

Table 2: Estimated Parameters of the Skill Aggregator

Parameter	Estimate
Φ_0	0.55
η	1.00
c	2.17
$\gamma_{0,t}$	$\{0.35, 0.34, 0.37, 0.88, 0.73\}$
$\delta_{0,t}$	$\{0.30, 0.43, 0.46, 0.44, 0.44\}$
$\delta_{1,t}$	$\{1.06, 1.12, 1.13, 1.08, 1.05\}$

The estimated value of c is slightly above 2, implying a near-quadratic "cost function" $\delta_t(x)$. This result can be explained by two theoretical properties of our model. First, as stated in Proposition (3), the condition that $c \geq 2$ is sufficient for complexity-wage profiles to be positively sloped throughout its domain. It is therefore not surprising that we find an estimate above this value. Second, the degree of complementarity between $\tilde{a}$ and x is given by $\left[\gamma'(x) \cdot \delta(x) - \gamma(x) \cdot \delta'(x)\right]$ and is bounded below by $\delta_1 + (c-1) \cdot (x + \delta_0)^c$ when using our parameterization.[67] For relatively complex occupations[68] this term is strictly increasing in c for $c > 1$. Our estimate of this parameter suggests

[66] One potential concern with this result is that we infer η exclusively from the distribution over occupational choices and the corresponding wages rather than panel data with direct measures of skills, such as the National Longitudinal Study of Youth (NLSY). However, existing evidence using such data suggests that high substitutability between high-order skills on the one hand and skills useful for solving "everyday problems", such as diligence and raw physical prowess, is not an unreasonable assumption. For example, Cunha *et al.* (2006) state that "Highly motivated, but not very bright, people may be just as effective as bright but unmotivated people" and cite evidence from the GED program in support of this hypothesis. A recent correlational study by Palczyńska (2020) using Polish nationally representative data documents similar results. On the other hand, recent research on the returns to different skills in the US labor market, such as Weinberger (2014) and Deming (2023), finds increasing complementarity between cognitive and social skills. These studies focus on high-order skills however, which both enter our $a-$skill, and do not inform the complementarity with our $d-$skill.

[67] See the proof of Proposition 3.

[68] Namely, those occupations for which $x + \delta_0 > 1$.

that values of c that are substantially larger than 2 will generate too strong complementarities and thereby too steep a wage-complexity relationship.

Arguably the most important finding in the table is that the skill aggregator is subject to substantial technological change over time, with a trend break in the year 2000. In particular, the factor loading $\gamma_{0,t}$ on the comparative advantage term $\tilde{a}$ remained relatively stable between 1980 and 2000, while the cost function shifted upwards as $\delta_{0,t}$ increased substantially over this period of time. This pattern reverses after 2000, when $\gamma_{0,t}$ grew almost two-fold, while $\delta_{0,t}$ and $\delta_{1,t}$ remained essentially constant.[69] Our supply side estimates therefore uncover a similar (though slightly later) structural break as our demand-side estimates discussed above, suggesting that the skill aggregation technology plays an important role in explaining the evolution of the occupational wage- and employment structure since 1980.

We now turn to a more detailed investigation of this channel and how it is inferred from the data. In particular, we argue and demonstrate that the finding of supply-side technological changes working through the skill aggregator necessarily follows from the theoretical implications of PAM on the one hand, and from our empirical finding that the raw complexity premium and its selection-corrected analogue, Φ_t, evolved quite differently between 1980 and 2020 on the other hand.

The Importance of Supply-Side Technological Change Identification of the skill aggregation technology relies on the difference between raw occupational wages and the corresponding skill prices. To understand why its parameters change over time, we begin by plotting in figure 8 the difference in both average wages and skill prices between the most- and least complex occupations, which can be interpreted as raw wage and complexity (*ie.* selection-corrected) premia. As shown in the figure, the two premia evolved quite differently. Between 1980 and 2000, Φ_t grew by 38 log points, while the wage premium rose by just 21 log points. This pattern reversed starting in 2000, when the aggregate complexity premium rose by only 3 log points over the rest of the sample period, while the wage premium increased by 23 log points. As was the case with the estimated factor productivities on the demand side, the evolution of the occupational wage structure thus features two distinct episodes: in this case the decades between 1980 and 2000, when the aggregate

[69] For two of the years in our sample we find $\delta_{0,t} > \gamma_{0,t}$, which violates one of the sufficient conditions for PAM stated in proposition 3. However, they are not necessary, and the PAM property can still be satisfied for $\delta_{0,t} > \gamma_{0,t}$, provided the condition (A.4) is met. This is indeed the case for our estimated parameterization.

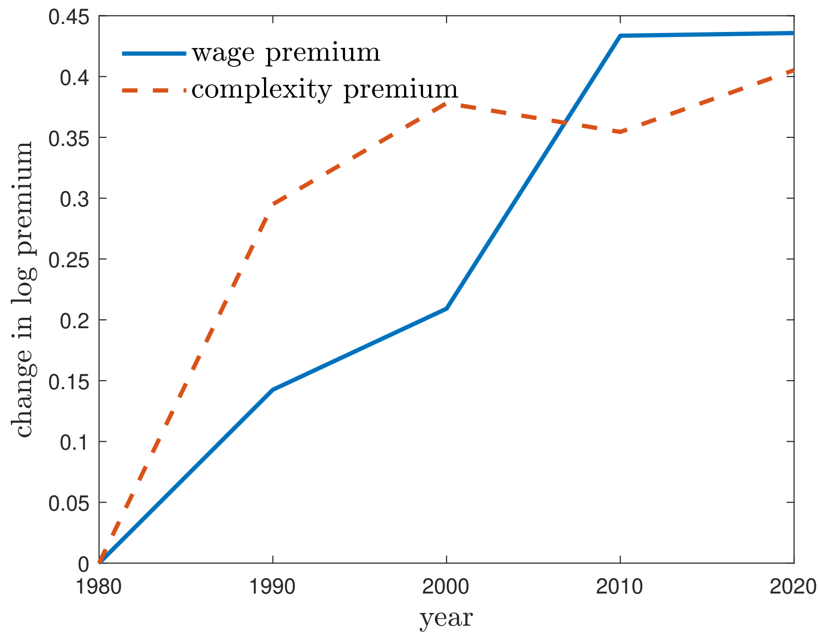
price of skill outpaced raw wages, and the period after 2000 when the converse was true.

Why does our model interpret these empirical patterns as the result of supply-side technological change? Notice that the estimation equation (21) implicitly imposes the restrictions implied by PAM because it conditions on *observed occupational choices being optimal.* To understand the implications of this conditioning, first consider the 1980s. During this decade the selection corrected complexity premium Φ_t experienced its largest increase, from about 55 log points in 1980 to approx. 85 log points in 1990. As a consequence, some workers who previously found it optimal to select into lower-complexity occupations now endogenously move up the complexity ladder. Due to PAM, these workers have a lower index of comparative advantage for complex work than workers who chose these occupations before the increase in the complexity premium. An increase in Φ_t, then, leads to a deterioration in an occupation's skill composition (as measured by the comparative advantage of its workers for complex work), pushing down the growth of raw occupational wages. Not correcting for occupational selection therefore leads to a downward bias in the growth of the complexity premium.

As shown in figure 8, this model prediction is *qualitatively* consistent with the data for the entire pre-2000 period. However, all things being equal the growth of Φ_t was so large that *quantitatively* the sorting mechanism described above, and therefore the difference between the growth of the raw- and selection-corrected complexity premia, should have been even larger. Indeed, as we will show in section 6.4.2, there was very little change in occupation selection across these years despite the substantial increase in Φ_t. Through the lens of our model, these results can only be reconciled by concurrent changes in $\gamma_t(x)$ or $\delta_t(x)$. Indeed, we estimate a sizeable increase in $\delta_{1,t}$ between 1980 and 1990, which has two potential interpretations. First, $\delta_t(x)$ can capture direct changes in the monetary rewards to less complex work. Since our model fits occupational wages over the entire complexity distribution in all years, these rewards in levels must have increased over the 1980s. Second, since $\delta_t(x)$ can be interpreted as a cost to selecting into more complex work, it will capture any frictions that temper labor reallocation out of lower-complexity occupations. One possible explanation is de-unionization, which may have impacted middle-complexity occupations in particular, thereby dis-incentivizing upward mobility into these occupations. We do not pursue this hypothesis any further, but merely emphasize that an increase in $\delta_t(x)$ captures any factors that partially mitigate the effect of dramatically increasing aggregate skill prices.

Starting in the year 2000, the model prediction that positive growth in Φ_t generates a selection

Figure 8: Change in the Wage and Complexity Premia Since 1980



Note: The figure plots the change, since 1980, in the difference between the log wage (blue) or skill price (red) of the most- and least complex occupations.

effect that causes raw occupational wage premia to grow less is not supported by the data anymore. In fact, as shown in figure 8 the opposite is the case. Since the gap between the skill price function and average occupational wages is entirely driven by the supply side, our model can explain these patterns only with an increase of the factor loadings on the comparative advantage in solving complex problems, $\gamma_t(x)$. This increases the efficiency units of labor in more complex occupations, driving up their wages at a faster rate than skill prices. As shown in table 2, our parameter estimates of $\gamma_{0,t}$ have indeed grown dramatically since 2000.

Is Supply-Side Technological Change Plausible? The importance of supply-side technological change for explaining the evolution of the occupational employment and wage structure over the last five decades is one of the two central quantitative findings of our paper. Our evidence comes from the combination of (a) pre-estimates of the rise in the selection-corrected aggregate complexity premium, which we obtain via the Flat Spot Method, with (b) the model assumption that skill-based sorting satisfies PAM. One may be concerned that the Flat Spot Method leads to pre-estimates that are biased. However, we have experimented with a wide range of implementation choices, some of which we discussed in section 6.2, and none of them deliver substantially different estimates than those used in our calibration. Moreover, the changes in Φ_t we feed into our structural estimation lie at the upper end of the estimates we have found in our robustness exercises, alleviating concerns that the estimated slowdown in the complexity premium reflects a downward bias. Finally, as shown in figure 3, our findings do not result from our parametric restrictions, as the non-parametric differences in occupational wages between high- and low complexity occupations have evolved virtually identically to those predicted by the parametric version of the skill-price function.

Another concern may be that the supply-side mechanism fundamentally depends on PAM. However, one of the primary objectives of this paper is to demonstrate that the evolution of the occupational wage- and employment structures over the last five decades can be rationalized quantitatively by a neoclassical model with skill-based monotonic sorting, and this sorting is equivalent to PAM in a framework with two-dimensional skills. That said, it is true that we could let other parameters in the skill aggregation technology vary over time to match wage-complexity profiles. For example, treating the conditional expectations of the second skill d given $\{x, \tilde{a}\}$ as free parameters that change over time would allow us to fit the evolution of the occupational wage structure over time as well. We do not pursue this avenue because, as discussed above, this would add degrees of freedom to our model, require direct skill measures for identification, and yield a model that no longer features pure comparative advantage.

Possibly the strongest support for our finding is that it is consistent with recent results in the existing literature that come from very different approaches to studying occupational wages. For example, several papers, such as Beaudry and Green (2005) and Beaudry *et al.* (2016), document substantial changes in worker composition within occupations in terms of observed characteristics, suggesting that there has indeed been a change in the assignment of worker skills to occupations. Our model can explain these empirical patterns if these worker characteristics are correlated with their comparative advantages in solving complex tasks. Similarly, Autor *et al.* (2024) provides strong evidence that within the 3-digit occupational codes we use in this paper there has been substantial labor-augmenting technological change that leads to fundamental changes in how labor is utilized. The γ_t and δ_t in our quantitative framework capture such within-occupation technological changes in a parsimonious way. According to this interpretation of our concept of supply-side technological change, task augmentation has led to a more effective utilization of worker abilities to solve complex problems.

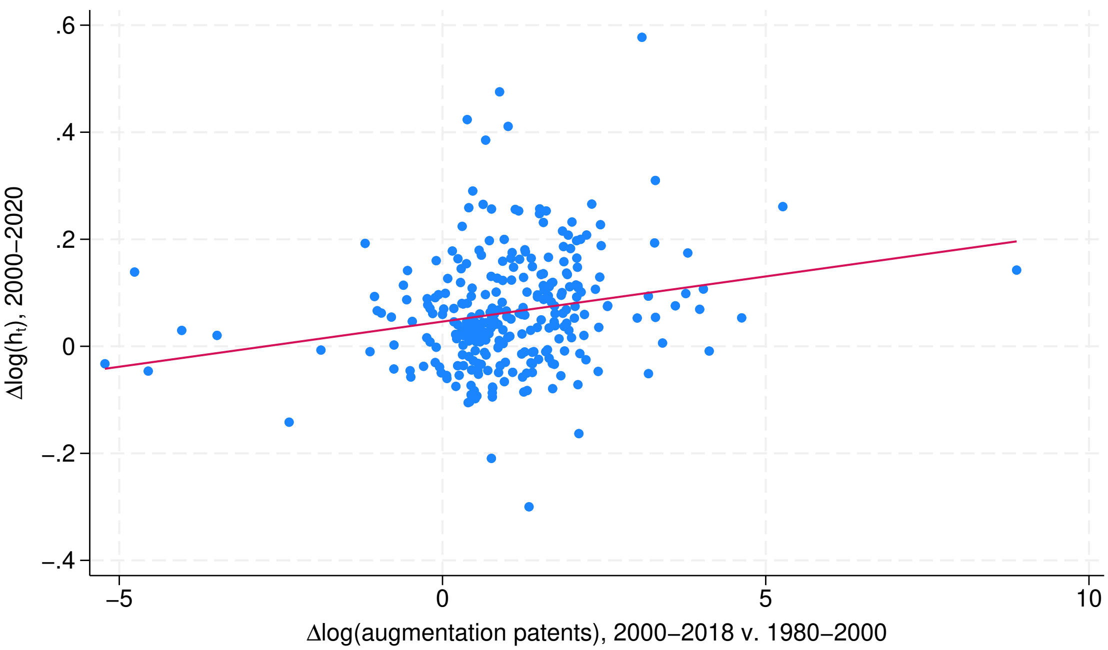
To provide suggestive evidence for this interpretation, we directly utilize the task augmentation measures on the three-digit occupational level from Autor *et al.* (2024).[70] We first calculate for each occupation j the difference between raw occupational log-wages, $\ln(w_t(x_j))$, and the corresponding aggregate skill price in logs $\ln(\pi_t(x_j))$. Under the assumption of our model, this difference is equal to the skill aggregation function evaluated at optimal choices $\ln(h_t(x, \tilde{a}))$. We then calculate the occupation-level growth rate of this object between 2000 and 2020, which we write has $\Delta_{00}^{20}(h_t)$. In absence of supply-side technological change, this difference is predicted to be unambigously negatively correlated with the aggregate price of complexity due to occupational choices having the PAM property, and this was indeed the case for the pre-2000 period. However, for the post-2000 period the correlation is positive instead, and the model can produce this result only with a change of the skill aggregator. This is the variation pinning down the evolution of the γ_t and δ_t. If an increase of the within-occupation efficiency in using worker skills for solving complex problems is partially due to task augmentation, as we hypothesize here, then we expect a measure of the latter to be positively correlated with $\Delta_{00}^{20}(h_t)$. Figure 9 shows a scatter plot of the two variables, where we use the growth in augmentation patents from Autor *et al.* (2024) as a measure of augmentation exposure. As predicted, we find a clear and pronounced positive relationship between the two. The line of best fit has a slope of 0.017, which is strongly significant. While these results do not have a causal interpretation, we still view them as strong external validation of supply-side technological change. First, our finding that the differential growth rates of the raw wage- and selection-corrected complexity premia are informative about the sources of technological change (if occupational choices have the PAM property) is, to the best of our knowledge, novel and *a priori* not obvious. Second, we have at no point of data construction and estimation utilized the augmentation measures from Autor *et al.* (2024), hence figure 9 is truly external validation.

6.4.2 The Skill distribution and Occupational Employment

Given our estimates of the skill aggregator, we recover the parameters of the skill distribution $\tilde{G}_t(\tilde{a})$ from the regression in equation (23). This equation is the structural labor market equilibrium condition and relates the predicted cdf of occupational employment over task complexity from

[70] We have downloaded the replication package of Autor *et al.* (2024) from the *Quarterly Journal of Economics* website, which includes the augmentation measures.

Figure 9: Supply-side Technological Change and Task Automation.

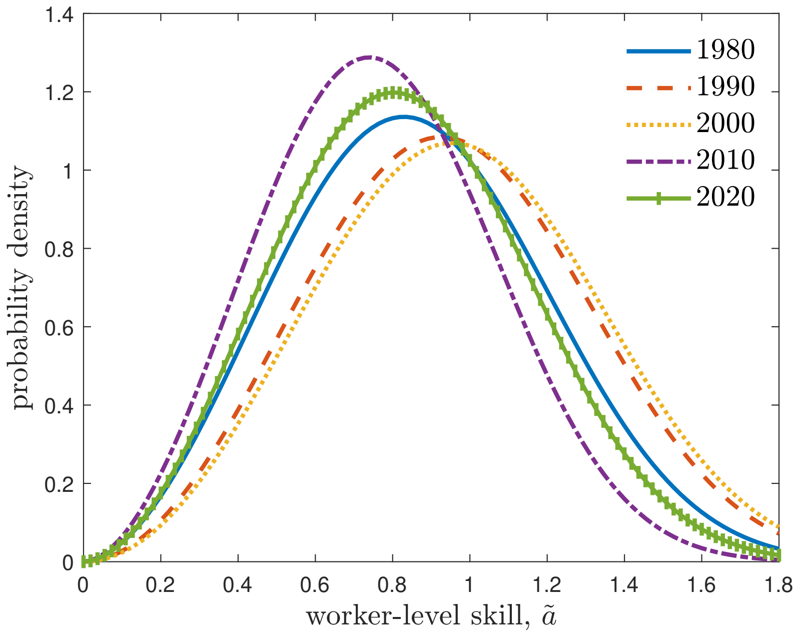


optimal worker choices to the empirical analogue. The model fit is shown in A.10, and justifies the restriction that the skill distribution follows a 2-parameter Weibull distribution. In particular, the assumption that the quantiles of worker ability, $\tilde{a}_t^q$, are log-linear in $\ln(1-q)$ is remarkably close to the predictions of the model's estimated inverse policy functions (which are themselves fully independent of our assumptions about $\tilde{G}_t(\tilde{a})$).

Figure 10 plots the estimated distribution of worker-level skill, $\tilde{G}_t(\tilde{a})$. Concurrent with the increasing returns to skill documented above, we find a rightward shift in the distribution of worker skills from 1980 to 2000. However, this is followed by a relatively sharp contraction in the supply of high-skill workers between 2000 and 2010. This shift likely reflects structural determinants of labor force participation that are external to our model. For example, the labor force participation rate increased between 1960 and 2000, and the substantial increase in female labor force participation in particular has been credited with increasing the relative supply of skilled workers during this period (Mulligan and Rubenstein, 2008). However, both female- and male labor force participation rates began to decline in the 2000s, which may well be reflected in the observed shifts in $\tilde{G}_t(\tilde{a})$. Our 2010 estimates for the skill distribution will also be influenced by any impacts of the Great

Recession on labor force participation, and these may be contributing to the estimated shift in $\tilde{G}_t(\tilde{a})$ between 2000 and 2010. Taken together, while there has been some decadal movement of the skill distribution, in the long run, that is between 1980 and 2020, its rightward shift has been relatively small. This is a reflection of our descriptive finding in section 3 that compared to the growth of the complexity premium, changes in high-complexity employment have been relatively muted.

Figure 10: Estimated PDF of Worker-Level Skill.



7 Counterfactuals

In this section, we use counterfactual analysis to quantify the importance of both demand- and supply-side technological changes and of the changes in the skill distribution for the rise of the complexity-wage premium. We will focus on counterfactual wage *premia* since this requires fixed point iteration on the skill-price function only, but not on the entire General Equilibrium correspondence, as shown in appendix C.3. This substantially facilitates conducting counterfactual experiments and is a consequence of the assumed exogeneity of the equilibrium prices of capital and the separability of consumption- and occupational choices.[71]

We present results in terms of Difference-in-Differences for raw occupational wages $w(x)$, the skill

[71] If one wants to compute counterfactuals on the wage distribution *in levels*, one still needs to solve for the entire equilibrium fixed point.

price function $\pi(x)$, and efficiency units $h(x, \tilde{a}^*(x))$. The first difference is between these functions evaluated at the 90th- and the 10th percentiles of the occupational complexity distribution and can be interpreted respectively as the raw complexity wage premium, the aggregate skill premium, and the selection effect due to sorting. The second difference is taken relative to the baseline period 1980. Results for long-run changes (between 1980 and 2020) are shown in columns (a)-(c) of table 3. The other columns of the table show the corresponding results separately for the time periods 1980-2000 and 2000-2020.[72] For the sake of comparison, the results for the baseline calibration are shown in line 1 of the table. The counterfactual experiments are carried out and presented in the same order as in section 6.

7.1 Counterfactuals: The Demand Side

7.1.1 Capital-Skill Complementarity

We start with quantifying the importance of capital-skill complementarity in driving the increase of the raw- and aggregate complexity premia over the last 5 decades. To this end, we set σ equal to the economy-wide average of our estimates for σ_j, so that the capital-labor substitution elasticity no longer depends on occupational complexity. As indicated by the results in the second row of table 3, the effect of this counterfactual parameter restriction is to almost halve the growth in both complexity premia. However, our estimates differ substantially between the periods before and after 2000. Most importantly, the counterfactual effect on the skill-price function is almost entirely driven by the pre-2000 period. The main reason is that under both the baseline and the counterfactual, growth in the capital stock slows markedly after 2000, a consequence of flattening user costs. The small effect on wage premia of eliminating capital-skill complementarity is another manifestation of our finding from the structural estimation that demand-side driven technological change has reduced explanatory power for the evolution of the occupational wage structure after 2000.

We also compute the corresponding employment effects, measured by the change in the employment share of complex occupations (defined here as the share of workers in occupations with

[72] Figures A.11 to A.13 show the impact of the counterfactual parameterization on the wage premium, complexity premium, relative efficiency units, and complex employment broken down by decade instead. Since the courser breakdown in table 3 captures the most important changes over time, we do not discuss these figures in the main text.

Table 3: Wage, Skill Price, and Labor Supply Changes under Counterfactual Parameterizations of the Model

	Difference-in-Difference x_{90}-x_{10}								
	1980 – 2020			1980 – 2000			2000 – 2020		
	$w(x)$	$\pi(x)$	$h(x,\tilde{a})$	$w(x)$	$\pi(x)$	$h(x,\tilde{a})$	$w(x)$	$\pi(x)$	$h(x,\tilde{a})$
Exercise	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
1. Baseline	0.292	0.255	0.037	0.142	0.260	-0.118	0.150	-0.004	0.154
Counterfactual Demand: Capital-Skill Complementarity									
2. No capital-skill complementarity	0.148	0.137	0.010	0.033	0.134	-0.101	0.114	0.003	0.111
Counterfactual Demand: User Cost									
3. Constant r_t^e	0.213	0.150	0.063	0.111	0.214	-0.102	0.101	-0.064	0.165
4. No flattening in r_t^e post-2000	0.318	0.289	0.029	0.142	0.260	-0.118	0.176	0.030	0.147
Counterfactual Demand: Factor Productivities									
5. No relative slowdown in complex λ_{jt}	0.379	0.366	0.013	0.164	0.292	-0.128	0.215	0.074	0.141
Counterfactual Supply: Skill Distribution									
6. Constant $G_t(\tilde{a})$	0.230	0.174	0.057	0.273	0.445	-0.173	-0.042	-0.272	0.229
Counterfactual Supply: Supply-side technology									
7. No post-2000 γ_t jump	0.416	0.590	-0.173	0.142	0.260	-0.118	0.275	0.330	-0.056

Notes: Columns (a)-(c) and are difference-in-differences, showing the change from 1980 to 2020 in the difference in (a) log hourly wages, (b) skill prices, and (c) occupation efficiency units between the occupations for which $x = 0.9$ and $x = 0.1$.

$x > 0.5$), which are shown in Table 4. As can be seen in line 2, the lower path of the complexity premium under the no capital-skill complementarity scenario reduces selection into complex occupations, with the complex employment share falling 2 percentage points (ppt) over the 1980-2020 period, compared with 3.5 ppt growth under the baseline.

7.1.2 The User Cost of Capital

Next, we quantify the impact of the decline in the user cost of equipment capital on wage- and employment patterns. As shown in figure A.2, r_t^e declined throughout the sample period, but at a decreasing rate in later decades. Line 3 of table 3 reports the results for a counterfactual where we

Table 4: Employment Changes under Counterfactual Parameterizations of the Model

	Change in Complex Employment Share		
	1980-2020	1980-2000	2000-2020
1. Baseline	0.035	0.026	0.009
Counterfactual Demand: Capital-Skill Complementarity			
2. No capital-skill complementarity	-0.020	-0.009	-0.011
Counterfactual Demand: User Cost			
3. Constant r_t^e	-0.059	-0.012	-0.047
4. No flattening in r_t^e post-2000	0.063	0.026	0.038
Counterfactual Demand: Factor Productivities			
5. No relative slowdown in complex λ_{jt}	0.124	0.052	0.072
Counterfactual Supply: Skill Distribution			
6. Constant $G_t(\tilde{a})$	0.001	0.039	-0.038
Counterfactual Supply: Supply-side technology			
7. No post-2000 γ_t jump	0.092	0.026	0.066

Notes: The table chose the change in the complex employment share, defined here as the share of individuals employed in occupations for which $x > 0.5$.

hold r_t^e at its 1980 value throughout the 1980-2020 period. This is the scenario where the marginal cost of expanding the stock of capital equipment has not declined and stands in for the case where automation has, at the margin, not become cheaper over time. Due to capital-skill complementarity in the production technology, the effect is to reduce growth in both the wage and complexity premia throughout the sample period. Since the aggregate complexity premium grows less, the PAM property of occupational choices implies that workers sort into higher complexity occupations at a lower rate. As this drives up the average occupation skill, $\tilde{a}$, in more complex occupations, their relative efficiency units increase relative to the baseline, and this effect is substantial. As can be seen in line 3 of table 4, this re-sorting of workers across occupations comes with a decline in the share of workers employed in relatively complex occupations by 6 ppt, compared to a 3.5 ppt increase under the baseline.

Row 4 of table 3 explores this mechanism further and quantifies the effect of the flattening of the user cost of capital in the later years of our sample. To this end we assume that r_t^e drops at the same rate in the 2000s and 2010s as it did in the 1990s.[73] The results from this scenario suggest that user costs are a minor contributor to the slowdown in labor demand, with both the wage and complexity premia growing only around 3 log points more than in the baseline after 2000 (line 4, table 3) and the complex employment share increasing by around 3 ppt after 2000 (line 4, table 4). Again we find that demand-side sources of technological change, here captured through the declining marginal cost of capital equipment, is not a major driver of the occupational wage- and employment structure in the post-2000 period.

7.1.3 Factor Productivities

In section 6 we documented a reversal in the pattern of factor productivity growth across occupations. Whereas growth in λ_{jt} was higher for more complex occupations in the 1980s, the reverse was true thereafter. To quantify the importance of this channel, we consider a counterfactual in which λ_{jt} continues to grow at its 1980 pace throughout the sample period. The results are shown in line 5 of table 3 and show that the relative slowdown in the factor productivity of complex labor has been considerably more important than flattening user costs in explaining recent occupational wage changes. Absent the reversal in the growth of relative factor productivities, the complexity premium would have grown by around 37 log points, compared with 26 log points under the baseline. The post-2000 period accounts for a substantial share of this counterfactual effect. We also find quantitatively important worker selection effects, with higher complexity premia drawing more workers into complex occupations. This pushes up the complex employment share by about 9 ppt relative to the baseline (line 5, table 4). Despite these selection effects, however, the impact of the factor productivity slowdown on the raw wage premium is comparable to its effect on the aggregate complexity premium (line 5, table 3).

While exogenous changes in complex labor demand have had a sizeable impact on complexity and wage premia over the period of our sample, as shown above, they cannot account for the slowdown in skill prices on their own. Indeed, no matter the demand-side counterfactuals we have conducted so far, the growth rate of the complexity premium is always substantially smaller in the later period of the sample (columns (h) vs. (e) of table 3). In order to account for this slowdown

[73] We need to impose a lower bound on r_t^e as it otherwise would go negative in the 2010s.

we must instead consider the impact of the labor supply changes discussed in section 6, to which we now turn.

7.2 Counterfactuals: The Supply Side

The results in section 6 show two sources of labor supply shifts that are exogenous to our model: (i) changes in distribution of worker-level skills and (ii) supply-side technological change, most notably after 2000, when technological change lead to a more effective utilization of workers' comparative advantages in performing complex tasks. To quantify the importance of these two mechanisms, we compute two counterfactuals. First, we compute model outcomes under the assumption that the distribution over worker skills, $\tilde{G}_t(\tilde{a})$, stays at its 1980 values throughout the sample. Second, we compute the model without supply-side technological change after 2000.[74]. The results for the wage and complexity premia are shown in lines 6 and 7, respectively, of table 3, and the results for employment are shown in the same lines of table 4.[75]

We first consider the impact of changes in the distribution of worker skills. Recall that $\tilde{G}_t(\tilde{a})$ shifts rightward between 1980 and 2000. Under the assumption of an unchanging distribution, the supply of high-skill workers will be relatively lower over these two decades. As can be seen in column (e) of table 3, this supply effects leads to a substantially larger increase of the complexity premium. After 2000 the effects are reversed as our baseline estimates for $G_t(\tilde{a})$ show it shifting to the left of the 1980 distribution in these years. In the counterfactual scenario, labor supply in complex occupations will therefore be higher, and this pushes down growth in the complexity premium. Indeed, column (h) of table 3 shows that, absent the decrease of skilled labor supply, the complexity premium would have contracted dramatically after 2000. Moreover, the elasticity of skill prices with respect to complex labor supply is sufficiently large for the counterfactual wage premium to have contracted after 2000 as well. Finally, line 6 of table 4 shows the effect of the counterfactual skill distribution on selection. Faster growth in the complexity premium increases the complex employment share relative to baseline period pior to 2000, while the contraction in the counterfactual complexity premium thereafter causes the complex employment share to decrease by almost 4 ppt over the same period. This re-sorting of workers is due to PAM and dampens the

[74]To be precise we hold $\gamma'_{0,t}$ constant after 2000, so that $\gamma'_{0,2010} = \gamma'_{0,2020} = \gamma_{0,2000}$

[75]For more detail, figure A.13 shows the impacts on the wage premium, complexity premium, relative efficiency units, and the complex employment share by decade

equilibrium effect of our counterfactual parameter restrictions.

What, then, accounts for the slowdown in the complexity premium after 2000? The last column of line 7 of table 3 shows that eliminating supply-side technological change by keeping the $\gamma_t(x)$ at its 2000 estimate has a first order effect on the efficiency units of labor supply in complex occupations. The lower values of these parameters essentially act as a negative supply shock in terms of efficiency units as they lower the skill aggregator, holding everything else constant. This leads to a jump of both the aggregate complexity premium and the raw wage premium after 2000. Furthermore, due to PAM, the higher complexity premium attracts more workers to complex occupations, leading to an increase of the complex employment share (line 7, table 4). For the entire 1980-2020 period, table 3 suggests that the estimated increase in $\gamma_t(x)$ after 2000 caused growth in the complexity premium to decrease by more than half relative to the counterfactual, while growth in the raw wage premium was reduced by about a quarter. If one interprets positive growth of $\gamma_t(x)$ as capturing task augmentation, within-occupational technological change of this kind had an outsized effect on the occupational wage- and employment distribution in the post-2000 period.

8 Conclusion

Understanding the forces behind cross-sectional labor market outcomes remains a central challenge in economics. In this paper we demonstrate that a single-dimensional measure of occupational complexity, based on problem-solving requirements, provides remarkable explanatory power for the evolution of the U.S. occupational wage structure over the past four decades. Our work yields a striking empirical regularity, namely, that occupational complexity exhibits a strong, monotonic relationship with wage growth since 1980. Moreover, once occupations are ordered by complexity, rather than automation potential, educational requirements, or some base period wage, the data reveal a pattern similar to the "race between education and technology" familiar from the SBTC literature–rising returns to complex skills accompanied by modest compositional shifts toward more complex work.

The quantitative model we develop to explain these patterns makes a number of conceptual contributions. First, our supply-side framework with positive assortative matching delivers a natural concept of a skill premium (in this case a complexity premium) while generating selection effects crucial for identification. Second, we introduce supply-side technological change through the skill aggregation technology, capturing in a parsimonious way mechanisms such as task augmentation, emphasized by Autor *et al.* (2024). Third, by treating occupational skill content and capital-labor substitutability as distinct dimensions, we create a framework capable of reconciling insights from both the SBTC and RRTC literatures. In particular, our production technology can generate returns to complex skill through changes in relative factor productivity as well as capital-skill complementarity in the spirit of Krusell *et al.* (2000). At the same time, it allows for occupation-specific automation effects whereby labor in less complex occupations is found to be substitutable with equipment capital.

Our identification strategy also represents a methodological contribution in its own right. In particular, by adapting the flat-spot method to estimate selection-corrected skill prices outside the model, we achieve point identification of all structural parameters without solving for the full general equilibrium.

Our empirical findings suggest that technological change can be characterized by two distinct periods since 1980. Prior to 2000, demand-side forces dominated: capital-skill complementarity interacted with declining equipment costs to boost the complexity premium substantially, while relatively higher factor productivity growth in complex occupations further amplified this trend. The post-2000 period tells a different story. Aggregate complexity premia essentially stagnated while raw occupational wage gaps continued expanding. Our structural estimates suggest that supply-side technological change has been the dominant force behind this pattern. Complex occupations became dramatically more efficient at utilizing worker skills, driving a plateau in their relative skill prices. This finding aligns with mounting evidence that tasks within occupations have evolved substantially, with new technologies augmenting rather than merely replacing labor.

We carry out a counterfactual analysis to quantify these mechanisms. Capital-skill complementarity accounts for roughly half the observed increase in the wage gap between the most- and least complex occupations. The relative slowdown in factor productivity growth for complex occupations after 2000 dampened inequality growth, while changes in the skill distribution and endogenous worker reallocation provided counterbalancing forces. Most strikingly, absent the post-2000 acceleration in supply-side technological change, the wage premium paid to complexity would have grown substantially faster than observed.

Our findings have important implications. In particular, our framework provides a natural foundation for analyzing the labor market impacts of artificial intelligence. Unlike the automation of routine tasks, AI has the potential to boost returns to different types of cognitive problem-solving while at the same time replacing labor in certain tasks. Our distinction between occupational skill content and capital-labor substitutability offers a more nuanced lens for evaluating AI exposure: occupations requiring complex problem-solving may experience wage gains if AI augments these capabilities, even if AI can perform similar tasks in isolation. The post-2000 experience with supply-side technological change offers a potential preview of AI's impact.

This work also suggest new avenues for life-cycle research on human capital investment and occupation selection. For example, our framework of sorting on the basis of comparative advantage in solving complex skills may offer explanations for persistent puzzles including mid-career occupational downgrading and rising within-occupation wage dispersion over the life cycle. More fundamentally, our identification of substantial shifts in occupation-specific technological change implies that cohorts entering the labor market during different technological epochs face different returns to skill investments, possible ground for further research.

References

Acemoglu, D. (1998). Why do new technologies complement skills? Directed technical change and wage inequality. *Quarterly Journal of Economics*, **113** (4), 1055–1089.

— (2024). The Simple Macroeconomics of AI, NBER Working Paper No. 32487.

— and Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, **108** (6), 1488–1542.

— and — (2022). Tasks, Automation, and the Rise in U.S. Wage Inequality. *Econometrica*, **90** (5), 1973–2016.

Althoff, L. and Reichardt, H. (2026). Task-Specific Technical Change and Comparative Advantage, Working Paper.

Athey, S. (2002). Monotone Comparative Statics under Uncertainty. *The Quarterly Journal of Economics*, **1** (117), 187â223.

Autor, D., Chin, C., Salomons, A. and Seegmiller, B. (2024). New frontiers: The origins and content of new work, 1940-2018. *Quarterly Journal of Economics*, **139** (3), 1399–1465.

Autor, D. H. (2014). Skills, Education, and the Rise of Earnings Inequality Among the Other 99 Percent. *Science*, **344** (6186), 843–851.

— and Acemoglu, D. (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, vol. 4, Elsevier, pp. 1043–1171.

— and Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, **103** (5), 1553–1597.

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

Bajari, P., Benkard, C. L. and Levin, J. (2007). Estimating dynamic models of imperfect competition. *Econometrica*, **75** (5), 1331–1370.

Beaudry, P. and Green, D. A. (2005). Changes in U.S. Wages, 1976-2022: Ongoing Skill Bias or Major Technological Change? *Journal of Labor Economics*, **23** (3), 609–648.

—, — and Sand, B. M. (2016). The Great Reversal in the Demand for Skill and Cognitive Tasks. *Journal of Labor Economics*, **34** (1), S199–S247.

Berlingieri, G., Boeri, F., Lashkari, D. and Vogel, J. (2024). Capital-Skill Complementarity in Firms and in the Aggregate Economy, UCLA Working Paper.

Biasi, B., Fu, C. and Stromme, J. (2025). Equilibrium in the Market for Public School Teachers: District Wage Strategies and Teacher Comparative Advantage, University of Wisconsin-Madison Working Paper.

Bowlus, A. J. and Robinson, C. (2012). Human Capital Prices, Productivity, and Growth. *American Economic Review*, **102** (7), 3483–3515.

Burstein, A., Morales, E. and Vogel, J. (2019). Changes in Between-Group Inequality: Computers, Occupations, and International Trade. *American Economic Journal: Macroeconomics*, **11** (2), 348–400.

Caines, C., Hoffmann, F. and Kambourov, G. (2017). Complex-task biased technological change and the labor market. *Review of Economic Dynamics*, **25**, 298–319.

Card, D. and Lemieux, T. (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *Quarterly Journal of Economics*, **116** (2), 705–746.

Caucutt, E. M. and Lochner, L. (2020). Early and Late Human Capital Investments, Borrowing Constraints, and the Family. *Journal of Political Economy*, **128** (3), 1065–1147.

Caunedo, J., Jaume, D. and Keller, E. (2023). Occupational Exposure to Capital-Embodied Technical Change. *American Economic Review*, **113** (6), 1642–85.

Costinot, A. (2009). An Elementary Theory of Comparative Advantage. *Econometrica*, **77** (4), 1165–1192.

Costrell, R. M. and Loury, G. C. (2004). Distribution of Ability and Earnings in a Hierarchical Job Assignment Model. *Journal of Political Economy*, **112** (6), 1322–1363.

Cunha, F., Heckman, J. J., Lochner, L. and Masterov, D. (2006). Interpreting the Evidence on Life Cycle Skill Formation. *Handbook of Economics of Education*, vol. 1, Elsevier, pp. 697–812.

—, — and Schennach, S. M. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, **78** (3), 883–931.

Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market. *Quarterly Journal of Economics*, **132** (4), 1593–1640.

— (2023). Multidimensional Human Capital and the Wage Structure. forthcoming in the Handbook of the Economics of Education, Elsevier.

Diamond, P., McFadden, D. and Rodriguez, M. (1978). Chapter IV.2 - Measurement of the Elasticity of Factor Substitution and Bias of Technical Change. In M. Fuss and D. McFadden (eds.), *Applications of the Theory of Production*, *Contributions to Economic Analysis*, vol. 2, Elsevier, pp. 125–147.

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000. *American Economic Review*, **106** (3), 479â524.

Eloundou, T., Manning, S., Mishkin, P. and Rock, D. (2023). GPTs are GPTs: An Early Look at the Labor Market Imapct Potential of Large Language Models, arXiv Working Paper.

Ericsson, K. A. (2003). "the acquisition of expert performance as problem solving: Construction and modification of mediating mechanisms through deliberate practice". In J. E. Davidson and R. J. Sternberg (eds.), *The psychology of problem solving*, Cambridge: Cambridge University Press, pp. 31–83.

Fan, T. (2025). The Labor Market Incidence of New Technologies, Working Paper.

Freund, L. B. and Mann, L. F. (2025). Job Transformation, Specialization, and the Labor Market Effects of AI, Working Paper.

Gathmann, C. and Schonberg, U. (2010). How General Is Human Capital? A Task-Based Approach. *Journal of Labor Economics*, **28** (1), 1–49.

GOLDIN, C. and KATZ, L. F. (2008). *"The Race Between Education and Technology"*. Cambridge, MA: Harvard University Press.

GOOS, M., MANNING, A. and SALOMONS, A. (2014). Explaining Job Polarization: Routine-Biased Technological Change and Offshoring. *American Economic Review*, **104** (8), 2509–2526.

GREENWOOD, J., HERCOWITZ, Z. and KRUSELL, P. (1997). Long-Run Implications of Investment-Specific Technological Change. *American Economic Review*, **87** (3), 342–362.

HAMPOLE, M., PAPANIKOLAOU, D., SCHMIDT, L. D. W. and SEEGMILLER, B. (2025). Artificial Intelligence and the Labor Market, Working Paper.

HASSAN, T., KALYANI, A. and RESTREPO, P. (2026). The Skill Premium in Times of Rapid Technological Change*, Working Paper.

HECKMAN, J. J., LOCHNER, L. and TABER, C. (1998). Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents. *Review of Economic Dynamics*, **1** (1), 1–58.

HOFFMANN, F., LEE, D. S. and LEMIEUX, T. (2020). Growing Income Inequality in the United States and Other Advanced Economies. *Journal of Economic Perspectives*, **34** (4), 52–78.

HUBMER, J. (2023). The Race Between Preferences and Technology. *Econometrica*, **91** (1), 227–261.

JAIMOVICH, N., SAPORTA-EKSTEN, I., SIU, H. and YEDID-LEVI, Y. (2021). Measuring and Interpreting Trends in Economic Inequality. *Journal of Monetary Economics*, **122**, 1–16.

— and SIU, H. E. (2020). Job Polarization and Jobless Recoveries. *The Review of Economics and Statistics*, **102** (1), 129–147.

KARABARBOUNIS, L. and NEIMAN, B. (2014). The Global Decline of the Labor Share*. *Quarterly Journal of Economics*, **129** (1), 61–103.

KATZ, L. F. and MURPHY, K. M. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *Quarterly Journal of Economics*, **107** (1), 35–78.

KLEINMAN, B., LIU, E. and REDDING, S. J. (2023). Dynamic spatial general equilibrium. *Econometrica*, **91** (2), 385–424.

Kopczuk, W., Saez, E. and Song, J. (2010). Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937. *Quarterly Journal of Economics*, **125** (1), 91–128.

Krusell, P., Ohanian, L. E., Rios-Rull, J.-V. and Violante, G. L. (2000). Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis. *Econometrica*, **68** (5), 1029–1053.

Lagakos, D., Moll, B., Porzio, T., Qian, N. and Schoellman, T. (2018). Life Cycle Wage Growth across Countries. *Journal of Political Economy*, **126** (2), 797–849.

Legros, P. and Newman, A. F. (2002). Monotone Matching in Perfect and Imperfect Worlds. *Review of Economic Studies*, **69** (4), 925–942.

León-Ledesma, M. A., McAdam, P. and Willman, A. (2010). Identifying the Elasticity of Substitution with Biased Technical Change. *American Economic Review*, **100** (4), 1330–57.

Lindenlaub, E. (2017). Sorting Multidimensional Types: Theory and Application. *Review of Economic Studies*, **84** (2), 718–789.

Milgrom, P. and Shannon, C. (1994). Monotone Comparative Statics. *Econometrica*, **62** (1), 157–180.

Moll, B. (2014). Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation? *American Economic Review*, **104** (10), 3186–3221.

Mulligan, C. B. and Rubenstein, Y. (2008). Selection, Investment, and Women's Relative Wages Over Time. *Quarterly Journal of Economics*, **123** (8), 1061–1110.

Oberfield, E. and Raval, D. (2021). Micro Data and Macro Technology. *Econometrica*, **89** (2), 703–732.

Palczyńska, M. (2020). Wage premia for skills: the complementarity of cognitive and non-cognitive skills. *International Journal of Manpower*, **42** (4), 556–580.

Pinker, S. (1997). *How the Mind Works*. New York, NY: W. W. Norton & Company.

Shimer, R. and Smith, L. (2000). Assortative Matching and Search. *Econometrica*, **68** (2), 343–369.

Weinberger, C. J. (2014). The Increasing Complementarity between Cognitive and Social Skills. *The Review of Economics and Statistics*, **96** (5), 849–861.

Wenke, D. and Frensch, P. A. (2003). Is success or failure at solving complex problems related to intellectual ability. In J. E. Davidson and R. J. Sternberg (eds.), *The Psychology of Problem Solving*, Cambridge: Cambridge University Press, pp. 87–126.

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

A Additional Tables and Figures

Figure A.1: Relative Prices of Equipment and Structures.

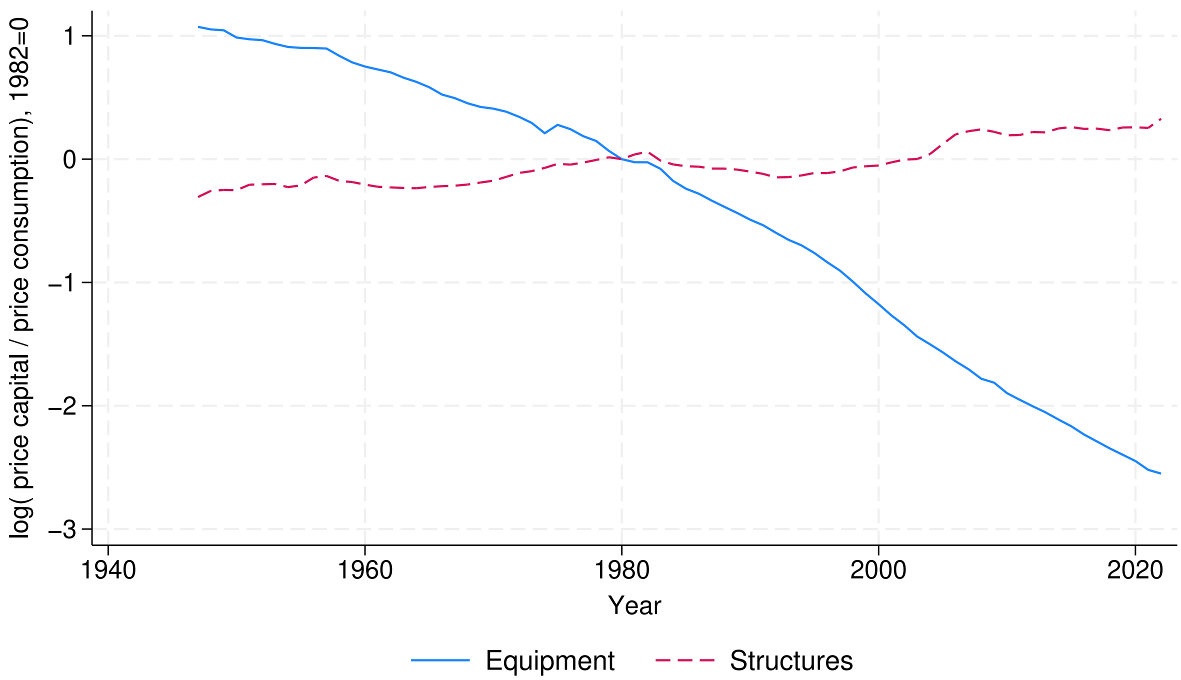


Figure A.2: Change in the User Cost of Equipment Capital Since 1980

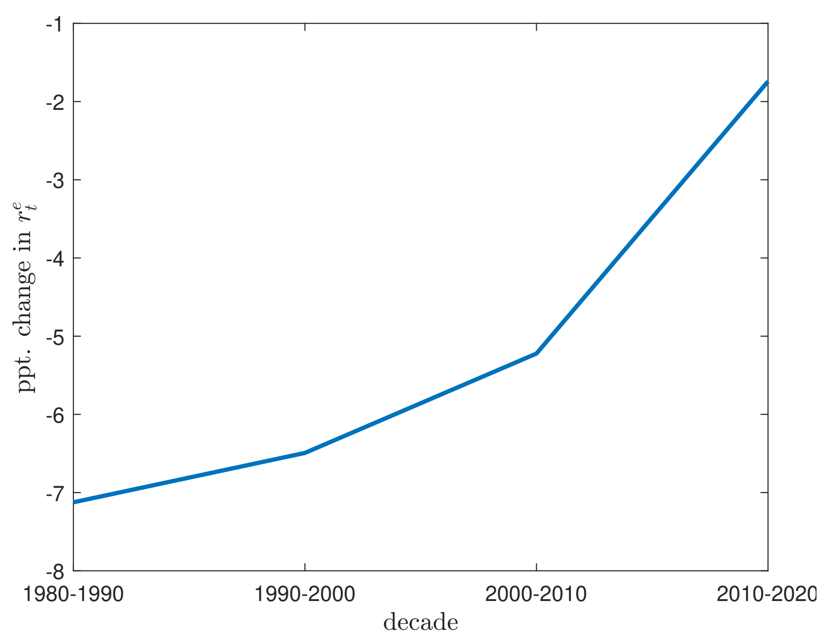


Figure A.3: Occupational Hourly Wage by Occupational Complexity.

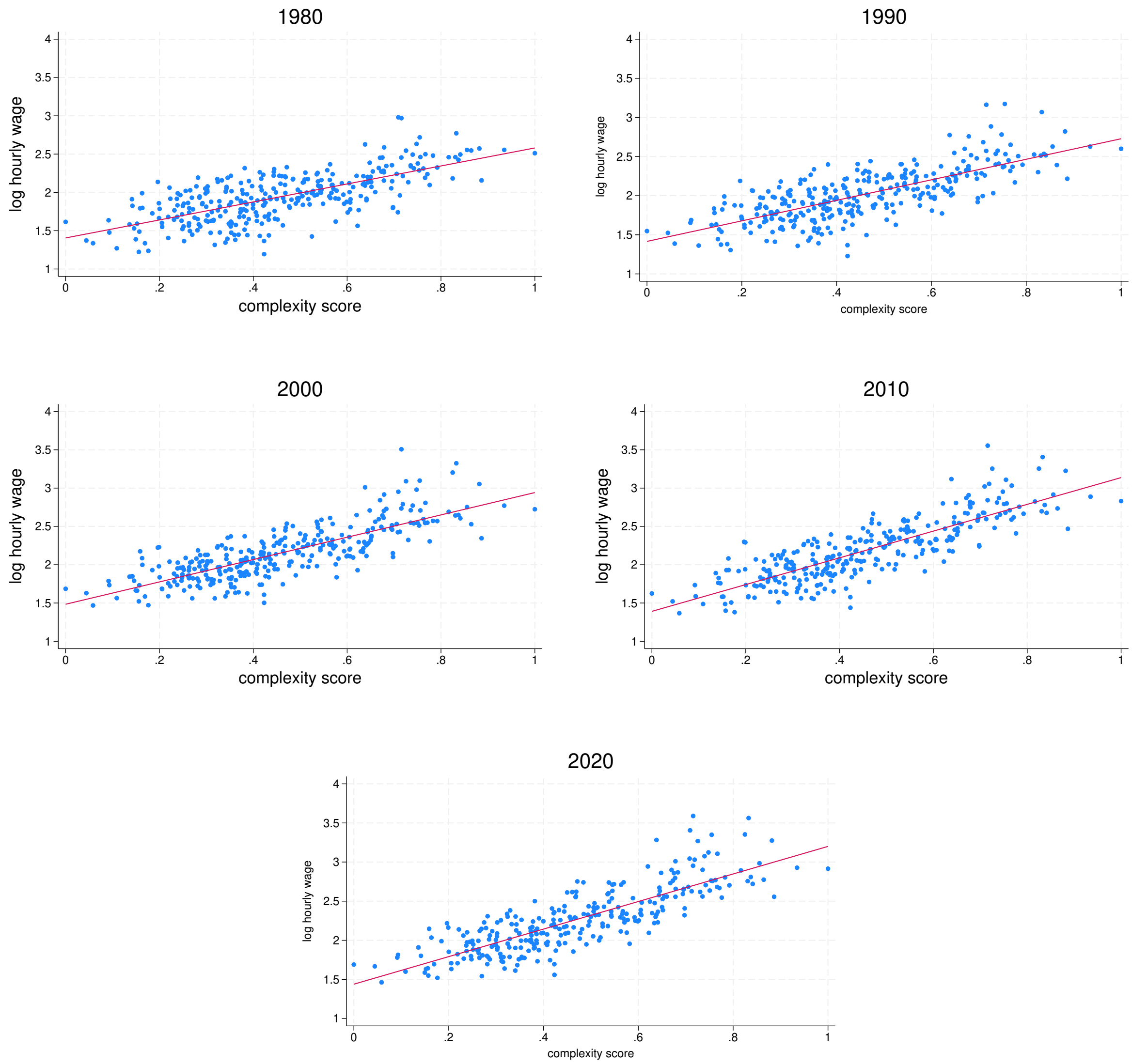


Figure A.4: Occupation Employment Share by Occupational Complexity.

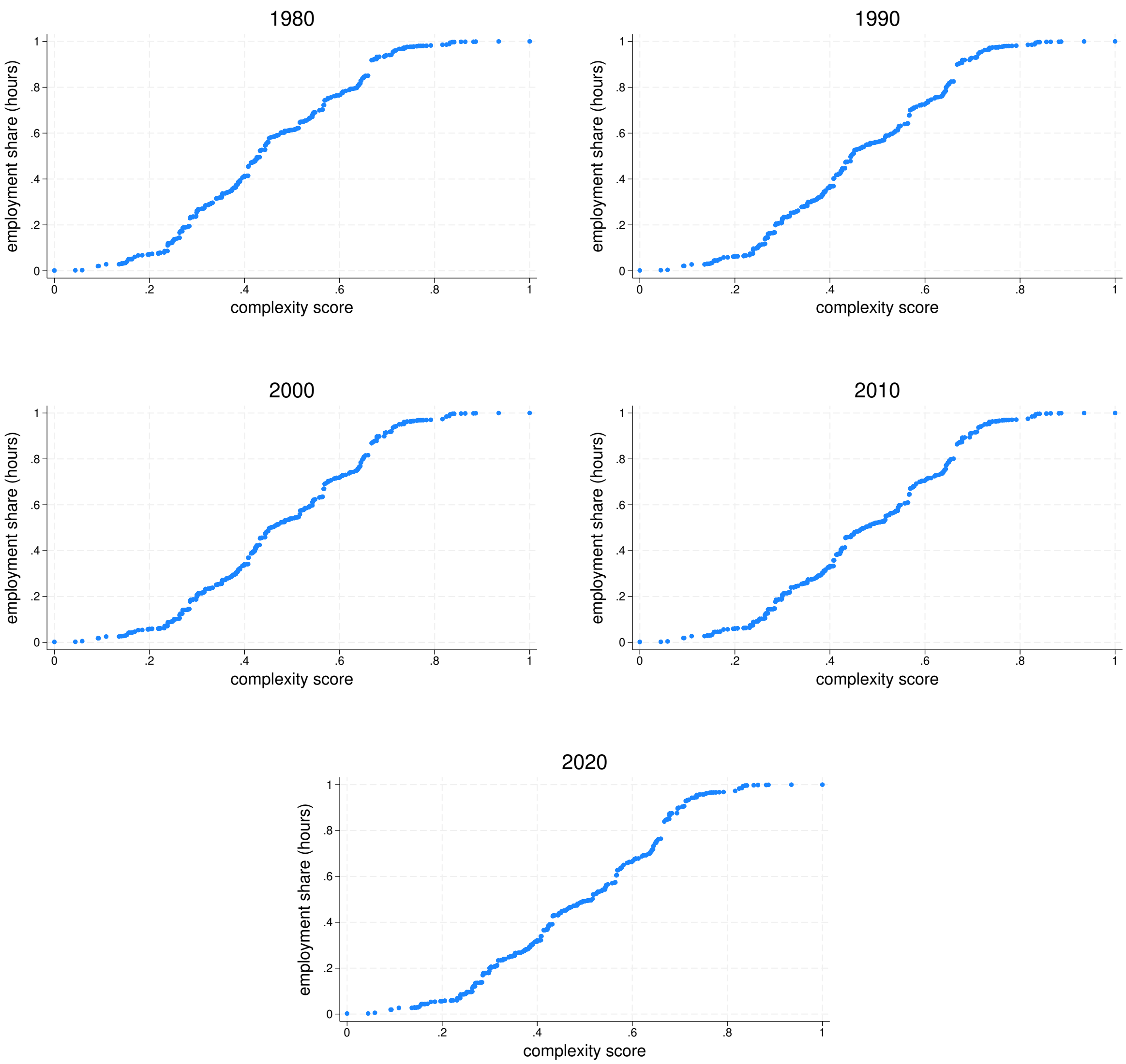


Figure A.5: Occupational Hourly Wage by Occupational Complexity.

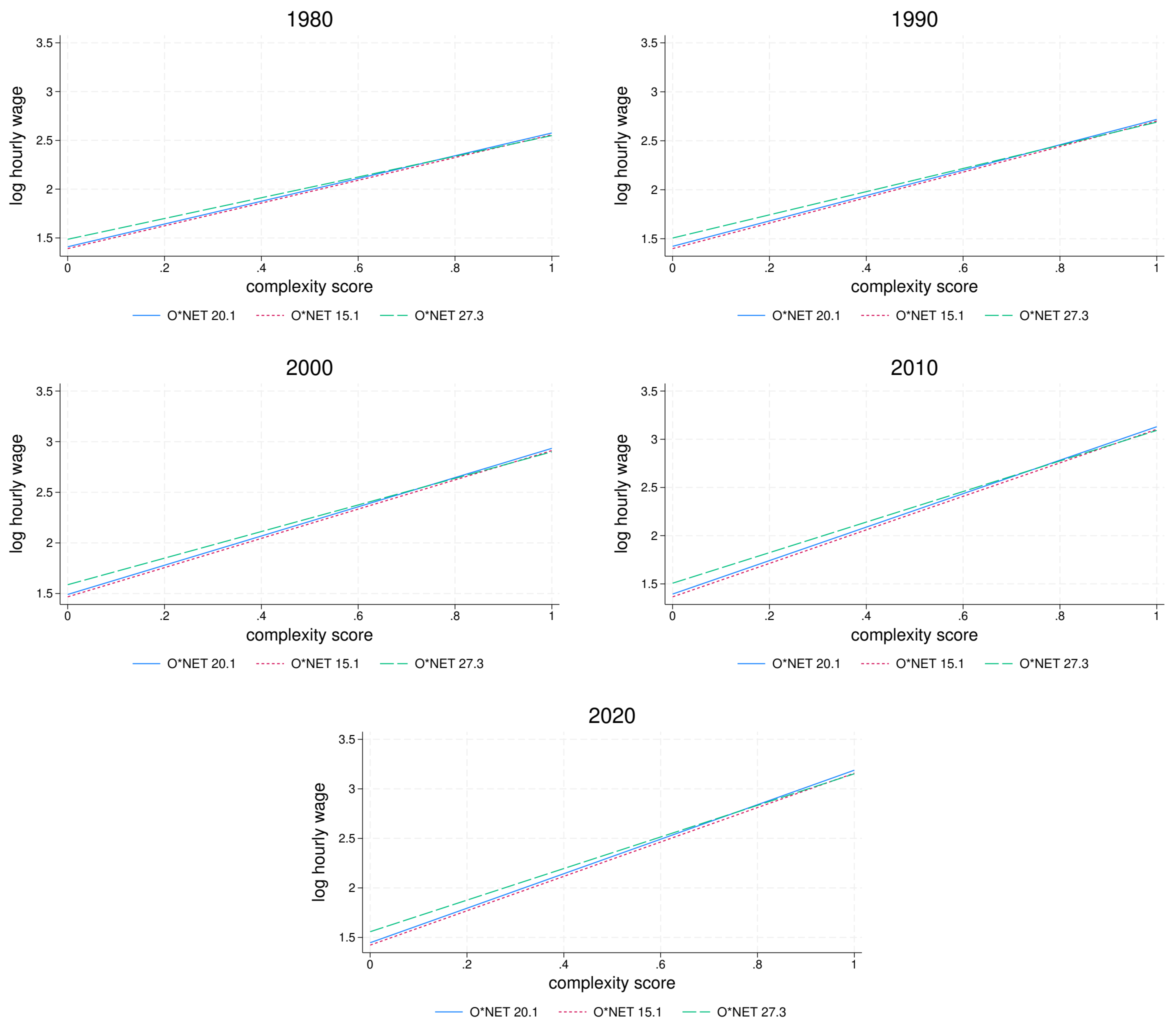


Figure A.6: Occupational Hourly Wage by Occupational Complexity Percentile.

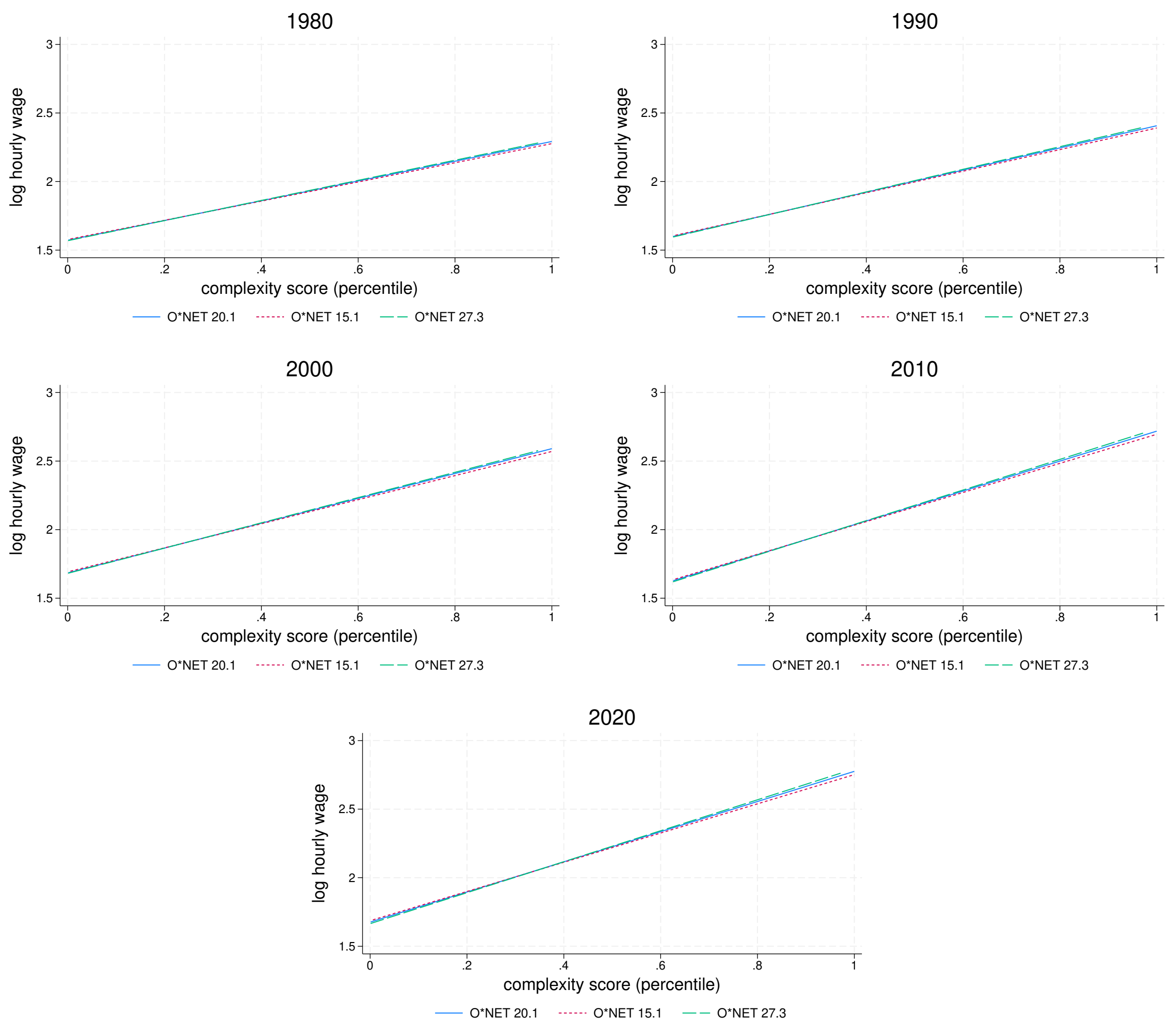


Figure A.7: Actual vs. Predicted Log-Wages, Parametric Skill Price Function.

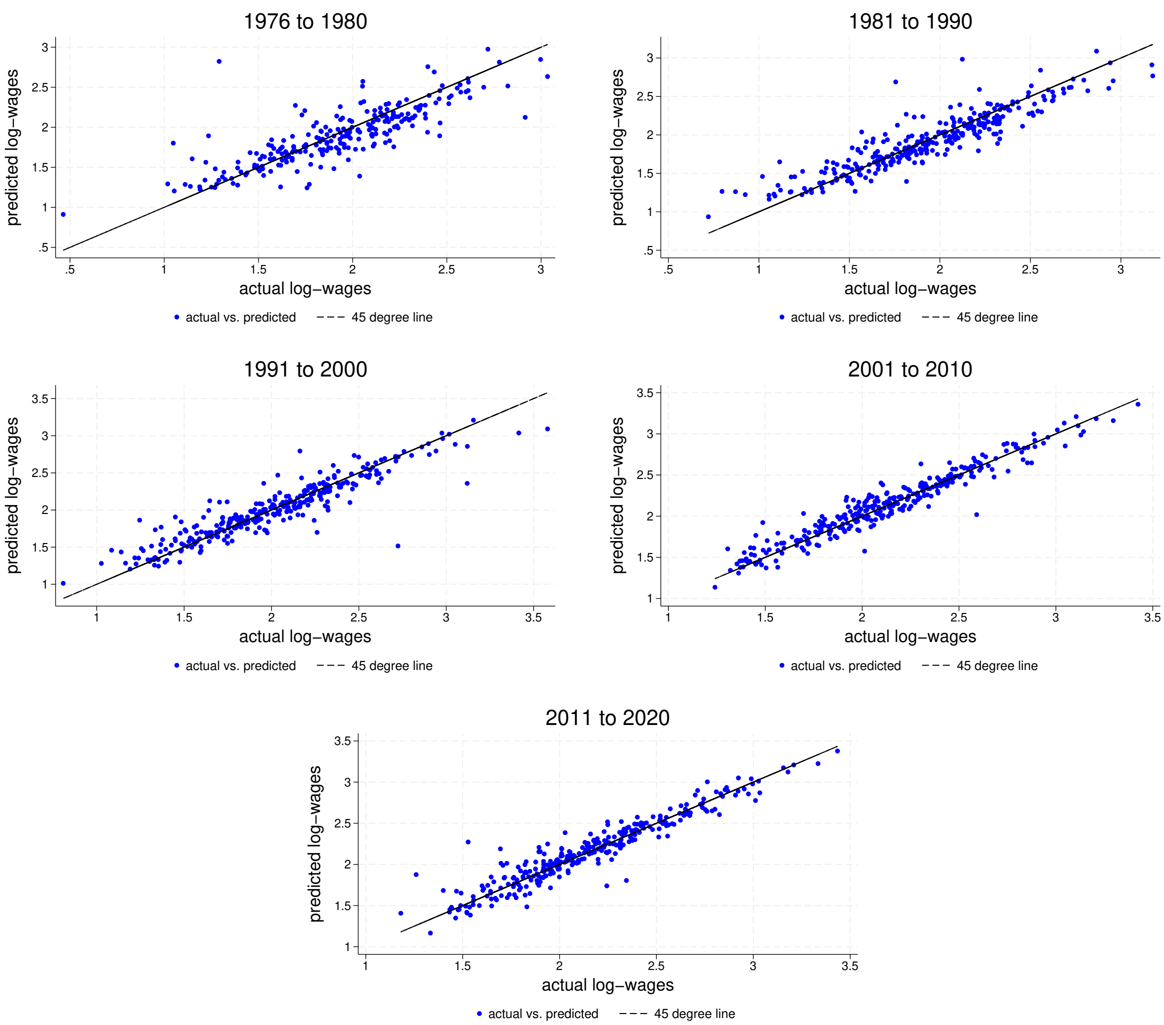


Figure A.8: Model Fit, Hourly Wage.

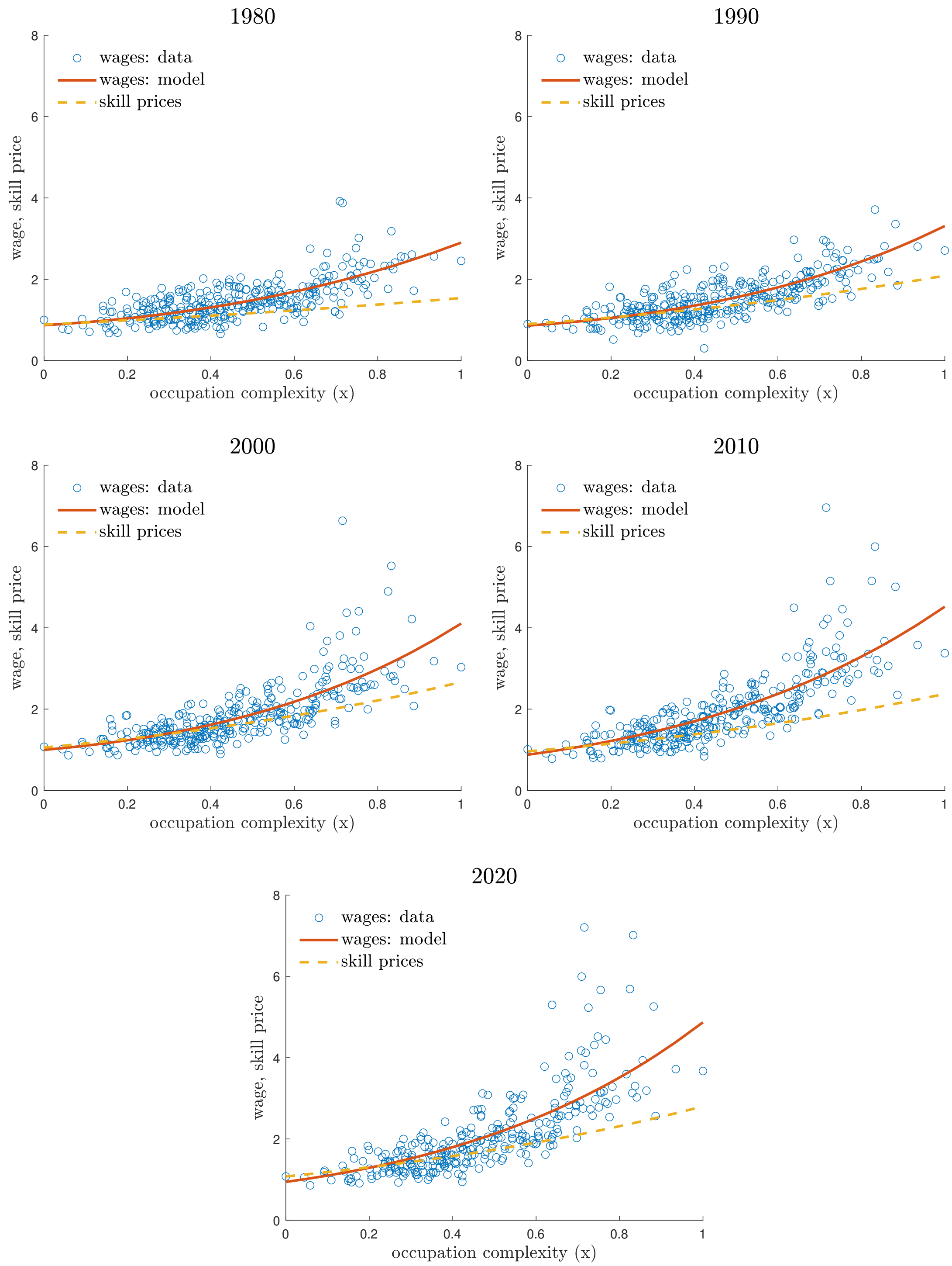


Figure A.9: Model Fit, Log Hourly Wage.

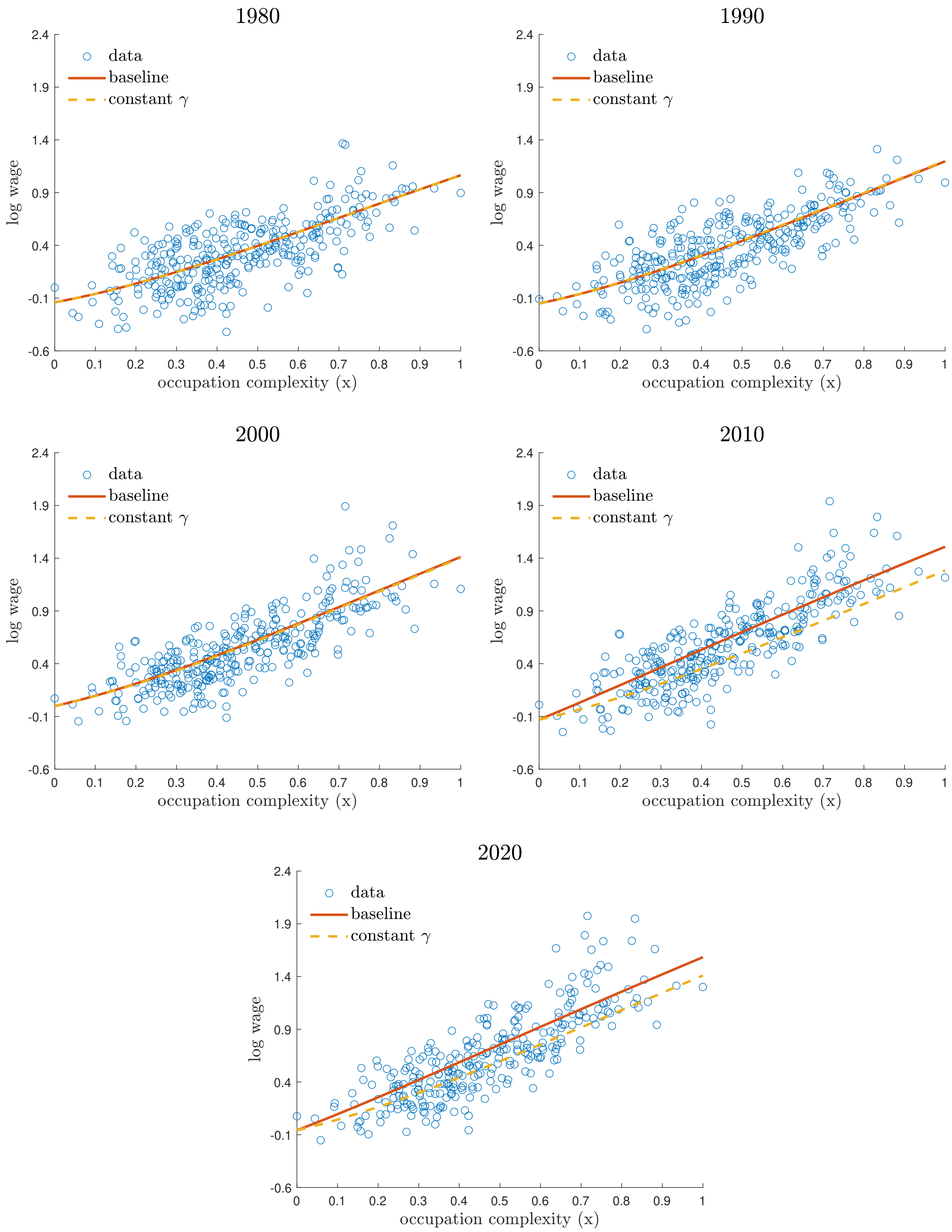


Figure A.10: Model Fit, Policy Function.

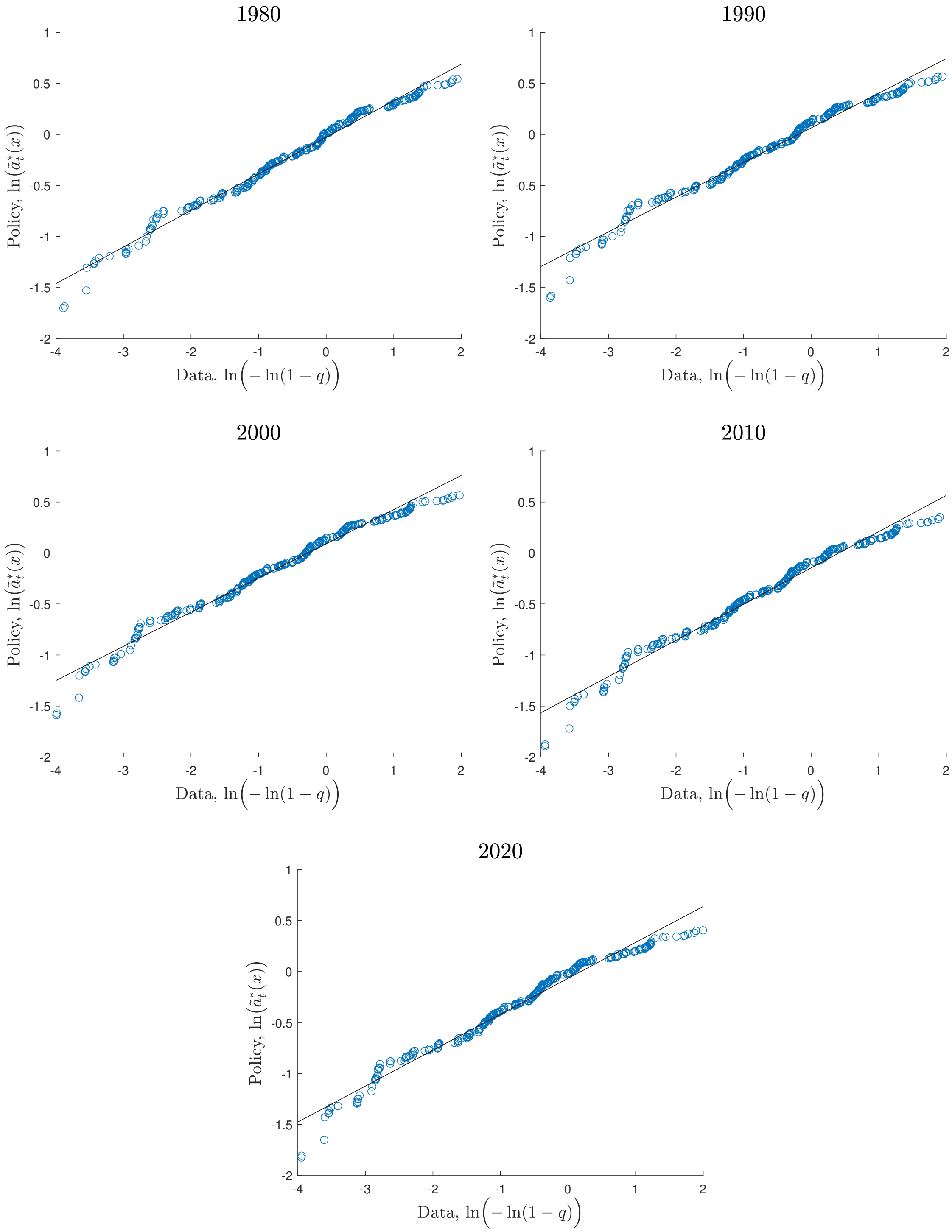
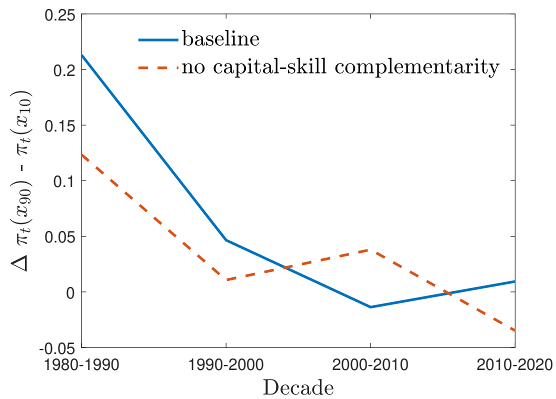
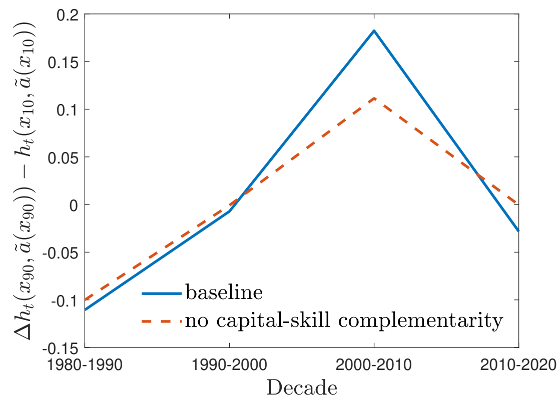


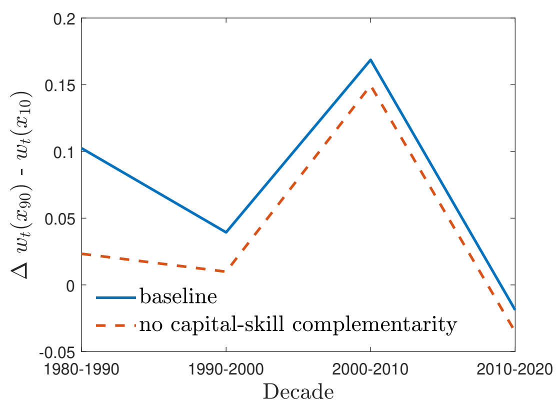
Figure A.11: Wage, Skill Price, and Labor Supply Changes by Decade without Capital-Skill Complementarity



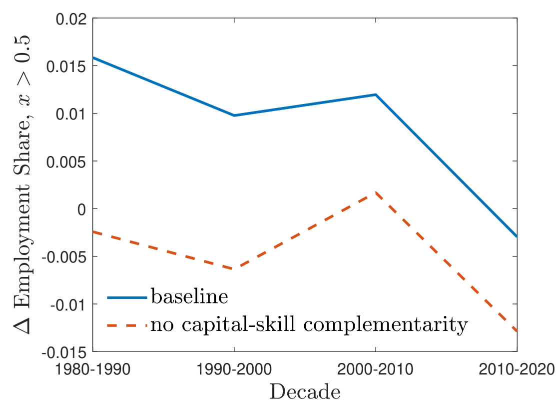
(a) Skill Prices



(b) Efficiency Units



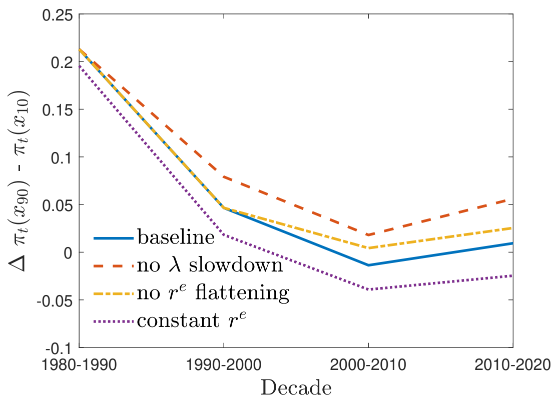
(c) Wages



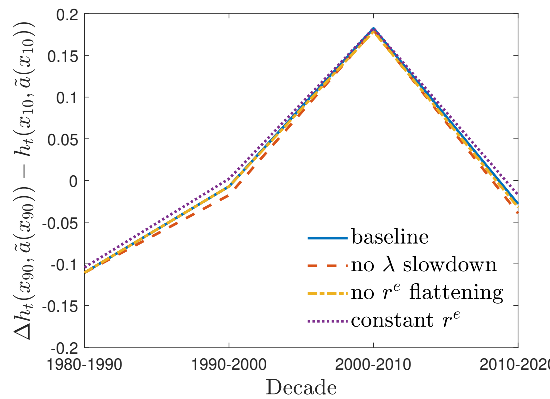
(d) Employment

Notes: The charts are difference-in-differences showing, for each decade, the change in the difference in (panel a) skill prices, (panel b) occupation efficiency units, and (panel c) log hourly wages between the occupations for which $x = 0.9$ and $x = 0.1$. Panel d shows, for each decade the change in the share of workers employed in occupations for which $x > 0.5$.

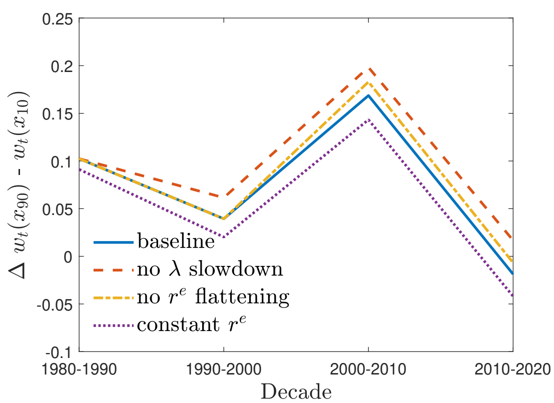
Figure A.12: Wage, Skill Price, and Labor Supply Changes by Decade under Counterfactual Demand Changes



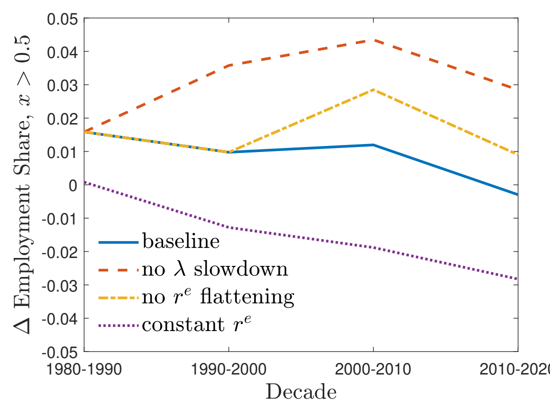
(a) Skill Prices



(b) Efficiency Units



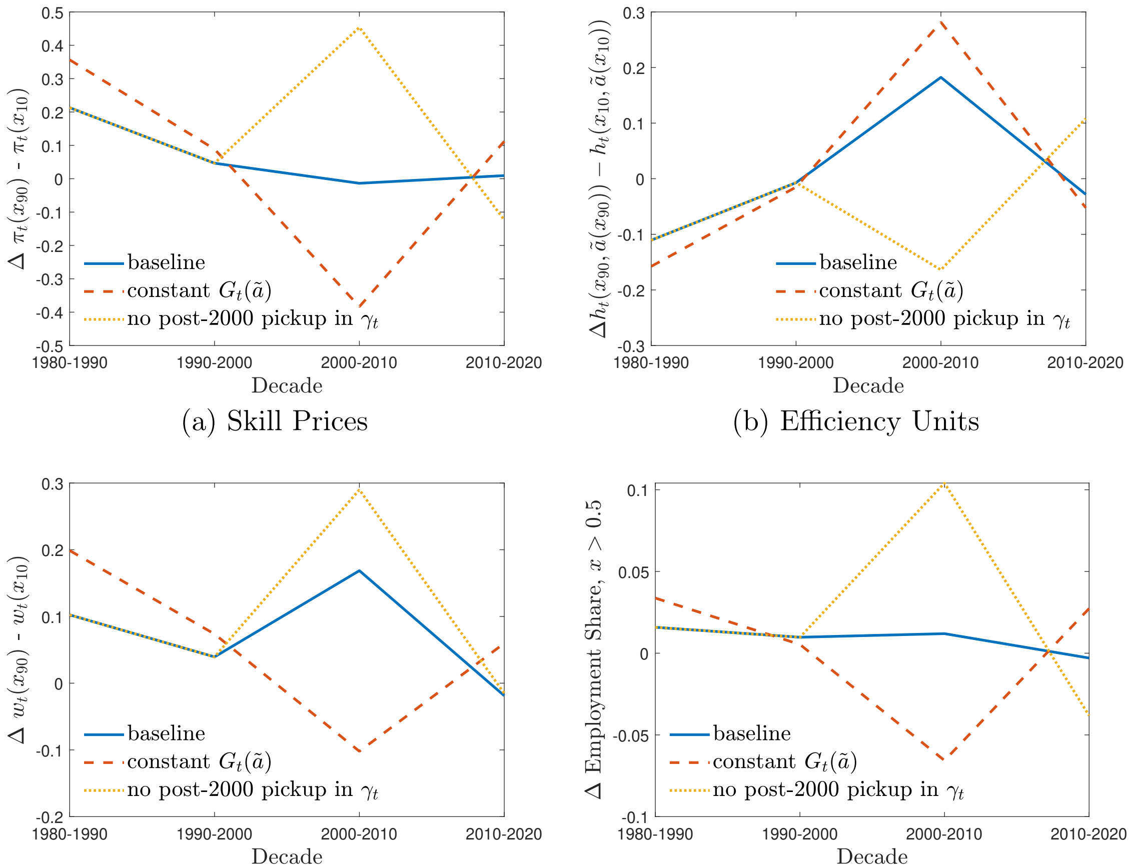
(c) Wages



(d) Employment

Notes: The charts are difference-in-differences showing, for each decade, the change in the difference in (panel a) skill prices, (panel b) occupation efficiency units, and (panel c) log hourly wages between the occupations for which $x = 0.9$ and $x = 0.1$. Panel d shows, for each decade the change in the share of workers employed in occupations for which $x > 0.5$.

Figure A.13: Wage, Skill Price, and Labor Supply Changes by Decade under Counterfactual Supply Changes



(a) Skill Prices

(b) Efficiency Units

(c) Wages

(d) Employment

Notes: The charts are difference-in-differences, showing, for each decade, the change in the difference in (panel a) skill prices, (panel b) occupation efficiency units, and (panel c) log hourly wages between the occupations for which $x = 0.9$ and $x = 0.1$. Panel d shows, for each decade the change in the share of workers employed in occupations for which $x > 0.5$.

Table A.1: O*NET Questions and PCA Factor Loadings

Descriptor	Factor Loading
O*NET Worker Abilities	
Oral Comprehension	0.1804
Written Comprehension	0.1848
Oral Expression	0.1753
Written Expression	0.1792
Fluency of Ideas	0.1812
Originality	0.1766
Problem Sensitivity	0.1808
Deductive Reasoning	0.1901
Inductive Reasoning	0.1850
Information Ordering	0.1804
Category Flexibility	0.1764
Mathematical Reasoning	0.1676
Number Facility	0.1582
Memorization	0.1613
Speed of Closure	0.1648
Flexibility of Closure	0.1415
Perceptual Speed	0.0858
O*NET Skills	
Mathematics	0.1567
Science	0.1420
Critical Thinking	0.1843
Active Learning	0.1879
Complex Problem Solving	0.1884
Programming	0.1328
Judgment and Decision Making	0.1868
Systems Analysis	0.1858
Systems Evaluation	0.1856
O*NET Activities	
Monitor Processes, Materials or Surroundings	0.1180
Judging the Qualities of Things/Services/People	0.1423
Processing Information	0.1675
Evaluating Information to Determine Compliance with Standards	0.1517
Analyzing Data or Information	0.1822
Making Decisions and Solving Problems	0.1788
Thinking Creatively	0.1620
Updating and Using Relevant Information	0.1764
Developing Objectives and Strategies	0.1654

Table A.2: Relationship between Complexity and Hourly Wage/Employment

	Δ log hourly wage 1980-2020			Δ log hours 1980-2020		
	(1)	(2)	(3)	(4)	(5)	(6)
Complexity	0.546***	0.549***	0.442***	0.000413	-0.000724**	-0.003558***
	(28.92)	(27.78)	(16.19)	(1.19)	(-2.00)	(-5.53)
RTI		0.011	0.021		0.00393***	-0.00367***
		(0.55)	(1.04)		(-10.58)	(-9.80)
Social Skill			0.111***			0.00292***
			(3.72)			(5.33)
N	15648	15648	15648	15782	15782	15782

Note: Regressions include fixed effects for sex, educational attainment, race and age
t-statistics are in parentheses. Significance levels are: *** 1%, ** 5%, * 10%.

Table A.3: Occupation-level Relationship between Complexity and Hourly Wage, by Decade

	Δ log hourly wage 1980-1990			Δ log hourly wage 1990-2000			Δ log hourly wage 2000-2010			Δ log hourly wage 2010-2020		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Complexity	-0.0187	-0.0147	-0.105	0.280***	0.269***	0.198**	0.262***	0.266***	0.313***	-0.0102	-0.0117	-0.0134
	(-0.36)	(-0.28)	(-1.65)	(4.46)	(4.24)	(2.59)	(5.54)	(5.61)	(5.50)	(-0.18)	(-0.21)	(-0.20)
RTI		0.0138	0.0277		-0.0394	-0.0298		0.0269	0.0200		-0.0110	-0.0108
		(0.44)	(0.87)		(-1.05)	(-0.79)		(1.04)	(0.76)		(-0.38)	(-0.36)
Social Skill			0.111**			0.0846*			-0.0550			0.00192
			(2.54)			(1.65)			(-1.48)			(0.04)
N	317	317	317	317	317	317	310	310	310	292	292	292

Note: Regressions include demographic controls: occupation-level, base-period means of the share of workers in an occupation with a college/high-school degree, the share of workers in an occupation who are non-white, the share of workers in an occupation who are married, the share of female workers in an occupation, the mean age of workers in an occupation, and the mean number of children of workers in an occupation. t-statistics are in parentheses. Significance levels are: *** 1%, ** 5%, * 10%.

Table A.4: Occupation-level Relationship between Complexity and Employment Share, by Decade

	Δ log hours 1980-1990			Δ log hours 1990-2000			Δ log hours 2000-2010			Δ log hours 2010-2020		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Complexity	0.00153	0.00136	-0.000228	0.000953	0.000679	-0.000634	0.000527	0.000355	-0.000580	0.00146	0.00129	0.00227**
	(1.56)	(1.37)	(-0.19)	(0.73)	(0.51)	(-0.40)	(0.55)	(0.37)	(-0.51)	(1.64)	(1.46)	(2.13)
RTI		-0.000576	-0.000331		-0.00103	-0.000858		-0.00114**	-0.000998*		-0.00123***	-0.00136***
		(-0.98)	(-0.56)		(-1.32)	(-1.09)		(-2.17)	(-1.88)		(-2.64)	(-2.90)
Social Skill			0.00195**			0.00156			0.00110			-0.00112
			(2.39)			(1.46)			(1.47)			(-1.64)
N	317	317	317	317	317	317	310	310	310	292	292	292

Note: Regressions include demographic controls: occupation-level, base-period means of the share of workers in an occupation with a college/high-school degree, the share of workers in an occupation who are non-white, the share of workers in an occupation who are married, the share of female workers in an occupation, the mean age of workers in an occupation, and the mean number of children of workers in an occupation. t-statistics are in parentheses. Significance levels are: *** 1%, ** 5%, * 10%.

Table A.5: Relationship between Complexity and Hourly Wage, by Decade

	Δ log hourly wage 1980-1990			Δ log hourly wage 1990-2000			Δ log hourly wage 2000-2010			Δ log hourly wage 2010-2020		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Complexity	0.0637***	0.0652***	-0.0148	0.111***	0.103***	-0.0110	0.295***	0.307***	0.337***	0.0385**	0.0296*	0.0434
	(4.15)	(4.03)	(-0.52)	(7.72)	(6.81)	(-0.41)	(20.12)	(19.92)	(12.18)	(2.31)	(1.70)	(1.39)
RTI		0.00475	0.0126		-0.0273*	-0.0163		0.0412**	0.0382**		-0.0313*	-0.0326*
		(0.28)	(0.74)		(-1.74)	(-1.03)		(2.56)	(2.36)		(-1.75)	(-1.81)
Social Skill			0.0828***			0.118***			-0.0313			-0.0412
			(3.43)			(5.20)			(-1.32)			(-0.53)
N	17350	17350	17350	17827	17827	17827	17152	17152	17152	15953	15953	15953

Note: Regressions include fixed effects for sex, educational attainment, race and age. t-statistics are in parentheses. Significance levels are: *** 1%, ** 5%, * 10%.

Table A.6: Relationship between Complexity and Employment, by Decade

	Δ log hours 1980-1990			Δ log hours 1990-2000			Δ log hours 2000-2010			Δ log hours 2010-2020		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Complexity	-0.000114	-0.000307**	-0.00103***	0.000738****	0.000540***	-0.00132***	-0.000821***	-0.00117***	-0.00247***	0.000541***	0.000149	0.00114***
	(-0.78)	(-1.99)	(-3.80)	(4.99)	(3.47)	(-4.81)	(-7.39)	(-10.04)	(-11.78)	(5.34)	(1.41)	(6.06)
RTI		-0.000635***	-0.000564***		-0.000661***	-0.000481***		-0.00118***	-0.00105***		-0.00137***	-0.00146***
		(-3.97)	(-3.49)		(-4.10)	(-2.96)		(-9.62)	(-8.52)		(-12.69)	(-13.43)
Social Skill			0.000748***			0.00192***			0.00134***			-0.00102***
			(3.24)			(8.24)			(7.44)			(-6.34)
N	17492	17492	17492	17934	17934	17934	17255	17255	17255	16049	16049	16049

Note: Regressions include fixed effects for sex, educational attainment, race and age. t-statistics are in parentheses. Significance levels are: *** 1%, ** 5%, * 10%.

B Proofs

B.1 Proof of Proposition 1

For notational clarity, we suppress the time subscripts in the following. Taking the cross-partials of $\ln(h(x,\tilde{a}))$ delivers, after some straightforward algebra, the expression

$$\frac{\partial \ln(h(x,\tilde{a}))}{\partial x \partial \tilde{a}} = [\gamma(x)\cdot(\tilde{a})^{\eta} + \delta(x)]^{-2}\cdot(\tilde{a})^{\eta-1}\cdot\left[\gamma^{'}(x)\cdot\delta(x) - \gamma(x)\cdot\delta^{'}(x)\right]. \tag{A.1}$$

For the twice-differentiable case considered here, log-supermodularity is equivalent with non-negativity of the cross-partial, and this is the case if $\frac{\gamma^{'}(x)}{\gamma(x)} \geq \frac{\delta^{'}(x)}{\delta(x)}$. This is a weaker condition than the combination of assumptions 1 and 2. The first statement in the Proposition is therefore established. With $\tilde{a}, \gamma(x), \delta(x)$ all assumed to be positive, $(h(x,\tilde{a})) \geq 0$ is trivially satisfied.

This leaves us to prove the implication of assumption 3 and of the full-support condition. Let $\varphi(x) \equiv \eta\cdot\Phi(x)$. The first-order condition in equation (4) yields the following expression for the inverse policy function (see also equation (6) in the text):

$$(\tilde{a}^{*}(x))^{\eta} = \frac{-\delta^{'}(x) - \varphi(x)\cdot\delta(x)}{\gamma^{'}(x) + \varphi(x)\cdot\gamma(x)}. \tag{A.2}$$

Assumptions 1 and 2 imply that the right-hand side is a non-constant function of x. The case with $\eta = 0$ yields a contradiction because the left-hand side would be equal to one for all x. Hence $\eta \neq 0$. The full-support condition evaluated at $x = 0$ is given by $\tilde{a}^{*}(0) = 0$, which implies that when $\eta < 0$ the left hand side of A.2 goes to ∞ as $x \to 0$. At $x = 0$ the denominator on the right hand side is positive, given assumptions 1 and 2. Consequently, the right hand side only goes to ∞ if the numerator goes to $-\infty$. Given assumptions 1 and 2, the first term in the numerator is positive, while the second is a bounded, non-positive term. Equation A.2 is therefore inconsistent with the full-support condition when $\eta < 0$. Since η is the parameter of the CES-aggregator $h_t(x,a,d)$ and is thus bounded above by 1, we have established that it must be in $(0,1]$.

It remains to be shown that the inverse policy function is strictly increasing on $x \in [0,1]$, and because $\eta > 0$ this is satisfied whenever $\frac{\partial((\tilde{a}^{*}(x))^{\eta})}{\partial x} \geq 0$. We therefore only need to take the derivative of the right-hand side of equation (A.2) and show that it is non-negative. After some algebra one gets the following equation:

$$\begin{aligned}\left[\gamma^{'}(x) + \varphi(x)\cdot\gamma(x)\right]\cdot\frac{\partial((\tilde{a}^{*}(x))^{\eta})}{\partial x} &= -\left(\gamma^{''}(x)\cdot(\tilde{a}^{*}(x))^{\eta} + \delta^{''}(x)\right) + \dots \\ &\quad \left(\varphi(x)^{2} - \varphi^{'}(x)\right)\cdot[h(x,\tilde{a}^{*}(x))]^{\eta}.\end{aligned} \tag{A.3}$$

PAM is therefore satisfied if and only if

$$-\gamma^{''}(x)\cdot(\tilde{a}^{*}(x))^{\eta} + \left(\varphi(x)^{2} - \varphi^{'}(x)\right)\cdot[h(x,\tilde{a}^{*}(x))]^{\eta} \geq \delta^{''}(x). \tag{A.4}$$

We show that the left-hand side of this inequality is positive, which proves the last part of the Proposition because $\delta^{''}(x) < 0$ by assumption 2.

First consider the term $-\gamma^{''}(x)\cdot(\tilde{a}^{*}(x))^{\eta}$. Since $\gamma^{''}(x) < 0$ by assumption, it is nonnegative if $(\tilde{a}^{*}(x))^{\eta} \geq 0$. To establish this inequality, define $f(x) = -\left(\delta^{'}(x) + \varphi(x)\cdot\delta(x)\right)$ on $x \in [0,1]$, which is the numerator of $(\tilde{a}^{*}(x))^{\eta}$ in equation (A.2). We showed above that the full-support condition for $\tilde{a}^{*}(x)$ requires $f(0) = 0$. Assumptions 2 and 3 together with $\eta > 0$ implies that $f^{'}(x) > 0$. Hence,

Table A.7: Parametric Skill Price Function: Estimates

	(1)	(2)	(3)	(4)	(5)
	benchmark, age $\in [53, 57]$	(a) without Occ-FEv	age $\in [50, 59]$	exper $\in [28, 32]$	with Occ-Educ FE
Intercept	1.865***	1.230***	1.867***	1.855***	1.922***
	(0.014)	(0.077)	(0.013)	(0.015)	(0.014)
Complexity		1.374***			
		(0.138)			
Complexity $\times$ Time FE					
1981-1985	0.134***	0.175***	0.113***	0.159**	0.106**
	(0.047)	(0.058)	(0.038)	(0.077)	(0.064)
1986-1990	0.295***	0.348***	0.244***	0.226***	0.179***
	(0.064)	(0.078)	(0.047)	(0.060)	(0.055)
1991-1995	0.260***	0.327***	0.247***	0.227***	0.152**
	(0.068)	(0.088)	(0.060)	(0.077)	(0.081)
1996-2000	0.378***	0.452***	0.353***	0.370***	0.287***
	(0.088)	(0.124)	(0.073)	(0.088)	(0.072)
2001-2005	0.421***	0.529***	0.381***	0.423***	0.312***
	(0.082)	(0.114)	(0.075)	(0.086)	(0.081)
2006-2010	0.354***	0.476***	0.341***	0.402***	0.292***
	(0.091)	(0.130)	(0.086)	(0.094)	(0.089)
2011-2015	0.319***	0.442***	0.328***	0.430***	0.302***
	(0.093)	(0.127)	(0.085)	(0.098)	(0.091)
2016-2020	0.405***	0.556***	0.378***	0.438***	0.303***
	(0.097)	(0.133)	(0.089)	(0.087)	(0.083)
Predicted DiD in Log-points	40.5	55.6	37.8	43.8	30.3
R-squared	0.27	0.20	0.27	0.27	0.33
N	282,439	282,439	577,753	342,149	342,149

Notes: The table shows estimates for two key parameters of the parametric skill-price function from the flat-spot method. “FE” stands for “Fixed Effects”; “DiD” stands for “Difference in Differences”. “DiD in Log-Points” is the predicted growth in the low-wage difference between the most- and least complex occupations from 1980 to 2020. Models (a), (c) and (d) include occupation FE, where we use occupational codes from David Dorn. Model (b) includes complexity as control. Model (e) includes occupation-by-education fixed effects. All Models include time effects.

$f(x) \geq 0$ on its support. The denominator of the inverse policy function is strictly positive because of assumption 1. This shows that $(\tilde{a}^*(x))^\eta \geq 0$.

Next consider the second term in the inequality. Assumption 3 together with $\eta > 0$ implies that $\left(\varphi(x)^2 - \varphi^{'}(x)\right) > 0$. Futhermore, $(h(x, \tilde{a}^*(x))) \geq 0$ is trivially nonnegative as long as $\tilde{a}^*(x)) \geq 0$, the latter of which was established above. Hence, the second term on the left-hand side of the condition for PAM is satisfied and the proposition is proven.

B.2 Proof of Proposition 2

We start with a technical Lemma:

Lemma 1. *Let $j = 2, 3, ..., J$, with $J \ll \infty$. Define the list $\omega_{j,t} = \langle \left\{ \lambda_{m,t}, \left(\frac{k_t^e}{l_{m,t}}\right) \right\}_{m=j+1}^{J}, \{\sigma_l\}_{l=j}^{J} \rangle \in \Omega_j$, where $\Omega_j = \left(\mathbb{R}_0^+\right)^{2(J-j)} \times (-\inf, 1]^{J-j+1}$ is of dimension $3 \cdot (J-j)+1$. Then there is a list of functions $\{H_j(\omega_{j,t})\}_{j=1}^{J} : \Omega_j \to \mathbb{R}_0^+$ such that*

$$\widetilde{l}_{j,t} \quad = \quad \lambda_{j,t} \cdot (l_{j,t})^{\sigma_j} + (k_t^e)^{\sigma_j} \cdot H_j(\omega_{j,t}) \tag{A.5}$$

.

Proof. We prove this result by induction from $j = J$ backwards to $j = 2$.
The equation holds by definition for $j = J$, with $H_{J,t} = 1$. For the definition of $H_{J,t}$ to be consistent with the Lemma, we define the function $H_J(\omega_{J,t}) = 1$ for any $\omega_{J,t} \in \mathbb{R}_0^+$.
Now assume that the equation holds for any $m = j+1, j+2, ..., J$. We can then write

$$\begin{aligned}
\widetilde{l}_{j,t} \quad &= \quad \lambda_{j,t} \cdot (l_{j,t})^{\sigma_j} + \left(\widetilde{l}_{j+1,t}\right)^{\frac{\sigma_j}{\sigma_{j+1}}} \\
&= \quad \lambda_{j,t} \cdot (l_{j,t})^{\sigma_j} + \{\lambda_{j+1,t} \cdot (l_{j+1,t})^{\sigma_{j+1}} + (k_t^e)^{\sigma_{j+1}} \cdot H_{j+1}(\omega_{j+1,t})\}^{\frac{\sigma_j}{\sigma_{j+1}}} \\
&= \quad \lambda_{j,t} \cdot (l_{j,t})^{\sigma_j} + (k_t^e)^{\sigma_j} \cdot \left\{ \lambda_{j+1,t} \cdot \left(\frac{k_t^e}{l_{j+1,t}}\right)^{-(\sigma_{j+1})} + H_{j+1}(\omega_{j+1,t}) \right\}^{\frac{\sigma_j}{\sigma_{j+1}}}
\end{aligned} \tag{A.6}$$

The first equation is the definition of $\widetilde{l}_{j,t}$, the second equation uses the induction argument, and the third equation follows from some simple algebraic manipulations. The equation in the Lemma then holds by defining the list of functions $\{H_j(\omega_{j,t})\}_{j=1}^{J}$ using the backward recursion

$$H_j(\omega_{j,t}) \quad = \quad \left\{ \lambda_{j+1,t} \cdot \left(\frac{k_t^e}{l_{j+1,t}}\right)^{-(\sigma_{j+1})} + H_{j+1}(\omega_{j+1,t}) \right\}^{\frac{\sigma_j}{\sigma_{j+1}}} \tag{A.7}$$

for $j < J$. As shown above, the initialization is $H_J = 1$. In each step of this recursion the list $\omega_{j,t}$ is constructed from attaching the list $\langle \lambda_{J+1,t}, \left(\frac{k_t^e}{l_{j+1,t}}\right), \sigma_j, \sigma j + 1 \rangle$ to $\omega_{j+1,t}$. The first two of these objects are in $\mathbb{R}_0^+$, while the $\sigma' s$ are contained in $(-\infty, 1]$. This finalizes the construction of the list of functions $\{H_j(\omega_{j,t})\}_{j=1}^{J}$ and their arguments satisfying the properties stated in the lemma. □

We now prove the proposition. Repeated applications of the chain rule to the production function delivers the following expression for the MRTS between type-j labor and capital equipment:

$$
\begin{aligned}
MRTS(l_{j,t}, k_t^e) &= \lambda_{j,t} \cdot (l_{j,t})^{\sigma_j - 1} \cdot (k_t^e)^{1-\sigma_J} \cdot \prod_{m=j+1}^{J} \left(\widetilde{l}_{m,t}\right)^{\frac{\sigma_m - \sigma_{m-1}}{\sigma_m}} \\
&= \lambda_{j,t} \cdot \left(\frac{k_t^e}{l_{j,t}}\right)^{1-\sigma_j} \cdot (k_t^e)^{\sigma_j - \sigma_J} \cdot \prod_{m=j+1}^{J} \left(\widetilde{l}_{m,t}\right)^{\frac{\sigma_m - \sigma_{m-1}}{\sigma_m}}. \qquad \text{(A.8)}
\end{aligned}
$$

Using the result in Lemma 1 allows us to write

$$
\widetilde{l}_{m,t} = (l_{m,t})^{\sigma_m} \cdot \left[\lambda_{m,t} + \left(\frac{k_t^e}{l_{m,t}}\right)^{\sigma_m} \cdot \cdot H_m(\omega_{m,t})\right] \qquad \text{(A.9)}
$$

and hence

$$
\left(\widetilde{l}_{m,t}\right)^{\frac{\sigma_m - \sigma_{m-1}}{\sigma_m}} = (l_{m,t})^{\sigma_m - \sigma_{m-1}} \cdot \left[\lambda_{m,t} + \left(\frac{k_t^e}{l_{m,t}}\right)^{\sigma_m} \cdot \cdot H_m(\omega_{m,t})\right]^{\frac{\sigma_m - \sigma_{m-1}}{\sigma_m}}. \qquad \text{(A.10)}
$$

For notational simplicity define $g_{m,t} = \left[\lambda_{m,t} + \left(\frac{k_t^e}{l_{m,t}}\right)^{\sigma_m} \cdot H_m(\omega_{m,t})\right]^{\frac{\sigma_m - \sigma_{m-1}}{\sigma_m}}$, which is a function of $\omega_{m,t}$. Putting together the equations above yields:

$$
\begin{aligned}
MRTS(l_{j,t}, k_t^e) &= \lambda_{j,t} \cdot \left(\frac{k_t^e}{l_{j,t}}\right)^{1-\sigma_j} \cdot (k_t^e)^{\sigma_j - \sigma_J} \cdot \prod_{m=j+1}^{J} (l_{m,t})^{\sigma_m - \sigma_{m-1}} \cdot \prod_{m=j+1}^{J} g_{m,t} \\
&= \lambda_{j,t} \cdot \left(\frac{k_t^e}{l_{j,t}}\right)^{1-\sigma_j} \cdot (k_t^e)^{(\sigma_j - \sigma_{j+1}) + (\sigma_{j+1} - \sigma_{j+2}) + \ldots + (\sigma_{J-1} - \sigma_J)} \cdot \ldots \\
&\quad \prod_{m=j+1}^{J} (l_{m,t})^{\sigma_m - \sigma_{m-1}} \cdot \prod_{m=j+1}^{J} g_{m,t} \\
&= \lambda_{j,t} \cdot \left(\frac{k_t^e}{l_{j,t}}\right)^{1-\sigma_j} \cdot \prod_{m=j+1}^{J} \left(\frac{k_t^e}{l_{m,t}}\right)^{\sigma_{m-1} - \sigma_m} \cdot \prod_{m=j+1}^{J} g_{m,t} \qquad \text{(A.11)}
\end{aligned}
$$

We notice that all objects appearing in the last two products are contained in the list $\omega_{j,t}$. We therefore establish the proposition by defining

$$
G_j(\omega_{j,t}) = \prod_{m=j+1}^{J} \left(\frac{k_t^e}{l_{m,t}}\right)^{\sigma_{m-1} - \sigma_m} \cdot \prod_{m=j+1}^{J} g_{m,t}. \qquad \text{(A.12)}
$$

B.3 Proof of Proposition 3

The condition for PAM in the case where only assumptions 1 to 3 are imposed is given by the inequality (A.4). Using the parametric assumptions 5 reduces this sufficient condition to

$$
\varphi(x)^2 \cdot [h(x, \tilde{a}^*(x))]^\eta \geq \delta''(x). \qquad \text{(A.13)}
$$

Since $\delta''(x) < 0$, we only need to show that $[h(x,\tilde{a}^*(x))]^\eta = \frac{\left(\delta(x)\cdot\gamma'(x)-\delta'(x)\cdot\gamma(x)\right)}{\gamma'(x)+\eta\cdot\Phi(x)\cdot\gamma(x)} > 0$ (see equation (7)), and, given that the denominator is positive, this is the case if $\delta(x)\cdot\gamma'(x) - \delta'(x)\cdot\gamma(x) > 0$. This log-supermodularity condition is indeed satisfied given our assumptions. This follows from using assumption 5, as shown by the following sequence of inequalities:

$$\begin{aligned}
\delta(x)\cdot\gamma'(x) - \delta'(x)\cdot\gamma(x) &= \delta_1 - (x+\delta_0)^c + (x+\gamma_0)\cdot c\cdot(x+\delta_0)^{c-1} && \text{(A.14)}\\
&\geq \delta_1 - (x+\delta_0)^c + (x+\delta_0)\cdot c\cdot(x+\delta_0)^{c-1} && \text{(A.15)}\\
&= \delta_1 + (c-1)\cdot(x+\delta_0)^c && \text{(A.16)}\\
&> 0. && \text{(A.17)}
\end{aligned}$$

The first inequality follows from $\gamma_0 > \delta_0$, and the last inequality follows from $c > 1$.

Next, we show that that $h(x,\tilde{a}^*(x))$ is an increasing function of x under the additional condition that $c > 2$. As before, since $\eta > 0$ we work with $[h(x,\tilde{a}^*(x))]^\eta$ instead. Differentiating equation (7) yields, after some algebra, the following equation:

$$\left[\gamma'(x) + \varphi\cdot\gamma(x)\right]\cdot\frac{\partial\,[h(x,\tilde{a}^*(x))]^\eta}{\partial x} = -\varphi\cdot\gamma'(x)\cdot[h(x,\tilde{a}^*(x))]^\eta - \gamma''(x)\cdot\gamma(x)\cdot[\tilde{a}^*(x)]^\eta - \delta''(x)\cdot\gamma(x). \tag{A.18}$$

Setting $\gamma'' = 0$ and multiplication by (positive) $\gamma'(x) + \varphi\cdot\gamma(x) > 0$ delivers the following sufficient condition for $h(x,\tilde{a}^*(x))$ to be an increasing function:

$$\gamma'(x)\cdot\left(\varphi\cdot\delta(x) + \gamma(x)\cdot\delta''(x)\right) + \varphi\cdot\gamma(x)\cdot\left(\gamma(x)\cdot\delta''(x) - \delta'(x)\right) \leq 0. \tag{A.19}$$

Using all conditions in assumption 5, this inequality implies the following:

$$\begin{aligned}
&\varphi\cdot(\delta_1-(x+\delta_0)^c) - (x+\gamma_0)\cdot c\cdot(c-1)\cdot(x+\delta_0)^{c-2} - \varphi\cdot(x+\gamma_0)\cdot c\cdot\left[(x+\gamma_0)\cdot(c-1)\cdot(x+\delta_0)^{c-2} - (x+\delta_0)^{c-1}\right]...\\
&< \varphi\cdot(\delta_1-(x+\delta_0)^c) - (x+\delta_0)\cdot c\cdot(c-1)\cdot(x+\delta_0)^{c-2} - \varphi\cdot(x+\gamma_0)\cdot c\cdot\left[(x+\delta_0)\cdot(c-1)\cdot(x+\delta_0)^{c-2} - (x+\delta_0)^{c-1}\right]...\\
&= \varphi\cdot(\delta_1-(x+\delta_0)^c) - c\cdot(c-1)\cdot(x+\delta_0)^{c-1} + \varphi\cdot(x+\gamma_0)\cdot c\cdot(x+\delta_0)^{c-1}\cdot[1-(c-1)]...\\
&< \varphi\cdot(\delta_1-(x+\delta_0)^c) - c\cdot(c-1)\cdot(x+\delta_0)^{c-1}...\\
&< \varphi\cdot(\delta_1-(x+\delta_0)^c) - c\cdot(x+\delta_0)^{c-1} = \varphi\cdot\delta(x) + \delta' \leq 0.
\end{aligned} \tag{A.20}$$

The first inequality uses the condition $\gamma_0 > \delta_0$; the second inequality holds because $c > 2$ implies that $1-(c-1) < 0$; the third inequality uses $c > 2$; and the last inequality is implied by the full-support condition and PAM.

B.4 Proof of Proposition 4

We need to show identification of $\{c, \Phi_0, \eta, \gamma_{0,t}, \delta_{0,t})$ ($\delta_{1,t}$ follows directly from $\delta_{0,t}$ and the full support condition). Starting with period 0, we can solve for the 5 parameters

$$\{c, \Phi_0, \eta, \gamma_{0,0}, \delta_{0,0}\}$$

from the following five equations:

$$\ln \tilde{w}_0(0) = \frac{1}{\eta} \cdot (\ln c - \ln \eta - \ln \Phi_0 + (c-1)\ln \delta_{0,0}) \tag{A.21}$$

$$\frac{\tilde{w}_0'(0)}{\tilde{w}_0(0)} = \Phi_0 + \frac{\Phi_0}{1+\eta\Phi_0\gamma_{0,0}}\left[\frac{c-1}{\delta_{0,0}} \cdot \gamma_{0,0} - 1\right] \tag{A.22}$$

$$\ln \tilde{w}_0(x_2) = \Phi_0 x_2 + \frac{1}{\eta}\ln\left[\frac{\delta_{1,0} - (x_2+\delta_{0,0})^c + c(x_2+\delta_{0,0})^{c-1}(x_2+\gamma_{0,0})}{1+\eta\Phi_0(x_2+\gamma_{0,0})}\right] \tag{A.23}$$

$$\ln \tilde{w}_0(x_3) = \Phi_0 x_3 + \frac{1}{\eta}\ln\left[\frac{\delta_{1,0} - (x_3+\delta_{0,0})^c + c(x_3+\delta_{0,0})^{c-1}(x_3+\gamma_{0,0})}{1+\eta\Phi_0(x_3+\gamma_{0,0})}\right] \tag{A.24}$$

$$\ln \tilde{w}_0(x_4) = \Phi_0 x_4 + \frac{1}{\eta}\ln\left[\frac{\delta_{1,0} - (x_4+\delta_{0,0})^c + c(x_4+\delta_{0,0})^{c-1}(x_4+\gamma_{0,0})}{1+\eta\Phi_0(x_4+\gamma_{0,0})}\right], \tag{A.25}$$

where $x_2 \neq x_3 \neq x_4$ and equation (A.21) follows from the full support condition

$$\delta_{1,t} = (\delta_{0,t})^c \cdot \left(1 + \frac{c}{\eta \cdot \Phi_t \cdot \delta_{0,t}}\right) \Longrightarrow h_t(0, \tilde{a}_t^*(0)) = \left(\frac{c\,(\delta_{0,t})^{c-1}}{\eta \cdot \Phi_t}\right)^{\frac{1}{\eta}},$$

while equations (A.23) - (A.25) are the reduced-form wage equations, where

$$\ln \tilde{w}_t(x_j) = \ln w_{j,t} - \beta_{t,0} - \beta_{t,1}x_j.$$

Identification of $\{\gamma_{0,0}, \delta_{0,0}\}$ from $\{c, \eta, \Phi_0\}$. We first show that $\gamma_{0,0}$ and $\delta_{0,0}$ can be expressed as functions of $\{c, \eta, \Phi_0\}$. From equation (A.21),

$$\delta_{0,0} = \exp\left(\frac{1}{c-1} \cdot (\eta \cdot \log \tilde{w}_0(0) - \log c + \log \eta + \log \Phi_0)\right) \equiv F_0(c, \eta, \Phi_0). \tag{A.26}$$

From equation (A.22),

$$\gamma_{0,0} = \frac{\frac{\tilde{w}_0'(0)}{\tilde{w}_0(0)}}{\left(\frac{c-1}{\delta_{0,0}}\right)\Phi_0 - \left(\frac{\tilde{w}_0'(0)}{\tilde{w}_0(0)} - \Phi_0\right) \cdot \eta\Phi_0} \equiv G_0(\delta_{0,0}, c, \eta, \Phi_0), \tag{A.27}$$

which implies

$$\gamma_{0,0} = G_0(F(c, \eta, \Phi_0), c, \eta, \Phi_0) \equiv \tilde{G}_0(\delta_{0,0}, c, \eta, \Phi_0). \tag{A.28}$$

So, identification of $\{\gamma_{0,0}, \delta_{0,0}\}$ follows from identification of $\{c, \eta, \Phi_0\}$

Identification of $\{\gamma_{0,t}, \delta_{0,t}\}$ from $\{c, \eta, \Phi_0\}$. Jumping ahead, if $\{c, \eta, \Phi_0\}$ are identified we can analogously identify all $\{\gamma_{0,t}, \delta_{0,t}\}_{t>0}$ if, for each $t > 0$, we observe the wage for $x = 0$ and some

$x = \hat{x}_t \neq 0$. In this case we have the following two equilibrium conditions for each t,

$$\ln \tilde{w}_t(0) = \frac{1}{\eta} \cdot (\ln c - \ln \eta - \ln \Phi_t + (c-1)\ln \delta_{0,t}) \tag{A.29}$$

$$\ln \tilde{w}_t(\hat{x}_t) = \Phi_0 \cdot \hat{x}_t + \frac{1}{\eta} \ln \left[\frac{\delta_{1,t} - (\hat{x}_t + \delta_{0,t})^c + c(\hat{x}_t + \delta_{0,t})^{c-1}(\hat{x}_t + \gamma_{0,t})}{1 + \eta(\Phi_0 + \beta_{t,1})(\hat{x}_t + \gamma_{0,t})} \right]. \tag{A.30}$$

Equations (A.21) and (A.29) imply

$$\delta_{0,t} = \exp\left(\frac{1}{1-c} \cdot (\eta \cdot \log \tilde{w}_t(0) - \log c + \log \eta + \log \Phi_t) \right) \equiv F_t(c, \eta, \Phi_0), \tag{A.31}$$

while equation (A.30) implies

$$\begin{aligned}
\underbrace{\exp\left(\eta(\ln \tilde{w}_t(\hat{x}_t) - \Phi_0 \cdot \hat{x}_t)\right)}_{\theta_{0,t}} &= \frac{\delta_{1,t} - (\hat{x}_t + \delta_{0,t})^c + c(\hat{x}_t + \delta_{0,t})^{c-1}(\hat{x}_t + \gamma_{0,t})}{1 + \eta(\Phi_0 + \beta_{t,1})(\hat{x}_t + \gamma_{0,t})} \\
\Longrightarrow \theta_{0,t}\left(1 + \eta(\Phi_0 + \beta_{t,1})(\hat{x}_t + \gamma_{0,t})\right) &= \underbrace{\delta_{1,t} - (\hat{x}_t + \delta_{0,t})^c + c(\hat{x}_t + \delta_{0,t})^{c-1}\hat{x}_t}_{\theta_{1,t}} + ... \\
&\quad ...\ c(\hat{x}_t + \delta_{0,t})^{c-1}\gamma_{0,t} \\
\Longrightarrow \theta_{0,t}\left(1 + \eta(\Phi_0 + \beta_{t,1})\hat{x}_t\right) + \eta\theta_{0,t}(\Phi_0 + \beta_{t,1})\gamma_{0,t} &= \theta_{1,t} + c(\hat{x}_t + \delta_{0,t})^{c-1}\gamma_{0,t} \\
\Longrightarrow \theta_{0,t}\left(1 + \eta(\Phi_0 + \beta_{t,1})\hat{x}_t\right) - \theta_{1,t} &= \underbrace{\left(c(\hat{x}_t + \delta_{0,t})^{c-1} - \eta\theta_{0,t}(\Phi_0 + \beta_{t,1})\right)}_{\theta_{2,t}} \gamma_{0,t} \\
\Longrightarrow \gamma_{0,t} = \frac{\theta_{0,t}\left(1 + \eta(\Phi_0 + \beta_{t,1})\hat{x}_t\right) - \theta_{1,t}}{\theta_{2,t}} &\equiv G_t(\delta_{0,t}, c, \eta, \Phi_0), \qquad \text{(A.32)}
\end{aligned}$$

which in turn implies

$$\gamma_{0,t} = G_t(F_t(c, \eta, \Phi_0), c, \eta, \Phi_0) \equiv \tilde{G}_t(c, \eta, \Phi_0). \tag{A.33}$$

It therefore remains to show identification of $\{c, \eta, \Phi_0\}$.

Identification of $\{c, \eta, \Phi_0\}$, existence. Combining equations (A.23 - A.25) with (A.26) and (A.28) gives the system

$$\ln \tilde{w}_0(x_j) = \Gamma_j(c, \eta, \Phi_0) \qquad j \in \{2, 3, 4\}. \tag{A.34}$$

We are going to show that system has a solution (c^*, η^*, Φ_0^*) in the interior of the parameter space

$$\Theta = \{(c, \eta, \Phi_0) : c \in [2, \bar{c}], \eta \in [\underline{\eta}, 1], \Phi_0 \in [\underline{\Phi}, \bar{\Phi}]\},$$

where $\bar{c}$ and $\bar{\Phi}$ are some arbitrarily large values such that $\bar{c}, \bar{\Phi} < \infty$. We reformulate as a mapping $\Psi(c, \eta, \Phi_0) : \Theta \to \mathbb{R}^3$

$$\Psi(c, \eta, \Phi_0) = \begin{pmatrix} \Gamma_2(c, \eta, \Phi_0) - \ln \tilde{w}_0(x_2) \\ \Gamma_3(c, \eta, \Phi_0) - \ln \tilde{w}_0(x_3) \\ \Gamma_4(c, \eta, \Phi_0) - \ln \tilde{w}_0(x_4) \end{pmatrix}.$$

In order to show existence we construct a function whose fixed points coincide with solutions to

$\Psi(c,\eta,\Phi_0) = 0$ and show that the conditions for Brouwer's Fixed Point Theorem are satisfied. Importantly, in doing so we don't show existence over Θ itself, but rather over some interior subspace of Θ, arbitrarily close to the boundry. Define the ϵ-interior of Θ,

$$\Theta_\epsilon = \{(c,\eta,\Phi_0) \in \Theta : \delta(c,\eta,\Phi_0) \geq \epsilon\}, \tag{A.35}$$

where the distance to boundary function is

$$\delta(c,\eta,\Phi_0) = \min\{c-2, \bar{c}-c, \eta-\underline{\eta}, 1-\eta, \Phi_0-\underline{\Phi}, \bar{\Phi}-\Phi_0\}.$$

Θ_ϵ is clearly a non-empty, compact, and convex set. Moreover $\Theta_\epsilon \subseteq \Theta$ and $\Theta_\epsilon \subseteq \text{int}(\Theta)$. Continuity of $\Psi(\cdot)$ over Θ_ϵ implies that $\Psi(\cdot)$ is bounded on Θ_ϵ. That is,

$$\exists K < \infty \quad st. \quad \sup_{x\in\Theta_\epsilon} \|\Psi(x)\| \leq K.$$

Now define $f : \Theta_\epsilon \to \Theta_\epsilon$:

$$f(c,\eta,\Phi_0) = P_{\Theta_\epsilon}\left((c,\eta,\Phi_0)' - \alpha \cdot \Psi(c,\eta,\Phi_0)\right),$$

where P_{Θ_ϵ} is the projection operator onto Θ_ϵ,

$$P_\Theta(c,\eta,\Phi_0) = \begin{pmatrix} \max\{2+\epsilon, \min\{c, \bar{c}-\epsilon\}\} \\ \max\{\underline{\eta}+\epsilon, \min\{\eta, 1-\epsilon\}\} \\ \max\{\underline{\Phi}+\epsilon, \min\{\Phi_0, \bar{\Phi}-\epsilon\}\} \end{pmatrix}$$

and $\alpha = \epsilon/(2\cdot K)$. The role of $P_\Theta(c,\eta,\Phi_0)$ is to "force" the output of $(c,\eta,\Phi_0)' - \alpha\cdot\Psi(c,\eta,\Phi_0)$ onto the space Θ_ϵ, thereby ensuring that f maps Θ_ϵ onto itself. Note that our parameterization implies

$$\|\alpha_0 \cdot \Psi(x)\| \leq \alpha_0 \cdot \|\Psi(x)\| \leq \frac{\epsilon}{2K}\cdot K < \epsilon.$$

The elements of $\Psi(c,\eta,\Phi_0)$ correspond to

$$\ln\left(h_0\left(x_j, \tilde{a}_0^*(x_j)\right)\right) - \Phi_0 \cdot x_j,$$

which are continuous functions provided $h_0\left(x_j, \tilde{a}_0^*(x_j)\right) > 0$, as was shown to be true in the proof to proposition 3 (equations A.14 - A.17). Since the projection operator P_Θ is clearly continuous on Θ_ϵ, it follows that f is a continuous function from a non-empty, compact, and convex set, Θ_ϵ, to itself. Brouwer's Fixed Point Theorem thereby implies the existence of a fixed point for f in Θ_ϵ.

It only remains to be shown that a fixed point for f must correspond to a solution to $\Psi(\cdot)$. This can be shown by demonstrating that at any fixed point (c^*,η^*,Φ_0^*) the projection P_{Θ_ϵ} is not binding. To see this, let

$$y^* = (c^*,\eta^*,\Phi_0^*)' - \alpha\cdot\Psi(c^*,\eta^*,\Phi_0^*).$$

We then have

$$\|y^* - (c^*,\eta^*,\Phi_0^*)'\| = \|\alpha\cdot\Psi(c^*,\eta^*,\Phi_0^*)\| \leq \alpha\cdot K = \frac{\epsilon}{2K}\cdot K = \frac{\epsilon}{2}.$$

Since $(c^*,\eta^*,\Phi_0^*)' \in \Theta_\epsilon$, we have $\delta(c^*,\eta^*,\Phi_0^*) \geq \epsilon$, meaning it must be the case that $y^* \in \Theta_\epsilon$.

Therefore:

$$
\begin{aligned}
&P_{\Theta_\epsilon}(y^*) = y^* \\
&\implies x^* = P_{\Theta_\epsilon}(y^*) = y^* = x^* - \alpha_0 \cdot \Psi(x^*) \\
&\implies \alpha_0 \cdot \Psi(x^*) = 0 \\
&\implies \Psi(x^*) = 0 \quad (\text{since } \alpha > 0).
\end{aligned}
$$

This establishes the existence of a solution over Θ_e, for any ϵ arbitrarily close to zero.

Local Uniqueness We now turn to proving local uniqueness of a solution given by the above. Formally, we show that for almost all observed data (in the Lebesgue measure sense), any solution to the identification problem is locally unique. In particular, we can show that the set of realized values for the system of wage equations whose corresponding Jacobian is singular is measure zero (whether they correspond to solutions to $\Psi(\cdot) = 0$ or not). It follows that the set of points where the Jacobian is singular *and* which are solutions to $\Psi(\cdot) = 0$ also have measure zero. As a result, for any solution the conditions of the Implicit Function Theorem, and by extension local uniqueness, hold almost surely.

To show this explicitly, define the mapping $\Gamma(c, \eta, \Phi_0) : \Theta_\epsilon \to \mathbb{R}^3$ as follows,

$$
\Gamma(c, \eta, \Phi_0) = \begin{pmatrix} \Gamma_2(c, \eta, \Phi_0) \\ \Gamma_3(c, \eta, \Phi_0) \\ \Gamma_4(c, \eta, \Phi_0) \end{pmatrix}.
$$

Let $\mathcal{C}$ denote the set of critical points of $\Gamma(c, \eta, \Phi_0)$, that is

$$
\mathcal{C} = \{\theta \in \Theta_\epsilon : \det(J_{\Gamma(\theta)}) = 0\},
$$

where $J_{\Gamma(\theta)}$ is the Jacobian of $\Gamma(c, \eta, \Phi_0)$. Since $\Gamma \in C^1$ (that is, $\Gamma(c, \eta, \Phi_0)$ is continuously differentiable across Θ_ϵ), Sard's theorem immediately implies that the set of critical values $\Gamma(\mathcal{C})$ has Lebesgue measure zero in $\mathbb{R}^3$.

Now, let $D^* = (\ln \tilde{w}_0(x_2), \ln \tilde{w}_0(x_3), \ln \tilde{w}_0(x_4))$ denote observed data. Our identification problem amounted to finding some $\theta^* = (c^*, \eta^*, \Phi_0^*)$ such that $\Gamma(\theta^*) = D^*$ (*ie.* $\Psi(\theta^*) = 0$). From Sard's theorem,

$$
\mathcal{L}(\{\text{data } D : \exists \theta \text{ with } \Gamma(\theta) = D \text{ and } \det(J_{\Gamma(\theta)}) = 0\}) = 0,
$$

where $\mathcal{L}$ denotes the Lebesgue measure. Put another way, for almost all data $D^* \in \mathbb{R}^3$, if there exists a solution θ^* with $\Gamma(\theta^*) = D^*$, then:

$$
\det(J_{\Gamma(\theta)}) = \det(J_{\Psi(\theta^*)}) \neq 0.
$$

It therefore follows from the Implicit Function Theorem that for almost all data (in a Lebesgue measure sense), any such solution is locally unique.

B.5 Proof of Proposition 5

The proof is constructive. Occupational wages for the continuous approximation are written $w_j = M(\tilde{a}^*(x_j))$. Let $w(x_j, \tilde{a})$ be wages as a function of arbitrary values of task complexity and worker comparative advantage, which is also the workers' objective function. As shown in the text, under assumptions 3 and 4 this function is log-supermodular. Occupational choices in the discrete case therefore satisfy the Interval Partition Rule, implying that the "true" occupational wages given optimal choices are $E\left[w\left(x_{j+1},\tilde{a}\right)|\tilde{a}\in(\tilde{a}_j,\tilde{a}_{j+1}]\right]$. Let $\hat{w}_j$ be the wage observed in the data for occupation j. The model implies that $\hat{w}_j = E\left[w\left(x_{j+1},\tilde{a}\right)|\tilde{a}\in(\tilde{a}_j,\tilde{a}_{j+1}]\right]$. On the other hand, the continuous approximation imposes the equality $\hat{w}_j = w_j$. The approximation error is therefore $\varepsilon(x_j) = |w_j - \hat{w}_j| = |M\left(\tilde{a}^*(x_j)\right) - E\left[w\left(x_{j+1},\tilde{a}\right)|\tilde{a}\in(\tilde{a}_j,\tilde{a}_{j+1}]\right]|$, as defined in the main text. To characterize this error term, we consider a limit process along which the initial occupational classification and its corresponding complexity scores $\{x_j : j = 1, ..., J\}$ are fixed at the values observed in the data, while more and more occupational categories are added as $J \to \infty$. In the limit, x takes all values in $[0,1]$, but the occupations indexed by j retain their original complexity scores. Occupations that are added as $J \to \infty$ will get a different index, as described below.

We first derive the distribution-free and computable bound on $\varepsilon(x_j)$ for an arbitrary occupation j in $\{1, ..., J\}$. The Interval Partition Rule for occupational choices implies that workers with $\tilde{a}\in(\tilde{a}_{j-2},\tilde{a}_{j-1}]$ choose occupation $j-1$, and workers with $\tilde{a}\in(\tilde{a}_{j-1},\tilde{a}_j]$ choose occupation j. Since $w(x_j,\tilde{a}) > 0$, its log-supermodularity implies supermodularity. Hence, for $\tilde{a}^1 \in (\tilde{a}_{j-2},\tilde{a}_{j-1}]$ and $\tilde{a}^2 \in (\tilde{a}_{j-1},\tilde{a}_j]$ we have the following inequalities:

$$w(x_{j-1},\tilde{a}^1) \le w(x_{j-1},\tilde{a}_j) < w(x_j,\tilde{a}_j) \le w(x_j,\tilde{a}^2). \tag{A.36}$$

Integration delivers the empirical lower bound $\hat{w}_{j-1} = E\left[w\left(x_{j-1},\tilde{a}\right)|\tilde{a}\in(\tilde{a}_{j-2},\tilde{a}_{j-1}]\right]$. The upper bound $\hat{w}_{j+1} = E\left[w\left(x_{j+1},\tilde{a}\right)|\tilde{a}\in(\tilde{a}_j,\tilde{a}_{j+1}]\right])$ can be derived using the same arguments, but with $j+1$ replacing j. These bounds are empirical and distribution-free because they are the "true" occupational wages and are directly observed in the data.

Next we study the convergence properties of the "true" occupational wages $\hat{w}_j$ to w_j as J grows, holding fixed the original grid indexed by j. We start with noting that $\tilde{a}^*(x_j) \in (\tilde{a}_{j-1},\tilde{a}_j]$, as discussed in the text. Inequalities (A.37) imply that

$$M\left(\tilde{a}^*(x_{j-1})\right) = w_{j-1} < E\left[w\left(x_j,\tilde{a}\right)|\tilde{a}\in(\tilde{a}_{j-1},\tilde{a}_j]\right] < w_{j+1} = M\left(\tilde{a}^*(x_{j+1})\right). \tag{A.37}$$

The complexity score $x_j \in [0,1]$ is the limit of some Cauchy-sequence of rational numbers in $[0,1]$ because of Cantor's construction of reals. Hence, the process of constructing the limit continuous occupational classification creates subsequences $\{x^-_{j,k}\}_{k=1}^{\infty}$ and $\{x^+_{j,k}\}_{k=1}^{\infty}$ of additional occupations indexed by k, with $k \to \infty$ as $J \to \infty$, that converge to x_j from below and above, respectively, because the continuous case of the space of task complexity $[0,1]$ itself is the completion of its countable analogue we are starting from. (Alternatively, one can view the discrete classification observed in the data as the result of removing these occupations (and uncountably many more).). Corresponding to these sequences are wages and cutoff skill levels $w^-_{j,k-1} < w^-_{j,k} < w^+_{j,k} < w^+_{j,k-1}$ and $\tilde{a}^-_{j,k-1} < \tilde{a}^-_{j,k} < \tilde{a}^+_{j,k} < \tilde{a}^+_{j,k-1}$ such that $w^-_{j,k} < E\left[w\left(x_j,\tilde{a}\right)|\tilde{a}\in(\tilde{a}^-_{j,k},\tilde{a}^+_{j,k}]\right] < w^+_{j,k}$. This follows from using the same arguments as those preceding inequalities (A.37), but with $(x^-_{j,k}, x^+_{j,k})$ taking on the roles of (x_{j-1}, x_{j+1}). Given assumption 5, the wage function in equation (7) is continuous and bounded on $x \in [0,1]$. Hence, the bounds $(w^-_{j,k}, w^+_{j,k})$ converge to w_j from below and above. With the true occupational wage $E\left[w\left(x_j,\tilde{a}\right)|\tilde{a}\in(\tilde{a}^-_{j,k},\tilde{a}^+_{j,k}]\right]$ "sandwiched" between the two bounds for any k, it must also converge to w_j. This completes the proof of the Proposition.

B.6 Proof of Proposition 7

We prove this proposition by backward induction on j. The main text provides the initialization for $j = J$. Now consider any $j \in \{1, 2, ..., J-1\}$. Using equation (A.38) together with $g_{m,t} =$

$\left[\lambda_{m,t} + \left(\frac{k_t^e}{l_{m,t}}\right)^{\sigma_m} \cdot H_m(\omega_{m,t})\right]^{\frac{\sigma_m - \sigma_{m-1}}{\sigma_m}}$ yields:

$$MRTS(l_{j,t}, k_t^e) \;=\; \lambda_{j,t} \cdot \left(\frac{k_t^e}{l_{j,t}}\right)^{1-\sigma_j} \cdot \left(\frac{k_t^e}{l_{j+1,t}}\right)^{\sigma_j - \sigma_{j+1}} \cdot g_{j+1,t} \cdot G_{j+1}(\omega_{j+1,t}) \qquad \text{(A.38)}$$

and thus

$$\ln\left(\frac{\overline{w}_{j,t}}{r_t}\right) - \ln G_{j+1}(\omega_{j+1,t}) = \widetilde{\lambda}_j + \theta_j \cdot t + (1-\sigma_j) \cdot \ln\left(\frac{k_t}{l_{j,t}}\right) + \beta_1^\sigma \cdot \ln\left(\frac{k_t^e}{l_{j+1,t}}\right) + \beta_2^\sigma \cdot \ln\left(\tilde{g}_{j+1,t}\right), \qquad \text{(A.39)}$$

where $\tilde{g}_{j+1,t} \equiv \left[\lambda_{j+1,t} + \left(\frac{k_t^e}{l_{j+1,t}}\right)^{\sigma_{j+1}} \cdot H_{j+1}(\omega_{j+1,t})\right]$, $\beta_1^\sigma \equiv (\sigma_j - \sigma_{j+1})$ and $\beta_2^\sigma \equiv \left(\frac{\sigma_{j+1}-\sigma_j}{\sigma_{j+1}}\right)$. The induction hypothesis implies that all objects in $\tilde{g}_{j+1,t}$ and in $G_{j+1}(\omega_{j+1,t})$ can be taken as data. As a consequence, this equation is a linear regression of $\left(\ln\left(\frac{\overline{w}_{j,t}}{r_t}\right) - \ln G_{j+1}(\omega_{j+1,t})\right)$ on occupation fixed effects, a linear time trend, $\ln\left(\frac{k_t}{l_{j,t}}\right)$, $\ln\left(\frac{k_t^e}{l_{j+1,t}}\right)$, and $\ln\left(\tilde{g}_{j+1,t}\right)$, a highly non-linear function of all capital-labor ratios and structural parameters for occupations $m > j$. Except from the case in which all labor inputs evolve at an identical log-linear trend, the regressors are not perfectly collinear, and the coefficients $(\theta_j,\ (1-\sigma_j),\ \beta_1^\sigma,\ \beta_2^\sigma)$ together with the occupation fixed effects $\widetilde{\lambda}_j$ are identified. As a consequence, not only is σ_j idenfied, but so are all parameters that will be used as data in the next induction step. This finishes the proof.

C Data Appendix & Additional Computational Details

C.1 Top Coded Earnings in the Census, ACS and ASEC

Top-coding in our earnings data is potentially problematic because how top-coded observations are treated in the samples we rely on changes over time. In particular, since 2000 the Census and the ACS report average states earnings above the top code for top-coded earnings observations, but for 1990 it reports the corresponding median and for 1980 it does not use any imputation and simply reports the top-code. We make wave earnings data consistent across our samples by replacing Census-imputed values in 1980 and 1990 with an imputed value that mimics the Census approach since 2000. To this end we simulate earnings observations above the top-code by drawing from a 2-parameter Weibull distribution and use the average over these draws as imputed values. The parameters of the distribution are estimated by matching distributional characteristics near the top-code, which are allowed to vary over time and across U.S. states. We use this procedure for imputation because it can be validated on Census/ACS data after 2000 by comparing the true imputed value with our simulated estimates. We find that it replicates the reported averages above the top-code near-perfectly, while a more common imputation based on a one-parameter Pareto-distribution, as used for example in Kopczuk *et al.* (2010) on Social Security Microdata, tends to underestimate these averages.

C.2 User cost of capital

In order to estimate returns on capital we calculate the user cost of capital equipment and capital structures, largely following the method in Hubmer (2023). In particular, the individuals's portfolio choice problem requires a non-arbitrage between holding capital of type $k = \{e, s\}$ and bonds, which implies the user cost formula

$$r_t^k = q_t^k \cdot \left(1 + \bar{r}_t - (1 - d_t^k)\frac{\mathbb{E}\left[q_{t+1}^k\right]}{q_t^k}\right) \tag{A.40}$$

Where r_t^k is the required return on a single *unit* of capital of type k

- $\bar{r}_t$ is the real interest rate, taken to be a weighted average of the return on debt (the 10-year US Treasury yield plus a 5 percent risk premium) and the return on equity (the Moody's AAA corporate bond yield), minus inflation. The weight on debt is the sum of debt securities and loans, taken from Table S.5 of the Integrated Macroeconomic Accounts, while the weight on equity is the equity and investment fund shares, also from Table S.5 of the Integrated Macroeconomic Accounts. Inflation is a five-year moving average of the CPI.
- d_t^k is depreciation, constructed as the current cost of depreciation divided by the current cost net stock of capital (both series taken from the BEA's fixed asset tables)
- q_t^k is the quality-adjusted price of capital of type k, relative to the price of consumption
  - q_t^e is the capital equipment deflator, taken as the FRED series "EDEF", divided by the consumption deflator (taken to be the FRED series "CONSDEF")
  - q_t^s is the BEA's nonresidential structures deflator (series A009RD3Q086SBEA), divided by the consumption deflator
- As in Hubmer (2023), the expected price growth term is approximated by calculating a five-year moving average of realized price growth.

C.3 Counterfactual Procedure

We now sketch out our method for recovering skill prices, wages and labor supply under counterfactual parameterizations of the model.

Given a counterfactual parameterization $\{\eta^*, \sigma_j^*, \lambda_{j,t}^*, r_t^{e*}, \kappa_t^*, \rho_t^*, \delta_t^*(\cdot), \gamma_t^*(\cdot)\}_{j=1}^J$:

1. Guess the parameters $\beta_t^0 = \{\beta_{t,0}^0, \beta_{t,1}^0\}$ of the skill price function. We restrict the functional form of the counterfactual skill price function to be consistent with our estimated equation (16)

$$\begin{aligned} log(w_{t,x}^*) &= \beta_{t,0}^0 + \beta_{t,1}^0 \cdot x, \\ \Phi_t^*(x) &= \beta_{t,1}^0. \end{aligned}$$

2. Compute the counterfactual inverse policy function, per-worker efficiency units, cumulative

employment shares, and total occupation efficiency units

$$
\begin{aligned}
\tilde{a}_t^{**}(x) &= \left[\frac{-\left(\delta_t^{*'}(x)+\eta^*\cdot\Phi_t^*(x)\cdot\delta_t^*(x)\right)}{\gamma_t^{*'}(x)+\eta^*\cdot\Phi_t^*(x)\cdot\gamma_t^*(x)}\right]^{\frac{1}{\eta^*}}, \\
h_t^*\left(x,\tilde{a}_t^{**}(x)\right) &= \left[\frac{\left(\delta_t^*(x)\cdot\gamma_t^{*'}(x)-\delta_t^{*'}(x)\cdot\gamma_t^*(x)\right)}{\gamma_t^{*'}(x)+\eta^*\cdot\Phi_t^*(x)\cdot\gamma_t^*(x)}\right]^{\frac{1}{\eta^*}}, \\
\tilde{G}_t^*(\tilde{a}) &= 1-e^{-\left(\frac{\tilde{a}}{\rho_t^*}\right)^{\kappa_t^*}}, \\
l_{j,t}^* &= h_t^*\left(x_j,\tilde{a}_t^{**}(x_j)\right)\cdot\left(\tilde{G}_t^*(\tilde{a}_t^{**}(x_j))-\tilde{G}_t^*(\tilde{a}_t^{**}(x_{j-1}))\right).
\end{aligned}
$$

3. Compute counterfactual capital demand

$$
k_t^{e*}=l_{J,t}^*\cdot\left(\frac{w_{J,t}^*}{\lambda_{J,t}^* r_t^{e*}}\right)^{\frac{1}{1-\sigma_J^*}}
$$

4. Compute counterfactual MRTS

$$
\begin{aligned}
MRTS^*(l_{j,t}^*,k_t^{e*}) &= \lambda_{j,t}^*\cdot\left(\frac{k_t^{e*}}{l_{j,t}^*}\right)^{1-\sigma_j^*}\cdot\left(k_t^{e*}\right)^{\sigma_j^*-\sigma_J^*}\cdot\prod_{m=j+1}^{J}\left(\widetilde{l}_{m,t}^*\right)^{\frac{\sigma_m^*-\sigma_{m-1}^*}{\sigma_m^*}} \\
&\quad\text{with} \\
\widetilde{l}_{j,t}^* &= \begin{cases}\lambda_{j,t}^*\cdot\left(l_{j,t}^*\right)^{\sigma_j^*}+\left(\widetilde{l}_{j+1,t}^*\right)^{\frac{\sigma_j^*}{\sigma_{j+1}^*}} & \text{for } j\in\{2,...,J-1\}\\ \lambda_{j,t}^*\cdot\left(l_{j,t}^*\right)^{\sigma_j^*}+\lambda_{k,t}^*\cdot\left(k_t^{e*}\right)^{\sigma_j^*} & \text{for } j=J.\end{cases}
\end{aligned}
$$

5. Loop over β_t^0 to match

$$
\left\|\frac{w_{j,t}^*}{r_t^{e*}}-MRTS^*(l_{j,t}^*,k_t^{e*})\right\|.
$$