Technology Adoption and the Slowdown in Skilled Labor Demand

Aniket Baksy *

November 25, 2022

Updated Daily, latest version here.

Abstract

Between 1980 and 2000, growth in the skill premium and a decline in the relative price of capital led economists to conclude that capital-embodied technical change was driving up the relative demand for skilled labor. Given the continued steady decline in capital prices post 2000, these models predict a continual rise in the skill premium. However, post 2000, growth in the skill premium has slowed down. I argue that as the skill premium increased, firms adopted new technologies economizing on the use of skilled labor. I quantify this force using an equilibrium model with costly technology adoption. As capital prices fall, capital-skill complementarity initially drives up the skill premium. Firms respond by investing in new technologies which are less skilled-labor-intensive. The model successfully accounts for the slowing skill premium and the behavior of the labor share. Without technology adoption, the model predicts a skill premium in 2019 that is 5 percentage points higher and a labor share that is almost 12 percentage points higher. I provide microeconomic evidence for my mechanism by showing that accountants relatively more exposed to the adoption of accounting software saw slower wage growth.

*I am extremely grateful to Patrick Kehoe, Nicholas Bloom, Luigi Bocola and Elena Pastorino for their extensive guidance, assistance in obtaining data, generosity with their time and invaluable comments. I also thank Adrien Auclert for his help throughout the process of writing this paper. I thank Mohamad Adhami, Ross Batzer, Rafael Berriel, Daniele Caratelli, Nick Cao, Calvin He, Brian Higgins, Pete Klenow, Ernest Liu, Cedomir Malghieri, Frédéric Martenet, Brendan Moore, Monika Piazzesi, Martin Schneider, Martin Souchier, and participants at the Bloom Reading Group, the Stanford Macroeconomics Seminar and Lunch for comments. I am grateful for generous support from the Patricia Liu McKenna and Kenneth McKenna Graduate Fellowship and the E.S. Shaw and B.F. Haley Fellowship for Economics through grants to the Stanford Institute for Economic Policy Research.

1 Introduction

Rising labor income inequality in the United States is a matter of lively academic and policy discussion. A key dimension of this rise in income inequality is the increase in the relative hourly earnings of skilled workers to unskilled ones, the *skill premium*. Between 1980 and 2000, the skill premium rose about 19 percentage points, explaining up to two-thirds of the increase in the dispersion of labor earnings. Over this same period, the relative price of capital goods to consumption goods fell steadily. The combination of these trends has led economists to develop models in which ongoing technical change, as measured by this declining relative price of capital goods, has raised the relative demand for skilled workers and steadily eroded the relative wages of unskilled workers (Krusell et al. (2000), Acemoglu and Restrepo (2020b)). As the prices of capital goods continue to fall, such models predict an ever growing skill premium and therefore a continuous rise in inequality[1].

Interestingly, even though in the twenty years since 2000 the rate of decline in the relative price of capital–the fundamental force behind the growing skill premium in traditional models of capital-skill displacement–has continued unabated, the actual skill premium has grown by much less. Indeed, rather than rising almost 20% as it did in the twenty years from 1980 to 2000, this premium instead rose less than 6% for the nineteen years from 2000 to 2019. In this paper, I argue that when the price of one type of labor, say skilled labor, gets much more expensive than another, say unskilled labor, there are natural competitive forces that lead firms to direct their efforts at developing and adopting new technologies that economize on the very expensive type of labor. This idea builds on the insight of Hicks (1932) that relative prices govern the direction of technology adoption by firms. I argue that this is exactly what happened in the post-2000 period. In a nutshell, the idea is that as the relative price of workers such as accountants, lawyers, and financial analysts continues to rise between 1980 and 2000, firms begin to find it worthwhile to invest in technologies, such as new computer software, that can accomplish many of the tasks performed by these workers. This directed technical change decreased the relative demand for such high-skilled workers and greatly slackened the rate of growth of their relative wages.

The idea that firms will respond to a rising skill premium by directing their technology choices toward reducing their reliance on skilled workers is closely related to the idea of labor displacement via automation of the tasks they perform. The existing literature on automation largely documents the adverse impacts of specific automation technologies such as industrial robots on low skilled workers[2]. While the possibility of displacement of lower skilled workers by physical capital-embodied technology has thus been well established, I argue that the capabilities of modern

[1] The large increases in the skill premium driven by rising automation and capital-labor substitution has led to some backlash, with technological progress garnering a lion's share of blame for recent rises in inequality. This has led some to call for policy responses to endogenously direct technological change in certain directions. For instance, see Qureshi (2020), Korinek et al. (2021) and Lohr (2022).

[2] See, for instance,Acemoglu and Restrepo (2020b), Acemoglu et al. (2020), Acemoglu and Restrepo (2020a), and Hémous and Olsen (2022).

technologies embodied in software capital and ICT products can in fact allow them to displace relatively high skilled workers. For instance, "E-discovery" software programs used in the legal domain use advanced language-based inference techniques to identify wrongdoing and white-collar crime, dramatically reducing the number of trained lawyers required in document review[3]. Software like SAP/R3, TurboTax and Oracle's NetSuite have dramatically reduced business' reliance on accountants for resource planning, invoicing and tax preparation. Solutions by companies like Synopsys have displaced the work of computer chip designers by substantially automating skilled-labor-intensive tasks like circuit design. My paper allows for both channels of displacement, so that firms can use capital in the form of machinery to displace both low-skilled workers - as in the existing literature - and also use capital in the form of software technologies to displace high skilled workers.

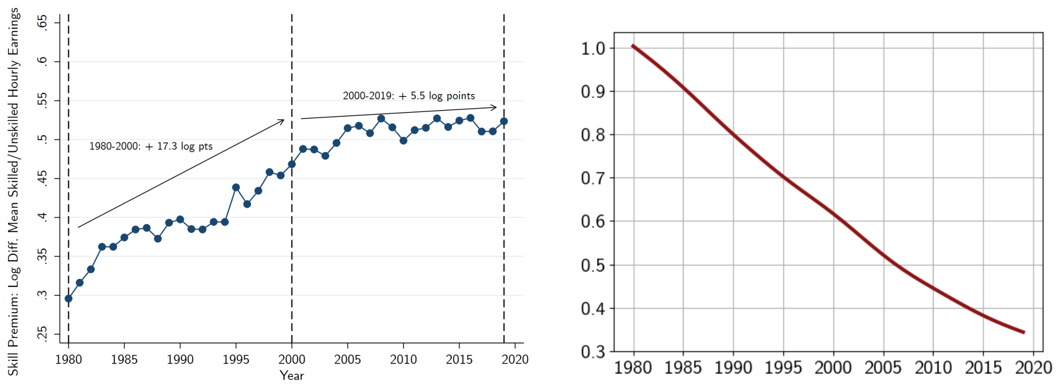


Figure 1: LEFT: Rise and Slowdown in the skill premium, defined as the log difference in mean skilled wages to unskilled wages. The sample is civilian male full-time-full-year employed workers aged 18-65. Wages are first residualized on race, experience and region. Group mean wages are computed as the weighted average of wages using labor supply weights as in Autor (2019) for workers in each of five education bins. Skilled wages are the labor-supply-weighted mean wages of groups with college degrees or post-college education plus half of workers with some college education. RIGHT: Declining price of capital goods relative to consumption goods from DiCecio (2009), normalized to one in 1980.

This paper has two parts. First, I propose a simple model of dynamic technology-skill substitution in which both unskilled and skilled workers are exposed to displacement by capital. I show that the model can explain the entire time path of both the skill premium and the labor share, and that shutting down the main mechanism I add generates counterfactual behavior for both series, especially the labor share. Second, having shown that the model is consistent with macro data series, I provide microeconomic evidence that growth in the relative demand for skilled labor has slowed down by studying the behavior of wages and employment of one group of skilled workers, accountants. I also show that between 1980 and 2019, there has been a shift in the direction of patenting toward the development of technologies which perform tasks more similar to the tasks that skilled workers perform.

[3]These include software such as Cataphora, created by Cataphora LLP, which mines information about individual actions and communications between individuals to identify patterns of "digital anomalies" associated with wrongdoing. (Markoff, 2011)

I now explain the model I construct in more detail. The model features firm dynamics with monopolistic competition, augmented with costly technology adoption decisions. Production by individual firms requires the completion of tasks specific to skilled labor and tasks specific to unskilled labor. Some tasks of each type can also be produced by capital. Technologies are indexed by the shares of tasks of each type that are capital-feasible. Firms enter each period with a pre-determined technology parametrized by these shares of capital-feasible tasks. Given these shares, capital and skilled labor are more complementary than capital and unskilled labor.

The key innovation I introduce here is the possibility of firms adopting new technologies over time by making upfront technology adoption investments. That is, after production occurs, firms have the opportunity to invest in costly technology adoption, allowing them to enter the next period with a new technology. New technologies adopted by firms involve a weakly larger share of each type of tasks that are capital feasible. The model therefore features two distinct notions of capital-labor substitution: a short run elasticity, conditional on the technology being operated, and a long-run elasticity, which incorporates the possibilities of changing the technology being operated. The former corresponds to the standard notion of movement along an isoquant. The latter corresponds to long-run shifts in factor shares as factor prices themselves change, which involves shifts in the isoquants themselves.

I study the ability of the model to explain the behavior of the skill premium and the labor share of income using a transition experiment driven by both the declining path for the relative price of capital and the increasing path for relative skill supplies. To take the simple mechanism I model to the data, I augment the model in two main dimensions. First, I allow for heterogeneity in firm productivity, with firms facing persistent productivity shocks. Second, I allow for technologies to diffuse via entry and exit. Given the paths for the relative price of capital and the increase in the relative supply of skilled labor, I find that the quantitative model can account for the paths of both the skill premium and the labor share, despite not being calibrated to target either of these series[4].

How does the model accomplish this? Along the initial part of the transition path, capital-skill complementarity drives up the skill premium by driving up the relative demand for skilled labor. As the skill premium rises, firms slowly begin directing their technology adoption choices toward displacing skilled labor performing skilled labor-intensive tasks. As in Hubmer and Restrepo (2021), more productive firms, which are larger, make larger investments in technology adoption, thus enjoying lower costs and higher revenues. The new technologies they adopt mean that they also have lower labor shares. The aggregate labor share decline is thus driven by *reallocation* of value added toward low labor share firms, consistent with the evidence in Kehrig and Vincent (2021). Over time, the adoption of these new less skilled labor-intensive technologies slows growth in skilled labor demand. This manifests as a slowdown in the share of labor income going to skilled labor. As the unskilled labor share continues to decline almost unabated, the net effect on the labor

[4]The fact that the model is not targeted to explain the behavior of either time series distinguishes my exercise from that of Hémous and Olsen (2022), who choose the parameters of their model to target the behavior of both the labor share and the skill premium.

share is an eventual decline.

Having shown that the model is consistent with the macroeconomic facts, I next provide microeconomic evidence for the key mechanism in the model, that growth in the relative demand for skilled labor has slowed down. I do this via two main exercises. First, I perform a case study of one group of particularly skilled workers, accountants, and show that the adoption of accounting software is associated with declines in their relative wages. I use microdata on the adoption of accounting software at the establishment level to construct, at the commuting zone level, a measure of exposure of accountants to the use of accounting software, measured by the share of establishments within that commuting zone which have adopted software by a given date. I find robust evidence that exposure has negative effects on accountants' wage growth. Second, I use frontier methods in natural language processing, based on Webb (2020), to show that over time, newly created technologies are more likely to be associated with the tasks performed by relatively skilled workers. Rising exposure measured by the similarity of tasks performed by newly created technologies and tasks performed by given classes of workers robustly predicts a decline in wage growth.

This paper contributes to the following strands of literature.

First, I contribute to a literature on how technological change affects skill demands and inequality by proposing a model which simultaneously rationalizes the initial increase in skilled labor demand between 1980 and 2000 with the subsequent slowdown via endogenous technology adoption. A long literature including Katz and Murphy (1992), Bound and Johnson (1992), Acemoglu (1998), Autor et al. (1998), Goldin and Katz (1998), Katz and Autor (1999), Johnson (1997), Acemoglu (2002) and summarized by Acemoglu and Autor (2011) argues that explaining the joint increase in the skill premium and the relative supply of skilled workers between 1980 and 2000 requires an expansion in relative demand for skilled labor. Models of *skill-biased technical change* (SBTC) explain this increase via exogenously rising relative skilled labor productivity. KORV (2000) provide an underlying story for skill-bias in technology, by proposing that skilled labor is more complementary to capital than unskilled labor and noting that in the presence of investment-specific technical change (Greenwood et al., 1997), this mechanism can generate an increase in skilled labor demand[5][6]. He and Liu (2008) embed this mechanism in a general equilibrium model with endogenous factor supply and show that it can account for almost all of the increase in the skill premium between 1980 and 2000.

However, estimated models of capital-skill complementarity struggle to simultaneously match the decline in the labor share and the slowdown in the skill premium after 2000 (Maliar et al.

[5]The exercise performed by Krusell et al. (2000) is as follows. They specify a production function for consumption goods and estimate the key parameters governing substitutability between capital and labor. They manipulate the first order conditions for profit maximization by the firm to obtain three equations: one for the wage bill ratio, one for the labor share of value added and one no-arbitrage condition between capital structures and equipment. They estimate the model by a variant of maximum likelihood using data on factor inputs, output and capital prices. Using the fact that wages equal marginal products of labor in their competitive model, they then compute the implied path for the skill premium, finding that it closely replicates the dynamics of the skill premium in the data.

[6]As Violante (2022) argues, in the presence of a declining capital equipment price, capital-skill complementarity acts as a skill-biased demand shifter, and hence can be viewed as a microfoundation for SBTC.

(2020), Ohanian et al. (2022), Castex et al. (2022)). Given fixed elasticities of substitution across the factors of production consistent with the data, the continued decline in the relative price of capital generates persistent increases in skilled labor demand. For a given growth in skilled labor supply, this increase drives up its share of value added and the skill premium. Castex et al. (2022) show that a variant of the original KORV model in which the elasticity of substitution between equipment and skilled labor increases over time improves the model's ability to account for the data, but their model still struggles to match the decline in the labor share over the post-2000 period. A key puzzle for this entire literature, therefore, is simultaneously matching the initial increase and then slowdown of the skill premium while also matching the initial stability and the subsequent decline in the labor share. I contribute to this literature a resolution of this puzzle - costly endogenous technology adoption to displace skilled labor as it becomes more expensive - which rationalizes the entire path of *both* time series.

The slowdown in the skill premium has been documented previously in work including Autor et al. (2008), Mishel et al. (2013), and Valletta (2018). Beaudry et al. (2016) show that since 2000, there was a large slowdown in the growth of occupations that were intensive in cognitive tasks and therefore in the demand for skilled workers, who are overrepresented in such occupations. Castex and Dechter (2014) document a decline in the returns to cognitive skills, and with the declining proportion employed in cognitive and non-routine occupations and slower wage growth for newer cohorts of college graduates documented in Beaudry et al. (2014). To this literature, I contribute a model of technological change that explicitly incorporates a channel for the slowing demand for skilled labor, and evaluate its ability to account for aggregate measures of this slowdown.

Second, I contribute to a literature in macroeconomics and labor economics studying the impacts of modern technologies, including computers, ICT and software, on labor market outcomes by skill group. Early studies of the impacts of these technologies, including Bound and Johnson (1992), Juhn et al. (1993), Berman et al. (1994), Autor et al. (1998), Kaiser (2000), Autor et al. (2002), Bresnahan et al. (2002), Spitz-Oener (2006), Bartel et al. (2007), Akerman et al. (2015), and Atalay et al. (2018), generally found that this cluster of technologies was skill-biased, in that the adoption of these technologies raised the relative demand for skilled workers to unskilled workers[7]. I contribute to this literature a model in which the adoption of these technologies can slow the demand for *skilled* workers, show that this mechanism is quantitatively important to explain aggregate trends, and provide microeconomic evidence for this slowdown in demand. My case study of accountants associates the diffusion of ICT with adverse outcomes for relatively high skilled individuals. This is consistent with the findings of Jiang et al. (2021), who show that occupations most exposed to FinTech adoption saw declines in job postings. The occupations they identify as most exposed -

[7] An influential literature, including the work of Autor et al. (2003), Autor et al. (2006), Goos and Manning (2007), Autor et al. (2008), Goos et al. (2009), Boehm (2013), Goos et al. (2014), Barany and Siegel (2018) and Michaels et al. (2020), argues for a richer characterization of the impacts of computers, instead emphasizing that ICT and related technologies substitute for workers in occupations which are relatively intensive in routine tasks. This literature therefore identifies the impact of computers as most negative for workers in the middle of the skill distribution, who are likely to work in relatively routine occupations. My model will not be able to speak to the literature on polarization because it only admits two skill levels, and a multi-skill extension would be particularly interesting for future research.

credit analysts and information security analysts - are both occupations identified in O*NET as skill intensive. My results are also consistent with the findings of Deming and Noray (2020), who show that the STEM premium enjoyed by graduates majoring in tech-intensive subjects declines dramatically over the first years of their careers due to obsolesence of their skills. My model interprets part of this obsolence as the outcome of the tasks they perform being performed by capital instead. My case study of accountants is also similar to the one conducted in Dillender and Forsythe (2022), but my data allows for the study of the effect of adopting a more precisely defined group of technologies on a more precisely defined group of workers.

Third, I contribute to a rapidly expanding literature in macroeconomics and labor economics studying the effect of automation on labor markets with models of production involving tasks (Zeira (1998), Acemoglu and Autor (2011), Autor (2013), and Acemoglu and Restrepo (2018b)). Papers in this literature studying the impacts of automation on labor income inequality[8] include Autor et al. (2003), Autor et al. (2006), Acemoglu and Autor (2011), Acemoglu and Restrepo (2018a), Acemoglu and Restrepo (2022), and Hémous and Olsen (2022). I contribute to this literature a model in which dynamic technology adoption choices by firms shapes labor market consequences for skilled and unskilled workers, but in which these labor market outcomes feed back into adoption choices. Hubmer and Restrepo (2021) construct a model of costly technology adoption in response to declining capital prices to study changes in the distribution of labor shares across firms. My model's structure is similar to theirs, but distinguishes between skilled and unskilled labor, allowing me to study the skill premium and the labor share jointly. My model shares features with Holmes and Mitchell (2008), who construct a model in which plants perform tasks with capital, skilled and unskilled labor. In their model, firms spend setup costs to be able to displace skilled labor by unskilled and unskilled labor by capital, and motivate these costs by the generality of the underlying tasks and the ability of workers to specialize in them. Unlike their work, my model features dynamic forward-looking decisions to adopt technologies and allows for the idea that capital directly displaces skilled workers at tasks that are specific to skilled workers.

Fourth, my mechanism is similar to the one emphasized in a literature on directed technological change, including the work of Hicks (1932), Habakkuk (1962), Allen (2009), Acemoglu (1998), Acemoglu (2010), Acemoglu and Restrepo (2018a), Aum (2018) and Hémous and Olsen (2022). This literature has largely focused on the impacts of automation on relatively low-skilled workers. Notable exceptions are Acemoglu and Restrepo (2018a), Aum (2018) and Hémous and Olsen (2022). Acemoglu and Restrepo (2018a) features a three-factor production structure, but does not contain a detailed analysis of directed technology adoption, since they focus throughout on marginal changes to technical thresholds around a given allocation. Aum (2018) uses a model in which software capital complements high skilled workers and equipment capital complements medium skilled workers, showing that a decline in software prices can produce polarization in the labor market and

[8] Moll et al. (Forthcoming) study impacts of automation for income and wealth inequality, arguing that automation has an additional impact on inequality by raising capital incomes relative to labor incomes.

a slowdown in the demand for skilled labor[9]. While his mechanism is related to mine, Aum does not study dynamics of either the labor share or skill premia. Hémous and Olsen (2022) construct a model in which firms can respond to rising wages of unskilled workers by engaging in technology adoption to displace them at the tasks they perform with capital. My model generalizes this setup substantially by allowing skilled workers to also be displaced by capital, allowing it to account for the paths of both the skill premium and the labor share despite not targeting either series..

Fifth, I contribute to the literature on the evolution of the labor share of income, reviewed in Grossman and Oberfield (2022). There is extensive debate in the literature on the magnitudes of this decline due to measurement issues involving, among others, the treatment of stock options (Eisfeldt et al. (2021)), the treatment of the labor portion of proprietors' incomes (Gollin (2002), Elsby et al. (2013)), intangible capital and changes in accounting norms surrounding IT capital (Koh et al. (2020)), how to account for depreciation in measuring value added (Weitzman (1976), Hulten (1992)), and the role of housing (Rognlie (2015), Gutiérrez and Piton (2020)). However, there is a general consensus that this share has declined for the aggregate economy, and particularly so in manufacturing and retail, two critical sectors for US employment. A wide variety of explanations have been proposed for the decline of the labor share, including the effects of trade (Elsby et al., 2013), rising market power (Barkai (2020), Autor et al. (2020), (De Loecker et al., 2020)), and long-run substitution of capital for labor driven by investment-specific technical change (Karabarbounis and Neiman (2013), Eden and Gaggl (2019)). Kehrig and Vincent (2021) show that the decline in the labor share is driven by reallocation toward low labor-share firms, and not by a decline in the mean (or median) labor share. My paper is closest to explanations emphasizing the role of automation of production processes for the decline in labor demand, including Acemoglu and Restrepo (2018a), Aum and Shin (2020), Cheng et al. (2021), Hubmer and Restrepo (2021), Aum and Shin (2022), and Hémous and Olsen (2022). Much of this literature does not explore implications of automation for labor income inequality, focusing instead on the functional distribution of income between labor and capital. Aum and Shin (2020) propose that the steep descent in the labor share owes at least partly to the rising ability of software capital to displace relatively skilled workers, the mechanism I attempt to quantify in this paper, but do not discuss implications for the skill premium. To the best of my awareness, there is no work that directly connects the large literature on the declining labor share to the literature on relative skill demands. This paper fills this gap.

2 The Skill premium and the Labor Share

In this section I describe the slowdown in the skill premium and the decline in the labor share.

First, I describe the construction of the time series for the skill premium. Following much of the

[9]His model achieves this across steady states via endogenous technological change. Intuitively, a lower price of equipment raises the relative demand for complementary skilled labor, raising the demand for software which is itself complementary to skilled labor. But this raises the profits to automating skill-intensive tasks, which offsets the growth of skilled labor demand.

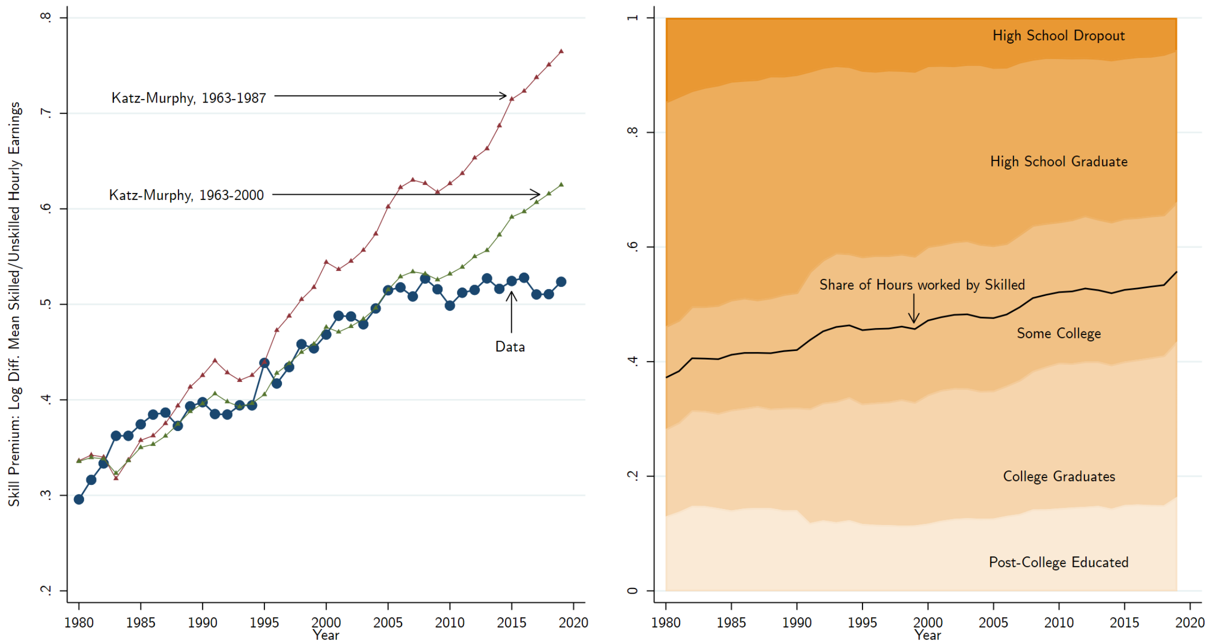


Figure 2: The behavior of the skill premium, 1980-2019 and the behavior of the relative supply of skilled workers. All data from CPS Annual Socio-Economic Complement for full-time-full-year employed men aged 16-64. Hourly earnings defined as market labor income divided by usual hours worked per year, itself the product of usual hours worked per week times weeks worked per year. Earnings are composition-adjusted for race and experience following Autor (2019). More details on data cleaning in Appendix section A.3. The relative supply of skilled workers, here plotted as the share of hours worked by skilled workers, is the share of hours worked by workers with a college degree or a post-college education plus half the share of workers with some college. The red line represents the predicted skill premium obtained by running the time-series regression $\log(w_{st}/w_{ut}) = \beta_0 + \beta_1 \log(\ell_s/\ell_u) + \beta_2 t + \varepsilon_t$ over the period 1963-1987, following Katz and Murphy (1992). The green line is the predicted skill premium obtained from running the same regression on data from 1963 to 2000.

literature linking technology to skills, I rely on the Current Population Survey's Annual Socio-Economic Complement Dataset[10], available from IPUMS (Flood et al., 2021). Throughout this paper, I focus on males aged 18-65 who are full-time-full-year employed (i.e. work at least 35 hours a week for at least 50 weeks a year). I drop workers with hourly wages under the real minimum wage and drop observations in the top 1% of the wage distribution by sex in each year to minimize the impact of outliers and the adjustments made for topcoding. Appendix section A.3 provides more details on the data cleaning procedure.

The left panel of figure 2 shows the slowdown in the growth rate of the skill premium. Between 1980 and 2000, the gap between composition-adjusted hourly earnings of skilled and unskilled male workers rose from about 32 log points to over 55 log points, an increase of 23 log points. Between 2000 and 2019, this gap rose by just 5.5 log points. This slowdown is not just driven by

[10]The CPS provides data on individual earnings in the form of market income earned from wages or self-employment over the reference year, and I define labor income as the sum of these two components, following Hoffmann et al. (2020). The CPS also provides data on usual hours worked, and I use this data to construct a measure of hourly earnings for each worker.

the wages of workers with just a bachelor's degree - the premium for having some postgraduate education rose by 39 log points between 1980 and 2000, and then rose just 5.8 log points between 2000 and 2019. This striking slowdown since about 2000 has also been documented in Beaudry et al. (2016) and discussed in Valletta (2018).

Can a rising relative supply of skilled labor by itself account for this slowdown? First, the right panel of figure 2 shows that while the relative supply of skilled labor has increased over this period, there is no clear trend break in 2000 that can account for this slowdown. Second, to make this point more formally, I replicate an exercise in Autor (2017) and plot the predicted skill premium one would obtain using the Katz and Murphy (1992) regression

$$\log(w_{st}/w_{ut}) = \beta_0 + \beta_1 \log(\ell_{st}/\ell_{ut}) + \beta_2 t + \varepsilon_t$$

estimated on different subsets of the data. Using the original Katz and Murphy sample of data from 1963-1987, one obtains a substantial over-prediction of the skill premium, by over 20 log points at the end of the sample period. Using an extended sample until 2000, the regression's fit on the in-sample period improves dramatically, but still predicts more rapid growth in the skill premium over this period than in the data. This data description device suggests that the combination of growth in skilled labor supply and substitution of skilled labor for unskilled labor driven by exogenous skill-biased technical change, proxied for by the time trend in this regression, cannot by itself account for the slowdown in the skill premium. Appendix A.1 considers some other possible forces for the slowdown in the skill premium, including composition effects across occupations and industries.

Figure 3 shows the path of the labor share, which exhibits a trend of stability prior to 2000 followed by a striking decline post 2000. This decline is not driven by a compositional shift towards relatively low or declining labor share sectors. Alternative measures of the labor share follow very similar trends[11].

Figure 4 decomposes the labor share by skill type. It shows that the stability of the labor share in the period 1980-2019 is explained by a rise in the skilled labor share that almost perfectly offsets the decline in the unskilled labor share. Post 2000, a noticeable slowdown in the growth rate of the skilled labor share means a weakening of this offsetting force. In appendix figure 18 I show that the slowing growth in the skilled labor share explains almost 70% of the 6.6 percentage point decline in the total labor share. Figure 19 shows that over the period 2000-2017, this slowdown is visible across most sectors of the economy. The fact that the decline in the labor share is largely explainable by the declining share of relatively skilled workers, raises questions about stories for the declining labor share which emphasize the role of factors which are neutral with respect to skills, such as the role of rising markups[12]. By contrast, my model will be able to qualitatively account for the decline

[11]See Koh et al. (2020), Grossman and Oberfield (2022) and Gutiérrez and Piton (2020) for discussions of measurement issues surrounding the labor share's decline.

[12]It is possible that rising markups can be non-neutral with respect to skills if skilled labor is disproportionately

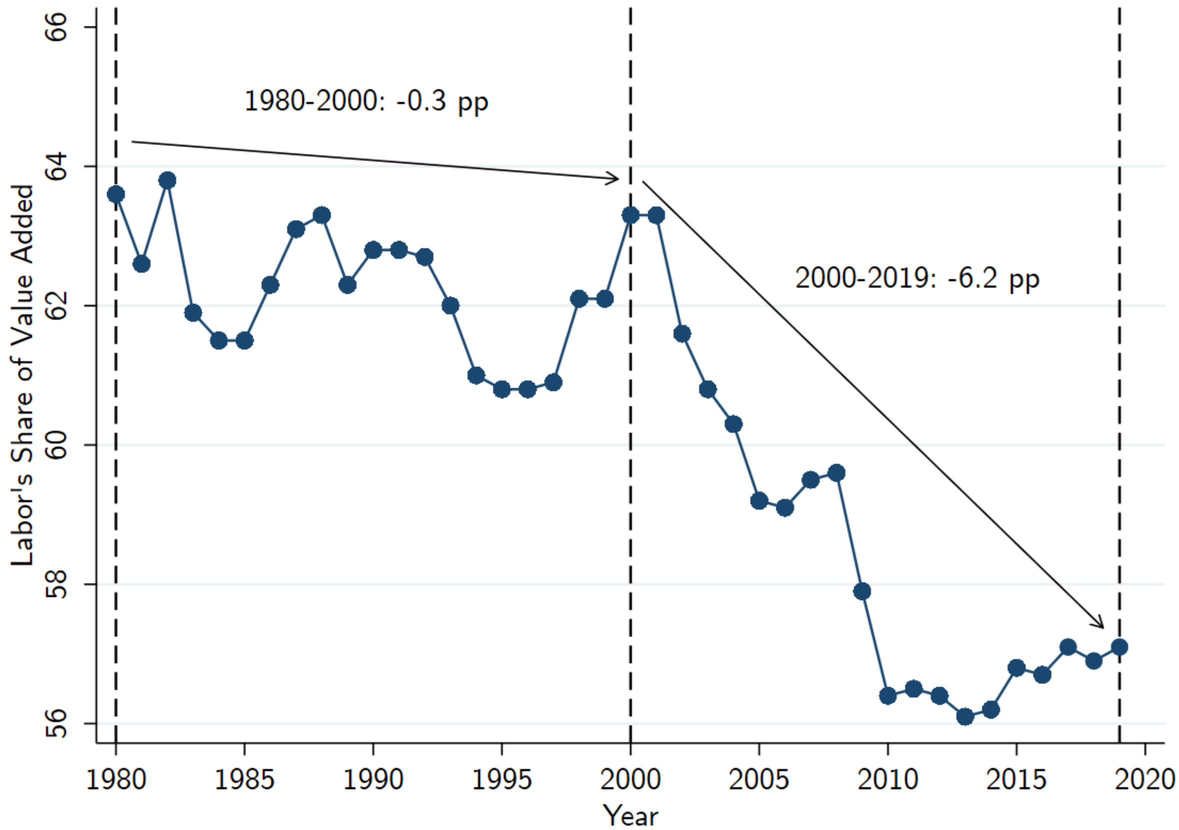


Figure 3: The decline in the labor share, here the non-farm business labor share of value added. Data from the Bureau of Labor Statistics and the BEA-BLS Integrated National Accounts.

of the labor share via a slowdown in the skilled share.

3 Model

Time is discrete and denoted by $t = 0, 1, \ldots$. I consider a small open economy in which a single risk-free asset can be traded in international markets promising a constant rate of return $\bar{r}$. There are no aggregate shocks in the economy, and I will focus on perfect foresight transitions or steady states throughout. There are three kinds of goods in the economy: a final good, a continuum of fixed mass of differentiated intermediate goods and capital. The final good is the numeraire in this economy and is the only good which is tradable. The economy has three agents: final good retailers, a fixed mass of intermediate good producers and representative households. The final good is produced competitively by final good retailers who purchase intermediate goods to produce it. The final good can be consumed or converted to capital for saving at the rate of q_{kt} final goods per

compensated by payments in stock options, which represent claims on the profits of firms. Eisfeldt et al. (2021) provide some evidence that this is the case.

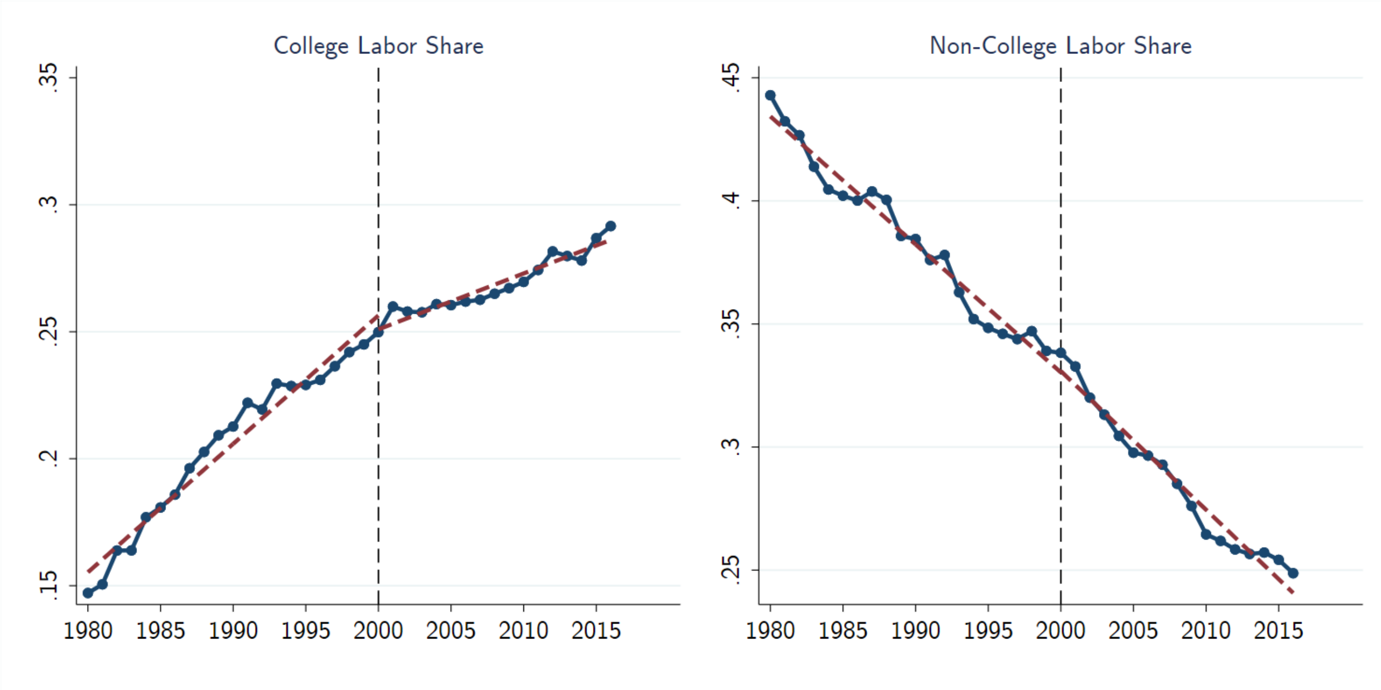


Figure 4: The decomposition of the labor share by skill level. Data from the Bureau of Labor Statistics and the BEA-BLS Integrated National Accounts.

unit of capital. I assume that the time path of q_{kt} is exogenous[13].

The representative household consumes final goods. It saves in capital and in final-good-denominated assets paying a world interest rate $\bar{r}$, assumed constant. I assume that the household's discount factor β satisfies $\beta(1+\bar{r}) = 1$. Households own all firms and retailers in the economy. They are endowed with a certain amount S_t of skilled labor and $U_t = H - S_t$ of unskilled labor, which they supply inelastically to intermediate goods producers each period. They can save by accumulating capital or in the internationally traded bond paying $\bar{r}$. To accumulate a unit of capital at date t, a household operates a technology that converts q_{kt} units of the final good into one unit of capital. The path for q_{kt}, which will reflect the relative price for capital, is the key driving force in this model and will be exogenously given. A unit of capital accumulated at date t is operated by firms at date $t+1$, and in the process of production, it depreciates at rate δ. The renting intermediate good firm pays a rental rate r_{kt+1} and returns the undepreciated fraction $1-\delta$ of the unit of capital rented to the household at date $t+1$. All wages and rental payments are paid in units of the final good, and all consumption takes place in the final good. For the purposes of this model, the only roles played by households are to pin down the supplies of skilled and unskilled labor to intermediate good firms, and to pin down the rental rate on capital.

[13]This model is isomorphic to one where capital goods are produced using final goods, in which case q_{kt} can be interpreted as the relative total factor productivity of the technology that produces final goods to the one that produces capital. Falls in q_{kt} correspond to increasing efficiency in the creation of capital goods and a corresponding decline in their relative price, as in Greenwood et al. (1997).

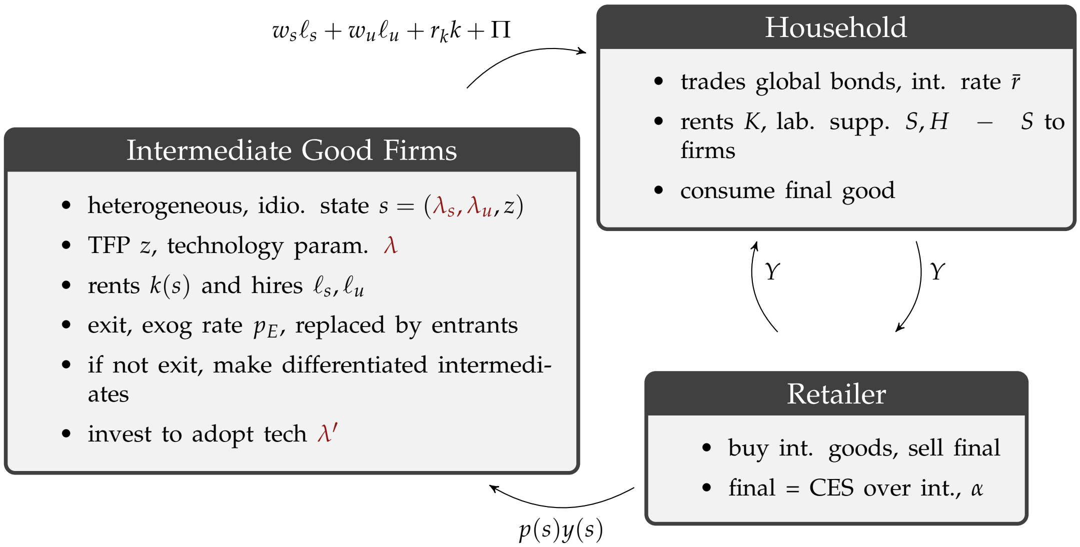


Figure 5: Model structure.

Final goods are produced by the final goods retailer, which purchases the output of all intermediate good firms in the economy and packages them into the final good. Since intermediate goods are differentiated and intermediate good firms are monopolists in the production of their specific intermediate, they behave monopolistically and make profits, taking the downward sloping demand for their specific intermediate goods as given. These profits will induce firms to invest in new technologies as the price of capital q_{kt} falls over time.

Each differentiated intermediate good is produced by a single intermediate good firm. Production of intermediate goods requires the completion of a measure 1 of tasks which only skilled labor can produce and a measure 1 of tasks which only unskilled labor can produce. Let $x_u \in [0,1]$ and $x_s \in [0,1]$ index skill and unskill-intensive tasks respectively. A unit of unskilled labor produces $\psi_u(x_u)$ units of the unskilled task x_u, while a unit of skilled labor produces $\psi_s(x_s)$ units of the skilled task x_s. Tasks of each type are ordered such that $\psi_s' > 0$ and $\psi_u' > 0$. All tasks can in principle be produced with capital as well, and the productivity of capital at each task x_u and x_s is assumed to be 1. Thus, $\psi_u(x_u)$ and $\psi_s(x_s)$ also characterize the comparative advantage that unskilled and skilled labor have at producing each task, and the assumptions that $\psi_s' > 0, \psi_u' > 0$ correspond to imposing that tasks are ordered in increasing order of labor's comparative advantage[14]. An intermediate good firm hires skilled and unskilled labor and rents capital from the households. It then decides on an allocation of these factors of production to tasks.

In each period, the technology an intermediate good firm operates is parametrized by three numbers $s = (\lambda_s, \lambda_u, z)$ where z is the firm's total factor productivity and λ_s, λ_u are parameters that govern the extent to which firms' technologies rely on skilled labor inputs and unskilled labor inputs. In

[14]One way to interpret this ordering is that it is increasing in the "non-routineness" of the tasks, i.e. higher x_i tasks are less routine than lower ones.

particular, at date t, it is only feasible for firms to perform unskilled tasks in the interval $[0, \lambda_u]$ and skilled tasks in the interval $[0, \lambda_s]$ with capital. The technology a firm operates therefore involves a constraint on the measure of each type of tasks that capital can perform. Henceforth, I will refer to λ_s, λ_u as the *capital feasibility cutoffs* characterizing a technology. After production, intermediate good firms choose whether or not to make innovation investments, and if they choose to, how much to invest. Intermediate good firms which make innovation investments will enter the next period with new capital feasibility cutoffs $\lambda_i' \geq \lambda_i, i = s, u$. In each period, I assume that intermediate good firms pay all of their profits net of innovation investments as dividends to the representative household. The vector s is therefore also the idiosyncratic state for all intermediate good firms.

3.1 Model Setup

3.1.1 Household

The economy is populated by a representative household which has H members. The representative household provides all of its members with a common level of consumption. At date t the household has S_t skilled members and $U_t = H - S_t$ unskilled members, each of whom supplies one efficiency unit of their respective labor type to the firm inelastically. At each date, the household provides labor services S_t, U_t, consumes C_t, and saves. It can save by either buying internationally traded bonds B_{t+1}, or by accumulating homogeneous capital[15][16] K_{t+1}^S. The household's problem is to choose sequences $\left\{C_t, K_{t+1}^S, B_{t+1}\right\}_{t=0}^{\infty}$ such that given initial capital holdings K_0, bond holdings B_0, the world interest rate $\bar{r}$ and perfectly foreseen paths for labor supplies $\{S_t, U_t\}_{t=0}^{\infty}$ and the price of capital $\{q_{kt}\}_{t=0}^{\infty}$, the household's chosen sequences maximize

$$
\begin{aligned}
\max_{\left\{C_t, B_{t+1}, K_{t+1}^S\right\}} \quad & \sum_{t=0}^{\infty} \beta^t \log C_t \qquad \text{subject to} \\
C_t + q_{kt} K_{t+1}^S + B_{t+1} \quad \leq \quad & w_{st} S_t + w_{ut} U_t + (r_{kt} + (1-\delta) q_{kt}) K_t^S + \Pi_t + (1+\bar{r}) B_t \\
& \lim_{s \to \infty} \frac{B_{s+1}}{(1+\bar{r})^s} \leq 0
\end{aligned}
$$

The household's optimality conditions for consumption give the Euler equation

$$
\frac{C_{t+1}}{C_t} = \beta (1 + \bar{r}) \tag{1}
$$

[15] I assume that bond payoffs are denominated in final goods and that the household trades these bonds with deep-pocketed risk-neutral international investors, who are endowed with a sufficiently large endowment of final goods to be able to meet any obligations associated with their bond position. The real rate of return on internationally traded bonds, $\bar{r}$, thus satisfies $\frac{1}{1+\bar{r}} = \beta^F$ where β^F is the one-period discount factor these investors apply to payoffs denominated in final goods. I assume that $\beta^F = \beta$.

[16] The superscript S on capital K^S is used to distinguish capital supply from capital demand, which will be denoted K.

The household's optimality conditions for the choices of K^S_{t+1} and B_{t+1} can be combined to obtain the no-arbitrage condition

$$1+\bar{r}=\frac{r_{kt+1}+(1-\delta)\,q_{kt+1}}{q_{kt}} \tag{2}$$

which equates the returns to saving on the two assets. The returns to saving in capital include the rental rate on capital relative to the cost of purchasing the unit of capital, as well as the capital losses, if any, on the undepreciated capital stock remaining after production. Holding q_{kt+1} fixed, a fall in the level of capital prices today q_{kt} requires a decline in r_{kt+1} to restore no-arbitrage.

3.1.2 Final Goods Retailers

Final goods are produced by a competitive final goods retailer which packages the output of intermediate good firms into final output. Denote the vector of idiosyncratic state variables that an intermediate good firm in the economy takes as given by s. Recall that $s=(\lambda_s,\lambda_u,z)$. The competitive final goods retailer operates the technology

$$Y_t=\left[\int y_t(s)^{\frac{\alpha-1}{\alpha}}dM_t(s)\right]^{\frac{\alpha}{\alpha-1}} \tag{3}$$

where $M_t(s)$ is the mass of actively producing firms with idiosyncratic state s at date t. The final good has a price $P_t=1$, i.e. final output is the numeraire. The final goods retailer therefore solves the profit maximization problem

$$\max_{Y_t,y_t(s)} Y_t-\int p_t(s)y_t(s)dM_t(s) \text{ subject to } [\lambda^F_t]: Y_t=\left[\int y_t(s)^{\frac{\alpha-1}{\alpha}}dM_t(s)\right]^{\frac{\alpha}{\alpha-1}}$$

In appendix section A.4, I show that the focs of this problem yield the following demand curves for firms with state s:

$$y_t(s)=(p_t(s))^{-\alpha}\,Y_t \tag{4}$$

and, since the final good is the numeraire, we require[17] that

$$\int p_t(s)^{1-\alpha}dM_t(s)=1 \tag{5}$$

3.1.3 Intermediate Goods Producers: Technology

In each period, the technology a continuing intermediate good firm operates is parametrized by three numbers $s=(\lambda_s,\lambda_u,z)$ where z is the firm's total factor productivity and λ_s,λ_u are parameters

[17] This equation is just the analogue of the profit maximization equation from any CES production function, and effectively states that since the final good retailer behaves competitively, the price it charges $P_t=1$ must equal the marginal cost of production, which is a nonlinear combination of the prices of the intermediate goods it aggregates. Acemoglu and Restrepo (2018a) and Hubmer and Restrepo (2021) call this equation the ideal price index condition.

that govern the extent to which firms' technologies rely on skilled labor inputs and unskilled labor inputs.

I now describe in detail how the technology that a firm enters a period t with works. Producing output involves completing a measure 1 of skill-specific tasks. Let $x_s \in [0,1]$ and $x_u \in [0,1]$ index these tasks. I assume that each type of tasks is specific to its labor type - that is, skilled labor only performs tasks $x_s \in [0,1]$ and unskilled labor only performs tasks in $x_u \in [0,1]$. This rules out the possibility of substitution of skilled labor for unskilled labor *within* a task[18]. At date t, it is feasible for a firm with technology parameters λ_s and λ_u to perform skilled tasks in the interval $[0,\lambda_s]$ and unskilled tasks in the interval $[0,\lambda_u]$ with capital. A unit of unskilled labor produces $\psi_u(x_u)$ units of the unskilled task x_u, while a unit of skilled labor produces $\psi_s(x_s)$ units of the skilled task x_s. I assume that tasks of each type are ordered such that $\psi_s' > 0$ and $\psi_u' > 0$ - that is, labor's comparative advantage at producing tasks strictly increases with the tasks's index[19].

At each date, the firm entering with state $s = (\lambda_s, \lambda_u, z)$chooses an allocation of each labor type and capital across each task

$$\left\{\{\ell_u(x_u), k_u(x_u)\}_{x_u\in[0,1]}, \{\ell_s(x_s), k_s(x_s)\}_{x_s\in[0,1]}\right\}$$

where $k_s(x_s), k_u(x_u)$ denote the quantities of capital allocated to the production of each set of tasks. Given this allocation, output produced is given by

$$y(\lambda_s,\lambda_u,z) = z\left[\mu G_u^{\frac{\sigma-1}{\sigma}} + (1-\mu)G_s^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{6}$$

where for $i = u, s$,

$$\begin{aligned} G_i &= \left[\int_0^1 \mathcal{Y}_i(x)^{\frac{\rho-1}{\rho}}dx\right]^{\frac{\rho}{\rho-1}} \\ \mathcal{Y}_i(x) &= \begin{cases} \psi_i(x)\ell_i(x) + k_i(x) & x \le \lambda_i \\ \psi_i(x)\ell_i(x) & x > \lambda_i \end{cases} \end{aligned}$$

and the functions ψ_u, ψ_s satisfy the following assumptions:[20]

- $\psi_u'(x) > 0, \psi_s' > 0$. That is, tasks are ordered so that labor's comparative advantage is increasing in the task index for each labor type.

[18]This assumption makes the technology firms operate in my model different from the one outlined in Acemoglu and Restrepo (2018a), and is analogous to assumption 1 in Acemoglu and Restrepo (2022) and is made for tractability. Relaxing it, possibly along the lines that Acemoglu and Restrepo (2022) do using a local approximation, is left for future research.

[19]This assumes that there is a unidimensional attribute of each task, say its "routineness", which fully determines the comparative advantage labor has at that task. See Lindenlaub (2017) for an analysis of sorting across tasks with multiple attributes.

[20]These assumptions are Inada-like conditions and guarantee that there is always some demand for capital, unskilled and skilled labor, no matter how skewed factor prices are in favor of one factor.

- $\frac{1}{\psi_i(x_i)}$ is convex in x_i. That is, as the task index rises, the unit labor requirement for a given task declines at a slower rate for a given increment in x_i. Section 5 discusses the importance of this assumption.
- $0 \leq \lim_{x \to 0} \psi_u(x), \lim_{x \to 0} \psi_s(x) < 1$. That is, capital is more productive than labor at at least some tasks.
- $\lim_{x \to 1} \psi_u(x) = \lim_{x \to 1} \psi_s(x) = \infty$. That is, labor is more productive than capital at at least some tasks.
- $\int_z^1 \psi_i(x)^{\rho-1} dx$ exists for all $z \in [0,1]$ and for both $i = u, s$. That is, a well-defined index of labor productivity exists for both types.

I assume that firm TFP z follows an AR(1) process in logs, so that

$$\log z_{t+1} = \rho_z \log z_t + \varepsilon_{t+1} \text{ where } \varepsilon \sim N\left(0, \sigma_\varepsilon^2\right) \tag{7}$$

Finally, in order to upgrade its technology parameter λ_i to λ_i', a firm pays the cost $\kappa_i\left(\lambda_i', \lambda_i\right) Y_t$ where

$$\kappa_i\left(\lambda_i', \lambda_i\right) = \begin{cases} 0 & \lambda_i' \leq \lambda \\ \kappa^0\left(\lambda_i' - \lambda_i\right) Y & \lambda_i' > \lambda \end{cases}$$

This formulation for the upgrading cost has two properties. First, the marginal cost of innovation is always positive, which will be useful in guaranteeing the existence of a steady state in which all firms use the same technology. Second, the normalization by GDP in the cost function is important to ensure that the firm's value function is homogeneous of degree 1 in aggregate output, a precondition for balanced growth when the relative price of capital is constant. I interpret the cost $\kappa_i(\cdot)$ as a stand-in for a wide range of costs that firms must pay in order to upgrade their technologies, including the cost of purchasing and installing new capital (Acemoglu and Restrepo (2020a)), the foregone output due to the time required to retool production processes (Kopytov et al. (2018)), as well as the managerial and organizational costs associated with the introduction of new technologies (Bresnahan et al. (2002)).

3.1.4 Exit and Entry

After production occurs, intermediate goods firms exit with an exogenously given probability p_E. If they exit, their realized value is zero. If they do not, they choose the technology they want to enter the next period with and pay the costs associated with this choice. A measure $p_E \bar{M}$ of firms exit in each period, and a mass $p_E \bar{M}$ new entrants enter the economy at this date to replace them, ensuring that the mass of firms is always constant. At entry, a new entrant draws an initial productivity parameter z_E from the stationary distribution of the $AR(1)$ process for TFP 7. I assume that new entrants' technologies are parametrized by λ_{Est} and λ_{Eut} where

$$\lambda_{Eit} = \frac{1}{\bar{M}} \int \lambda_{it}(s) dM_t(s)$$

for $i = s, u$ and $\bar{M}$ is the constant mass of intermediate good firms. This allows for technology diffusion in the economy via entry. While inconsequential for the main mechanism I model, allowing for technology diffusion via entry is important for the dynamics of the model and in order to establish the existence of a well-defined steady state.

3.1.5 Intermediate Goods Producers' Problem

The problem of a firm is to choose sequences of factor inputs $\{k_t, \ell_{st}, \ell_{ut}\}$, and the sequence of allocations of capital and each labor type over tasks at each date,

$$\left\{ \{\ell_{ut}(x_u), k_{ut}(x_u)\}_{x_u \in [0,1]}, \{\ell_{st}(x_s), k_{st}(x_s)\}_{x_s \in [0,1]} \right\}$$

and sequences of technology parameters $\lambda_t \equiv \{\lambda_{ut}, \lambda_{st}\}$ to maximize the present value of its profits net of costs of technology upgrading. When solving this problem, the firm takes the paths of factor prices w_{st}, w_{ut}, r_{kt} and aggregate demand Y_t as given. Given the aggregate state vector $\boldsymbol{s}_t = (w_{st}, w_{ut}, r_{kt}, Y_t, q_{kt})$ the firm's value V_t^F solves the following Bellman Equation.

$$\begin{aligned} V_t^F(\lambda, z; \boldsymbol{s}_t) &= \pi_t(\lambda, z; \boldsymbol{s}_t) \\ &+ (1 - p_E) \max_{\lambda_i' \geq \lambda_i} \left\{ - \sum_{i=u,s} \kappa_i(\lambda_i', \lambda_i) Y_t + \frac{\mathbb{E}_t\left[V_{t+1}^F(\lambda', z'; \boldsymbol{s}_{t+1})\right]}{1 + \bar{r}} \right\} \end{aligned} \tag{8}$$

where the period profit function $\pi_t(\lambda, z; \boldsymbol{s}_t)$ solves

$$\begin{aligned} \pi_t(\lambda, z; \boldsymbol{s}_t) = \max_{\left\{G_i, \{y_{Gi}(x), \ell_i(x), k_i(x)\}_{x=0}^1\right\}_{i=u,s}, p, y} \quad & py \\ & - \int_0^1 (r_{kt} k_u(x) + w_{ut} \ell_u(x))\, dx \\ & - \int_0^1 (r_{kt} k_s(x) + w_{st} \ell_s(x))\, dx \end{aligned}$$

subject to, for $i = s, u$,

$$
\begin{aligned}
& y \geq p^{-\alpha} Y_t \\
& y \leq z \left[\mu G_u \left(\ell_u, k_u \right)^{\frac{\sigma-1}{\sigma}} + (1-\mu) G_s \left(\ell_s, k_s \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \\
& G_i \leq \left[\int \mathcal{Y}_i \left(x_i \right)^{\frac{\rho-1}{\rho}} dx_i \right]^{\frac{\rho}{\rho-1}} \\
& \mathcal{Y}_i \left(x_i \right) = \begin{cases} \psi_i \left(x_i \right) \ell_i \left(x_i \right) + k \left(x_i \right) & x_i \leq \lambda_i \\ \psi_i \left(x_i \right) \ell_i \left(x_i \right) & x_i > \lambda_i \end{cases} \\
& k_i \left(x_i \right) \geq 0 \\
& \ell_i \left(x_i \right) \geq 0
\end{aligned}
$$

and TFP z follows the law of motion

$$
\log z_{t+1} = \rho_z \log z_t + \varepsilon_{t+1} \text{ where } \varepsilon \sim N(0, \sigma_\varepsilon^2)
$$

It is analytically convenient to split this problem into a static part and a dynamic part. First, we solve for the cost-minimizing allocation of capital and each type of labor across tasks associated with the production of a unit of the task intermediate G_i. Let $P_{Git}(\lambda_i; \boldsymbol{s}_t)$ be the minimized cost of producing a unit of the task intermediate $G_i, i = s, u$. Given $P_{Git} \left(\lambda_i; \boldsymbol{s}_t \right)$, we next solve for minimized cost of producing a unit of the intermediate good y. Let $C_{Ft} \left(\lambda, z; \boldsymbol{s}_t \right)$ be this minimum unit cost, which by constant returns is the marginal cost function. Given the marginal cost of producing intermediate good output, we can solve for the optimal price and output decision made by the firm to obtain the profit function. Second, given the profit function $\pi_t \left(\lambda, z; \boldsymbol{s}_t \right)$ we can solve the Bellman Equation 8 to obtain optimal technology adoption decisions. Let $g_{\lambda it} \left(\lambda, z; \boldsymbol{s}_t \right)$ denote the policy functions for choices of new technology.

3.1.6 Static Cost Minimization and Profit Maximization

In this section, I characterize the static cost function associated with the firm's choice of inputs and show that it is well-defined and isomorphic to that of a nested CES production function. I also show that the cost function is convex in the technology parameters, guaranteeing that the profit function is concave in the technology parameters and ensuring that a well-defined solution to the firm's dynamic problem exists.

Lemma 1. *Consider a firm with state $s = (\lambda_s, \lambda_u, z)$. The cost-minimizing allocation of factors of production to tasks is characterized as follows. For $i = s, u$ let*

$$
\lambda_i^* \left(\lambda_i; \boldsymbol{s}_t \right) = \min \left\{ \lambda_i, \hat{\lambda}_i \left(\boldsymbol{s}_t \right) \right\}, \textit{ and } \hat{\lambda}_i \left(\boldsymbol{s}_t \right) \textit{ solves } \frac{w_{it}}{\psi_i \left(\hat{\lambda}_i \left(\boldsymbol{s}_t \right) \right)} = r_{kt}
$$

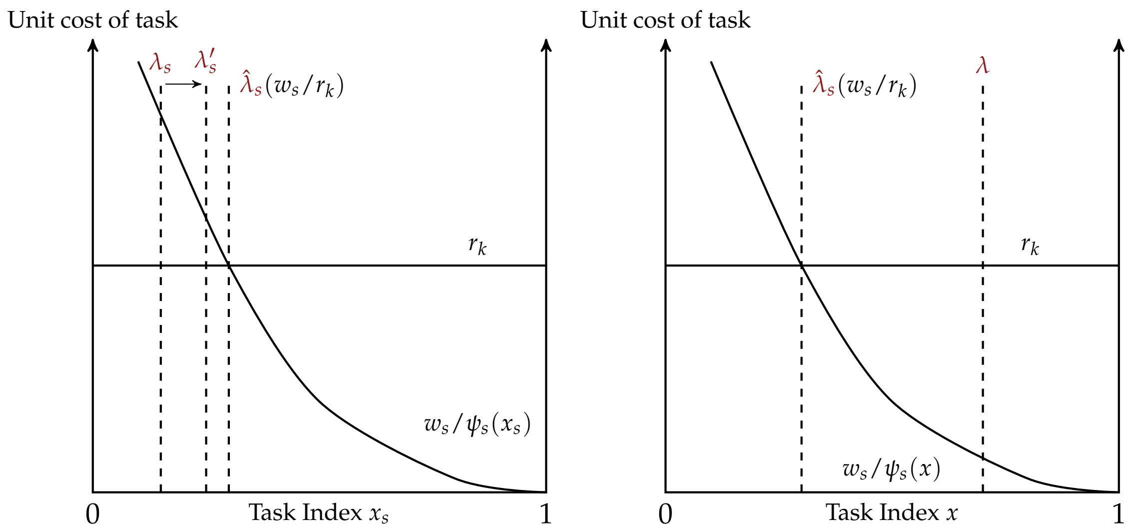


Figure 6: Left Panel: Determination of λ_s^* in the case where the constraint on technological possibilities is binding (the case for $i = u$ is completely symmetric). A cost-minimizing firm that could freely choose λ_s^* would choose to set $\lambda_s^* = \hat{\lambda}\,(w_s/r_k)$, but this is impossible because the technological possibilities at that firm only extend until $\lambda_s < \hat{\lambda}_s$. The gap between $w_s/\psi_s(\lambda)$ and r_k is a measure of how valuable additional automation investments are to the firm. If this gap is sufficiently large to outweigh the costs of investment in new technologies, all else equal, the firm will choose a higher λ in the next period. The Right Panel shows an analogous situation, except now $\lambda_s > \hat{\lambda}\,(w_s, r_k)$. Such a firm always chooses to set $\lambda_s^* = \hat{\lambda}\,(w_s, r_k)$.

Then, the firm optimally chooses to produce all tasks of type i in the interval $x_i \in [0, \lambda_i^*]$ using capital and all tasks $x_i \in (\lambda_i^*, 1]$ using labor ℓ_i.

Proof. In appendix. □

This allocation rule is intuitive. Given factor prices w_i, r_k and the monotonicity and Inada conditions on ψ_i, there is a cutoff task $\hat{\lambda}_i$ associated with each labor type, such that for any task $x_i < \hat{\lambda}_i$ it is optimal to execute the task using capital and for any task $x_i \geq \hat{\lambda}_i$ it is optimal to use labor (recall that the comparative advantage labor of type i enjoys over capital, $\psi_i\,(x_i)$, is increasing in x_i). If the current technology parameter of the firm λ_i exceeds this cutoff, the firm's choice of λ_i^* is unconstrained and equal to $\hat{\lambda}_i$. In this case, the firm's choice of technology is independent of its current technology state. However, if the parameter $\lambda_i < \hat{\lambda}_i$, it is technologically infeasible for the firm to use capital to perform tasks in $\left(\lambda_i, \hat{\lambda}_i\right)$, even though it would be profitable for the firm to do so. Again, since $\psi_i\,(x_i)$ is strictly increasing in x_i, it is clearly optimal for the firm in such cases to just set $\lambda_i^* = \lambda_i$. This will be the relevant case for the equilibria I will study. Figure 6 illustrates both situations.

Given this allocation rule, it is straightforward to show that for $i = s, u$, the minimized cost of producing the task intermediate G_i is given by

$$P_{Git}(\lambda_i; \boldsymbol{s}_t) = \left[r_{kt}^{1-\rho} \lambda_i^*\,(\lambda_i; \boldsymbol{s}_t) + w_{it}^{1-\rho} \int_{\lambda^*(\lambda_i; \boldsymbol{s}_t)}^{1} \psi_i(x)^{\rho-1} dx \right]^{\frac{1}{1-\rho}}$$

where, as above,

$$\lambda_i^*(\lambda_i; \boldsymbol{s}_t) = \min\left\{\lambda_i, \hat{\lambda}_i(\boldsymbol{s}_t)\right\}, \text{ and } \hat{\lambda}_i(\boldsymbol{s}_t) \text{ solves } \frac{w_{it}}{\psi_i\left(\hat{\lambda}_i(\boldsymbol{s}_t)\right)} = r_{kt}$$

Given the nested CES structure of the cost function, cost minimization over the choices of how much of each task intermediate G_i to use and an application of Shephard's Lemma give us the following characterization of the cost and conditional input demand functions.

Lemma 2. *Consider a firm with state $s = (\lambda_s, \lambda_u, z)$. The marginal cost function for such a firm is given by*

$$C_{Ft}(s; \boldsymbol{s}_t) = \frac{1}{z}\left[\mu^\sigma P_{Gut}(\cdot)^{1-\sigma} + (1-\mu)^\sigma P_{Gst}(\cdot)^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{9}$$

where

$$P_{Git}(\lambda_i; \boldsymbol{s}_t) = \left[r_{kt}^{1-\rho}\lambda_i^*(\lambda_i; \boldsymbol{s}_t) + w_{it}^{1-\rho}\int_{\lambda^*(\lambda_i; \boldsymbol{s}_t)}^1 \psi_i(x)^{\rho-1}dx\right]^{\frac{1}{1-\rho}}$$

where

$$\lambda_i^*(\lambda_i; \boldsymbol{s}_t) = \min\left\{\lambda_i, \hat{\lambda}_i(\boldsymbol{s}_t)\right\}, \textit{ and } \hat{\lambda}_i(\boldsymbol{s}_t) \textit{ solves } \frac{w_{it}}{\psi_i\left(\hat{\lambda}_i(\boldsymbol{s}_t)\right)} = r_{kt}$$

The conditional input demand functions for this firm satisfy

$$\begin{aligned}
k(s; \boldsymbol{s}_t) &= \frac{y_t(s; \boldsymbol{s}_t)\tilde{C}_{Ft}(s; \boldsymbol{s}_t)^\sigma}{z}\left[\mu^\sigma\frac{P_{Gut}(s; \boldsymbol{s}_t)^{\rho-\sigma}}{r_{kt}^\rho}\lambda_u + (1-\mu)^\sigma\frac{P_{Gst}(s; \boldsymbol{s}_t)^{\rho-\sigma}}{r_{kt}^\rho}\lambda_s\right] && (10)\\
\ell_s(s; \boldsymbol{s}_t) &= \frac{y_t(s; \boldsymbol{s}_t)}{z}\left(\frac{(1-\mu)\tilde{C}_{Ft}(s; \boldsymbol{s}_t)}{P_{Gst}(\lambda_s; \boldsymbol{s}_t)}\right)^\sigma\left(\frac{P_{Gst}(\lambda_s; \boldsymbol{s}_t)}{w_{st}}\right)^\rho\Psi_s(\lambda_s) && (11)\\
\ell_u(s; \boldsymbol{s}_t) &= \frac{y_t(s; \boldsymbol{s}_t)}{z}\left(\frac{\mu\tilde{C}_{Ft}(s; \boldsymbol{s}_t)}{P_{Gut}(\lambda_u; \boldsymbol{s}_t)}\right)^\sigma\left(\frac{P_{Gut}(\lambda_u; \boldsymbol{s}_t)}{w_{ut}}\right)^\rho\Psi_u(\lambda_u) && (12)
\end{aligned}$$

where I define[21] $\Psi_i(\lambda) = \int_\lambda^1 \psi_i(x)^{\rho-1}dx$ *and* $\tilde{C}_{Ft}(\lambda; \boldsymbol{s}_t) = \left[\mu^\sigma P_{Gut}(\cdot)^{1-\sigma} + (1-\mu)^\sigma P_{Gst}(\cdot)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$.

Proof. In appendix. □

Note that the unit cost function 9 is isomorphic to the cost function that emerges from a nested CES production function. In particular, when λ_s, λ_u are constants, so are the objects $\Psi_i(\lambda)$. In this case, it is easy to show that the marginal cost function above is identical to the marginal cost function that would be obtained for a firm operating the two-stage nested CES technology

$$y = z\left[\mu G_u^{\frac{\sigma-1}{\sigma}} + (1-\mu)G_s^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$$

[21] Note that $\Psi_i(\lambda)$ is an index of the productivity of factor input $i = \ell_s, \ell_u$ and that $\tilde{C}_{Ft}$ is just the marginal cost of production for a firm with unit TFP.

in which, for $i = u, s$

$$G_i = \left[\lambda_i k_{it}^{\frac{\rho-1}{\rho}} + \Psi_i\left(\lambda_i\right) \ell_{it}^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}}$$

This technology is isomorphic to the one presented in, for instance, Autor et al. (2022). The key distinction between the nested CES formulation and the model I present is that in the former, λ_i is a structural parameter which is typically assumed to be fixed whereas in my model, λ_i is flexible and an endogenous choice that firms make. Thus, while my model allows for capital-labor substitution via movements along an isoquant exactly as in standard models, it features the additional possibility of actually rotating the isoquants themselves to reach an even lower isocost line. Finally, given the cost function $C_{Ft}\left(s, \boldsymbol{s}_t\right)$, we can define the profit function as the maximized value of the profits the firm earns given its technology and the demand curve it faces for its output from the final goods retailer,

$$\pi_t\left(s; \boldsymbol{s}_t\right) = \max_{p,y} py - C_{Ft}\left(s; \boldsymbol{s}_t\right) y \quad \text{subject to} \quad p = \left(\frac{y}{Y_t}\right)^{-1/\alpha}$$

The solution to this static problem yields the standard constant markup pricing equation

$$p_t\left(s; \boldsymbol{s}_t\right) = \frac{\alpha}{\alpha - 1} C_{Ft}\left(\lambda, z; \boldsymbol{s}_t\right) \tag{13}$$

which in turn yields the profit function

$$\pi_t\left(s; \boldsymbol{s}_t\right) = \frac{Y_t}{\alpha^\alpha}\left(\frac{C_{Ft}\left(s; \boldsymbol{s}_t\right)}{\alpha - 1}\right)^{1-\alpha} \tag{14}$$

Since $C_{Ft}(s; \boldsymbol{s})$ is convex and $\alpha > 1$, the profit function is a decreasing transformation of C_{Ft} and is strictly concave.

3.2 Equilibrium and Steady State Characterization

I introduce the notation $\lambda \equiv (\lambda_s, \lambda_u)$ for exposition, so that $s = (\lambda, z)$. I also drop dependence on the aggregate states to conserve on space. Given an initial distribution of firms over their idiosyncratic states $M_0(s)$, initial capital holdings K_0, initial skilled labor supply S_0, a global interest rate $\bar{r}$, and an exogenous path $\{q_{kt}\}$ for capital prices, an **equilibrium** is

- an allocation
$$\left\{C_t, B_{t+1}, K_{t+1}, S_{t+1}, Y_t, \{k_t(s), \ell_{st}(s), \ell_{ut}(s), y_t(s)\}_{s=(\lambda,z)}\right\}$$
- a sequence of technology choices $\{\lambda_{t+1}(s)\}_{s=(\lambda,z)}$,
- a distribution of firms over the idiosyncratic state space at each date $\{M_t(s)\}$,
- and a set of prices $\{w_{st}, w_{ut}, r_t, r_{kt}, \{p_t(s)\}_s\}$

such that the following conditions all hold.

- Consumption growth follows the consumption Euler equation 1.
- The rental rate for capital r_{kt} and the world interest rate $\bar{r}$ are related by equation 2.
- Firms' input choices $k_t(s), \ell_{st}(s), \ell_{ut}(s)$ and price and output choices $p_t(s), y_t(s)$ satisfy the focs 10, 11 and 12, and the pricing equation 13 given the law of motion for TFP 7 and aggregates $w_{st}, w_{ut}, r_t, r_{kt}, Y_t$. Profits are given by 14.
- Firms' choices of technology parameter $\lambda'(s)$ are consistent with their values satisfying the Bellman equation 8.
- The distribution of firms over the state space $M_t(s)$ satisfies the law of motion

$$\begin{aligned} M_{t+1}\left(\lambda', z'\right) &= (1-p_E)\int \mathbf{1}\left\{g_{\lambda t}(s)=\lambda'\right\}\Pr\left(z' \mid z\right) dM_t\left(\lambda, z\right) \\ &+ p_E \bar{M}\mathbf{1}\left\{\lambda_{Et}=\lambda'\right\}\int \Pr\left(z'\right) d\phi^{stat}\left(z'\right) \end{aligned} \tag{15}$$

  where $\phi^{stat}\left(z'\right)$ is the value of the cumulative density associated with the stationary distribution of the law of motion for TFP 7 and $\bar{M}$ is the constant mass of intermediate good firms.

- Labor markets clear,

$$\int \ell_{st}(s)dM_t(s) = S_t \tag{16}$$

$$\int \ell_{ut}(s)dM_t(s) = H - S_t \tag{17}$$

- The economy's resource constraint is satisfied.

$$\begin{aligned} B_{t+1} - (1+\bar{r})B_t &= Y_t - C_t - q_{kt}\left(K^S_{t+1} - (1-\delta)K^S_t\right) \\ &- \sum_{i=u,s}\int \kappa_\lambda\left(g_{\lambda it}(s), \lambda\right) Y_t dM_t(s) \end{aligned} \tag{18}$$

I now characterize the steady state of the model which will constitute the starting point for the transition dynamics of the model.

Lemma 3. *If $\bar{r} > 0, \kappa^0 > 0$ are positive constants and $q_{kt} = q_k$, $S_t = S$ for all t, then there exists a steady state in which output Y and all factor prices are constant, wages w_s, w_u and the rental cost of capital r_k are constant, all firms operate technologies $\lambda_i \geq \hat{\lambda}_i(\boldsymbol{s}), i = s, u$ where $\boldsymbol{s} = (w_s, w_u, r_k, Y)$ and, recall that $\hat{\lambda}_i(\cdot)$ is defined by $\frac{w_i}{\psi_i\left(\hat{\lambda}_i(\boldsymbol{s})\right)} = r_k$, and there is no investment in further technological adoption.*

Proof. In appendix A.6. □

4 Quantification of Model

In this section I describe how I quantify the model. I first outline the functional form assumptions I make for the task productivity schedules, which are the key objects which discipline the substitutability between capital and labor of each type.

I follow Hubmer and Restrepo (2021) and set

$$\psi_u(x) = B_u \left[x^{\frac{1-\rho-\gamma_u}{\gamma_u}} - 1\right]^{\frac{1}{1-\rho-\gamma_u}} \qquad \text{and} \qquad \psi_s(x) = B_s \left[x^{\frac{1-\rho-\gamma_s}{\gamma_s}} - 1\right]^{\frac{1}{1-\rho-\gamma_s}}$$

where $0 < \rho < 1$ and $\bar{\rho}_i = \rho + \gamma_i > 1$ for both $i = s, u$. Recall that ρ is the elasticity of substitution across task inputs in the task aggregators G_i. To simplify notation, introduce the notation $\psi_{ki} = 1$ for capital's productivity at tasks of type i, so that $\Psi_{ki}(\lambda^*) = \lambda^*$. I show in appendix section A.7 that

$$\Psi_i\left(\lambda^*\right) = \int_{\lambda^*}^{1} \psi_i(z)^{\rho-1} dz = B_i^{\rho-1} \left[1 - (\lambda^*)^{a_i}\right]^{1/a_i}$$

where I define $a_i = \frac{\bar{\rho}_i - 1}{\gamma_i}$. We also have

$$\frac{w_i}{\psi_i\left(\hat{\lambda}\right)} = r_k \implies \frac{w_i/B_i}{r_k} = \left[\hat{\lambda}_i^{\frac{1-\bar{\rho}_i}{\gamma_i}} - 1\right]^{\frac{1}{1-\bar{\rho}_i}} \implies \hat{\lambda}_i = \left(\left(\frac{w_i/B_i}{r_k}\right)^{1-\bar{\rho}_i} + 1\right)^{\frac{\gamma_i}{1-\bar{\rho}_i}}$$

The parameters to calibrate now are $\beta, \bar{r}, \alpha, \rho, \gamma_s, \gamma_u, B_s, B_u, \sigma, \mu, \kappa, p_E$. I set some of these parameters externally, and jointly choose the remaining parameters to hit a set of moments for the initial steady state of the model. First, I set the real interest rate to be 4% a year, consistent with its average value over the sample period, and set $\beta = 1/(1+\bar{r})$ as noted above. I choose $\alpha = 7.7$ to match an average markup of 15% from Barkai (2020). I set $p_E = 6.2\%$ to match the exit rate in Lee and Mukoyama (2015). I choose parameters of the TFP process so that in the steady state the model produces an aggregate sales share for the top 1.1% of all firms equal to 40%, following Autor et al. (2020).

Next, to calibrate ρ, I note that ρ is the elasticity of substitution across worker tasks for both skilled and unskilled workers. I set $\rho = 0.49$, the value in Humlum (2021). To further validate the value I choose for ρ, note that we have[22] $\left.\frac{d\log(k/\ell_i)}{d\log(w_i/r_k)}\right|_{\lambda_s,\lambda_u \text{ fixed}} = \rho$. This implies that one interpretation of ρ is that ρ maps into the short-run elasticity of substitution between capital and each labor type. I confirm that my results are largely robust to values in the range of 0.4, corresponding to the short-run plant level elasticity computed in Oberfield and Raval (2021), to 0.5. Beyond this value, the model becomes numerically unstable. I am working on studying the model's behavior for values of ρ as high as 0.8, the aggregate value that Oberfeld and Raval obtain and close to the upper bound for the consensus aggregate estimate in Knoblach et al. (2020).

The parameters $\mu, \sigma, \gamma_s, \gamma_u, \kappa_s, \kappa_u, B_s, B_u$ are chosen jointly to match a set of moments, which I now

[22] This follows immediately from the factor demand functions in equations 10, 11 and 12.

enlist sequentially. While the value of each of these parameters affects the value of all of the moments I define below, I provide some intuition for which moments are key to help determine the values of the parameters. The value of μ is particularly informative for the steady-state labor share, and I choose it to hit the initial steady state labor share of 62%. The parameters B_s, B_u are chosen so that the model can match wages of skilled and unskilled workers in the 1980 steady state. Note that these parameters are not separately identified from the level of the capital price q_k, so I normalize the capital price series to be 1 in the initial steady state.

To calibrate κ_s and κ_u separately would require estimates of technology adoption costs split by the skill type that uses it. I could not find any such estimates, and therefore choose conservatively to set $\kappa_s = \kappa_u \equiv \kappa_0$. I choose κ_0 so that along the transition path, the model's implied share of GDP spent on technology upgrading is 3%. This is in between the share of GDP spent on ICT products in

To calibrate γ_i, I target estimates of the medium-run substitutability between capital and labor. I conduct the following experiment. For a given pair of values of (γ_s, γ_u), I solve for a steady state in which all firms operate the same technologies $\hat{\lambda}_s^{SS}, \hat{\lambda}_u^{SS}$. I then simulate the behavior of one firm f hit by an unexpected 1% shock to that specific firm's cost of capital, with all other firms remaining unaffected. I assume that this shock is permanent, and that the price of capital for this firm remains at this level forever. The firm therefore is surprised by learning that the paths of the aggregates that it must take into account are now $\left(w_s^{SS}, w_u^{SS}, 0.99 r_k^{SS}, Y^{SS}\right)$. That is, it perceives that except the lower rental cost of capital it faces, all other prices are unchanged forever[23].

I then solve for the firm's choices of its own states $(\lambda_{st}, \lambda_{ut})$ at dates going forward, by iterating on the firm's policy functions $g_{\lambda t}(s)$. With constant factor prices forever, the lower r_k implies that the firm chooses to raise the capital feasibility cutoffs λ_s, λ_u. However, due to the cost of raising the cutoffs, the firm will in general not immediately raise the cutoff to the newly optimal cutoffs $\hat{\lambda}'_s > \hat{\lambda}_s, \hat{\lambda}_u > \hat{\lambda}_u$, but will do so slowly. As it raises the cutoffs, its factor demands will change, leading to a new path for ℓ_s, ℓ_u, k for that firm. I store these paths. I repeat this exercise for each point in the state space, and compute the statistic $\hat{\varepsilon}_i(s) = \left(\frac{w_s^{SS}\ell_s^{SS}(s)+w_u^{SS}\ell_u^{SS}(s)+r_k^{SS}k^{SS}(s)}{r_k^{SS}k^{SS}(s)}\right)\frac{\Delta_{0\to 5}\log(\ell_i(s))}{\Delta_{0\to 5}\log q_k}$ for each point s in the state space. I then calculate the mean value of this elasticity over all firms in the economy using the initial stationary distribution as weights, and ask the model to match the estimates of the statistics σ_{le}, σ_{he} in Berlingieri et al. (2022) which are defined as the analogues of my statistics[24]. I obtain $\gamma_s = 0.79$ and $\gamma_u = 1.1$, which are consistent with greater substitutability between capital and unskilled labor in the medium run than between capital and skilled labor.

Finally, given all the remaining parameters, σ is particularly informative about the elasticity of substitution between skilled labor and unskilled labor. I choose σ to match an initial steady-state elasticity of substitution between skilled and unskilled labor, $\frac{d\log(\ell_s/\ell_u)}{d\log(MP_\ell/MP_u)} = 0.75$, the midpoint of

[23] This experiment is analogous to the one used by Humlum (2021) to study the impact of robot adoption.

[24] Berlingieri et al. (2022) construct an instrument capturing unanticipated variation in the price of capital faced by a firm using bilateral exchange rate shocks and the fact that different firms import capital goods from different sets of countries to estimate the statistic corresponding to $\hat{\varepsilon}_i$ at the firm level.

	Parameter	Value	Source/Target
Elast. Subst. across int. goods	α	7.67	Agg Markup 15% (Barkai 2020)
Production Function	ρ	0.49	Humlum (2019)
	σ	2.75	$\frac{d\log(\ell_s/\ell_u)}{d\log(MP_\ell/MP_u)}=0.75$
	μ	0.15	1980 labor share
	ρ_z	0.95	Estd. TFP Persistence
	σ_z	0.105	top 1% firms have 40% sales in 1982
Comp. Adv. Schedules	γ_s, γ_u	0.76,1.14	Estimates in Berlingieri et al (2022)
	B_s	4.41	w_s, w_u in 1980
	B_u	502.02	
Exit/Entry Rate	p_E	6.2%	Lee and Mukoyama (2015)
Adoption Costs	κ_0	2.3e3	Adoption costs 1% of GDP in 2000

Table 1: Parameter values for the model.

the range of estimates reported by[25] Havranek et al. (2020).

5 Mechanism

In this section I provide some intuition for the mechanism underlying the model's behavior over time. This intuition is largely captured by noting the first order condition of an intermediate good firm which is constrained - that is, for a firm with $\lambda_i \leq \hat{\lambda}_i\left(w_i/r_k\right)$. The first order condition for such a firm's choice of λ_i' is

$$\kappa_i Y_t = \frac{1-p_E}{1+\bar{r}}\left[\frac{\partial \tilde{\pi}_{t+1}\left(\lambda_s', \lambda_u'\right)}{\partial \lambda_{it+1}}\mathbb{E}\left(\left(z'\right)^{\alpha-1} \mid z\right) + \kappa_i Y_{t+1}\right]$$

where recall that $\tilde{\pi}\left(\lambda\right) = \pi\left(\lambda, 1\right)$, i.e. profits for a firm with unit productivity. The left side reflects the constant marginal costs of raising λ_i, while the right side reflects the two marginal benefits.

[25] Standard estimates of the elasticity of substitution between skilled labor and unskilled labor exploit time series variation in the shares of skilled labor and unskilled labor in a setting abstracting from capital accumulation, or fail to allow for flexible patterns of substitution between capital and either labor type. Virtually without exception, the notion of skilled-unskilled labor substitution these methods capture reflects movements along an isoquant, identified by plausibly exogenous variation in supplies of labor (note that in the standard Katz and Murphy (1992) approach to identifying the elasticity, identification is only possible because they assume that labor supply is exogenous and that the production function imposes a constant elasticity of substitution between skilled labor and unskilled labor.). Through the lens of my model, this identifying variation is confounded by two forces. First, in my model, the elasticity of substitution between capital and labor of type i differs across labor types, which means that the elasticity of substitution between skilled and unskilled labor depends on capital accumulation over time. Second, endogenous technology adoption implies that regressions based on time-series variation in aggregates will confound the effects of movement along an isoquant and shifts of isoquants when measuring the change in demand.

The first marginal benefit is that a higher λ_i raises profits for any constrained firm by reducing the cost of its relative intermediate good. It is easy to show that

$$\begin{aligned}\frac{\partial \tilde{\pi}_{t+1}}{\partial \lambda_{it+1}} &= \underbrace{\frac{Y}{\alpha^{\alpha}} \frac{(1-\alpha)}{(\alpha-1)^{1-\alpha}} \tilde{C}_{Ft+1}(\cdot)^{-\alpha}}_{\frac{\partial \tilde{\pi}}{\partial \tilde{C}_F}<0} \frac{\partial \tilde{C}_{Ft+1}(\cdot)}{\partial \lambda_{it+1}} \\ &= \underbrace{-\frac{Y}{\alpha^{\alpha}} \frac{\tilde{C}_{Ft+1}(\lambda, z)^{\sigma-\alpha}}{(\alpha-1)^{-\alpha}}}_{<0} \underbrace{\frac{\mu_i P_{Git+1}(\cdot)^{\rho-\sigma}}{1-\rho} \left[r_{kt+1}^{1-\rho} - \left(\frac{w_{it+1}}{\psi_i(\lambda_i)} \right)^{1-\rho} \right]}_{\leq 0}\end{aligned}$$

The first term in the underbrace is negative because $\alpha > 1$ (recall α is the elasticity of substitution across varieties) and the second underbrace is negative as long as the firm is choosing a value $\lambda_{it+1} < \hat{\lambda}_{it+1}$. Note that this is true irrespective of whether $\rho > 1$ or $\rho < 1$.

The second marginal benefit is that a higher technology parameter installed tomorrow reduces the cost associated with any *future* choice of an even higher technology parameter, by reducing the gap between the technology installed tomorrow and the higher technology parameter to be installed in the future.

I now explain how the model generates a shift in the direction of technology adoption. To build intuition, it is useful to consider the cost of performing a task using unskilled workers and the cost of performing it using skilled workers. Recall that these costs are given by

$$\chi_i(\lambda_i, w_i) = \frac{w_i}{\psi_i(\lambda_i)}, i = s, u$$

In the initial steady state, all firms operate the same technology with $\lambda_i = \hat{\lambda}_i(w_i/r_k)$. Starting from this steady state, given the form for $\psi_i(\cdot)$ and the calibrated values for B_i, γ_i, it is easy to show that the curve $\chi_u(\cdot)$ is steeper than[26] the curve $\chi_s(\cdot)$ $\left(\text{i.e. } \left|\frac{\partial \chi_u}{\partial \lambda_u}\right| > \left|\frac{\partial \chi_s}{\partial \lambda_s}\right|\right)$. Thus, at the initial equilibrium, a given increase in λ_s generates a *smaller* drop in total costs than an equivalent increase in λ_u. Since the marginal cost of raising λ_i is the same for both types of labor $i = s, u$, at the margin, constrained firms will tend to shift λ_u out by more than λ_s. Along the initial part of the transition path, this force contributes to slower growth in the relative demand for unskilled labor, relative to skilled labor. This faster increase in skilled labor demand manifests as an increase in the skill premium. This increase is driven by a rapid increase in wages of skilled labor, relative to unskilled labor. One interpretation of this is that firms initially see larger cost declines from the introduction of industrial robots which displace unskilled workers, than they do by introducing software or other technologies which displace skilled workers. This is in line with what the literature has found for the period 1980-2000.

[26]This steeper slope reflects a combination of two forces. First, the pure labor augmenting term B_u (which is fixed over time) is larger than B_s. Second, $\gamma_s < \gamma_u$. Note that the levels of B_u and B_s cannot be independently identified from the level of q_k, so one interpretation of the higher value of B_u is that the cost of capital used to perform unskilled tasks is lower by a constant factor than the cost of capital used to perform skilled tasks.

As the wages of skilled labor rise, however, two things happen. First, the curve $\chi_s(\lambda_s, w_s)$ becomes steeper since its numerator grows, implying a greater reduction in costs from a marginal increase in λ_s. Second, since $\frac{1}{\psi_u(x_u)}$ is convex in x_u, the marginal decline in unit costs from a marginal increase in λ_u falls. This weakens the incentives to raise λ_u. Together, these forces imply that the *relative* incentives to displace skilled labor by raising λ_s increase over time, compared to the incentives to raise λ_u. Two forces prevent firms from fully closing the gap between λ_s and $\hat{\lambda}_s$ immediately. First, there is the marginal cost of raising λ_s, which is $\kappa_s Y$. Second, since the cost of using skilled labor is also convex in λ_s, raising λ_s reduces the cost savings from marginal increases in λ_s and therefore -all else equal- reduces the incentives to further raise λ_s. The model thus produces a gradual transition from technology adoption focused on displacing unskilled workers to displacing skilled workers. Note that this argument does not rely on the specific functional form assumptions I make for $\psi_i(\cdot)$ - it only requires that the function satisfy the assumptions listed below the definition of the production function 6 and that the curve $\chi_u(x)$ be steeper than $\chi_s(x)$ at all points $x \in [0,1]$. However, the parametrization I use has the convenient property that it requires only two parameter γ_s, γ_u to be disciplined.

While the argument above conveys the key intuition of my model, there are also interaction effects that complicate firms' adoption choices. To understand these, consider a rise in λ_s for a constrained firm. All else equal, such an increase will reduce the cost of the skilled intermediate good G_s. I estimate that G_s and G_u are gross substitutes, which implies that a fall in the price of G_s will reduce the firm's input of the unskilled intermediate good G_u. This, in turn, reduces the firm's incentives to raise λ_u and is a force that further pushes in the direction of raising λ_s by more than λ_u. The force works in the other direction as well: a fall in λ_u induces firms to invest less in raising λ_s as well. In the initial period, these two forces offset each other. As the skill premium rises, however, the the former dominates and the incentives to raise λ_u decline by relatively more.

Finally, I explain the role of size heterogeneity in my model. While not directly a factor in the mechanism above, heterogeneity in firm sizes induced by the productivity shocks allows the model to capture the staggered adoption of technologies, and also allows the model to capture a key feature of the data: the central role of reallocation of value added towards firms with low labor shares.

Lemma 4. *All else equal, consider two constrained firms with states $s_1 = (\lambda, z_1)$ and $s_2 = (\lambda, z_2)$ where $z_2 > z_1$. Then the policy functions for next period's choice of the capital feasibility cutoffs satisfy, for both $i = s, u$, $\lambda'_i(s_1) \leq \lambda'_i(s_2)$.*

Proof. See proof of claim 5 in section A.6.

□

Lemma 4 shows that firms with higher TFP will be more likely to adopt new technologies. The mechanism by which this translates to larger firms having lower labor shares is analogous to the one discussed in Zolas et al. (2020) and modeled in Hubmer and Restrepo (2021): with a constant fixed cost of raising the capital feasibility cutoffs by a unit, constrained firms with higher TFP have more to gain from a marginal increase in their cutoffs. This is because their higher productivity will persist from date t, when they make their choice, to date $t+1$, when they will use the technology they are investing in adopting today. All else equal, higher productivity tomorrow equates to higher marginal increases in profits tomorrow from a marginal change in the cutoff parameters made today. However, with $\alpha > 1$, these higher TFP firms are *ceteris paribus* also larger firms, with larger factor input demands and intermediate good sales. Thus, in the model, larger firms are the first to adopt newer technologies. Since these new technologies are more capital intensive, their labor shares are lower as well. As I will argue in the model validation section, the model does a decent job at matching the implied size gradient in the labor share.

6 Results

In this section, I do two things. First, in section 6.1, I validate my model by showing that it is quantitatively consistent with several non-targeted moments, including output elasticities and changes in establishment-level skilled labor to unskilled labor ratios. I also show that the model generates the decline in the labor share via reallocation of value added toward large firms, and that the median firm's labor share actually rises along the transition path, which is consistent with recent evidence on the decline of the labor share being driven by reallocation forces. I show that the model is broadly consistent with the rising concentration of sales and employment in large firms.

Second, I perform the main transition experiment I consider, which works as follows. I assume that the economy is in a steady state of the form described above in 1980. I assume that in 1980, agents learn of a continuous decline in the price of capital from its 1980 value to its 2019 value and also learn of a persistent increase in the relative supply of skilled labor between these periods. I assume that after 2020, the growth rates of both series decline (in absolute value) linearly from their 2019 value to 0 over the next 10 years. After 2030, both series are constant forever[27].

I compute the perfect foresight transition of the economy in response to the new information thus obtained by economic agents. I solve for the initial steady state in 1980 and the final steady state using the terminal values of the exogenous aggregate time series for the price of capital q_{kt} and skill supplies S_t, U_t. I then solve for the paths of wages w_{st}, w_{ut} and output Y_t such that labor markets clear at each date, consistent with the definition of the dynamic equilibrium above. I compute the model implied equivalent of the skill premium $Skillprem_t = \log w_{st} - \log w_{ut}$ and the labor share in value added[28] $lshr_t = \frac{w_{st}S_t + w_{ut}U_t}{Y_t}$.

[27]My results are robust to choosing different horizons for the decline in the growth rates between 20 and 40 years. It is important to allow for this gradual adjustment since without this, the kink in the path for the relative price of capital produces a discontinuous jump in the rental rate of capital for one period which affects firms' incentives to produce output at all prior dates.

[28]In my model, the labor share in value added is proportional to the labor share in costs since markups are constant.

6.1 Model Validation on Non-Targeted Moments

I validate the model by checking its fit on a range of non-targeted moments.

First, I note that the model-implied output elasticities as of 1980 with respect to each factor of production are consistent with estimated value-added production functions. In particular, I estimate the output elasticities with respect to each factor, $\omega_f^Y = \frac{d \log y}{d \log f}$ for $f = \ell_s, \ell_u, k$ at the steady state. Note that in the steady state, all firms choose the same value of λ, and the elasticity is invariant to the productivity of the firm up to a constant. I compute these elasticities for ℓ_s, ℓ_u, k to be 0.29, 0.41 and 0.29 respectively, consistent with the values in Demirer (2020)and in Gandhi et al. (2020)[29][30].

Second, the model performs well at matching changes in the *median* of the ratio of skilled to unskilled labor input, $\ell_{st}(s)/\ell_{ut}(s)$. Even though the aggregate ratio S_t/U_t is an input to the model, this aggregate ratio is consistent with many different underlying distributions of this ratio across firms, which may have very different median values. The ability of the model to be consistent with the change in the median is therefore a success of the model.

To the best of my knowledge, publicly available data on firm inputs of skilled and unskilled labor are not available, and I provide one estimate of the distribution of this ratio. To construct it, I draw on data from the Computer Intelligence Technology Database, maintained by Aberdeen/Harte-Hanks, which contain establishment level information on shares of white collar and blue collar employees. I discuss this dataset further in section 7.1. I impute the number of skilled and unskilled workers by combining the Harte-Hanks data with the CPS ASEC as follows. In the CPS, I assign occupations to white or blue-collar status based on definitions used by the BLS: White-collar occupations include Professional, Managerial and Technical workers, some Sales occupations, and Office Administrative Support Occupations, and blue-collar occupations are the complement. I then use the CPS to calculate, for industry k in year t, the object

$$\Pr\left(Skilled \mid WhiteCollar, kt\right) = \frac{\sum_i \omega_i \mathbf{1}\left(i \in Skilled_t \bigcap WhiteCollar_t\right)}{\sum_i \omega_i \mathbf{1}\left(i \in WhiteCollar_t\right)}$$

where ω_i is individual i's demographic weight, the sets $Skilled_t, WhiteCollar_t$ indicate individuals who are skilled and who work in white-collar jobs respectively. I repeat this for other skilled/unskilled and blue/white-collar combinations. For establishment j in industry k at date t, I impute

$$N_{skilled,jkt} = N_{WhiteC,jkt} \Pr\left(Skilled \mid WhiteC, kt\right) + N_{BlueC,jkt} \times \Pr\left(Skilled \mid BlueC, kt\right)$$

[29]Demirer (2020) estimates that the average capital elasticity across the manufacturing industries he studies is 0.25, as compared to my estimate of 0.29. He also estimates that larger firms have higher capital elasticities. This is not true in the steady state of my model, since factor ratios are invariant to heterogeneity in firm TFP z, which is the only dimension of heterogeneity that exists in the steady state. However, it is true along the transition path, since larger firms are more likely to adopt technologies that are more skill intensive.

[30]Gandhi et al. (2020) estimate gross-output production functions and compute the average ratio of the capital to the labor elasticity to be 0.4 across a variety of specifications and datasets. I compute this to be 0.42.

Moment	Data	Model
$P50\left(\frac{\ell_{s,1998}}{\ell_{u,1998}}\right)$	0.96	0.89
$P50\left(\frac{\ell_{s,2008}}{\ell_{u,2008}}\right)$	1.1	1.14

Table 2: Data from Harte-Hanks CiTDB. Skilled and Unskilled labor imputed by allocating reported white collar and blue collar employment to skilled and unskilled categories proportionate to their respective ratios in CPS industry-year bins.

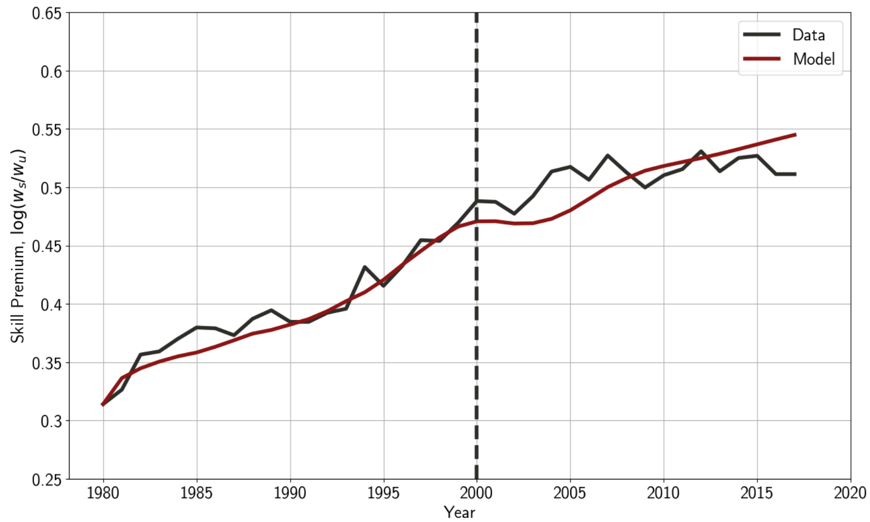
and similarly for other skilled/unskilled and blue/white collar combinations. I choose to do this in the years 1998 and 2008, a 10-year period, because I have sufficiently many observations to calculate these objects in these years[31].

Next, I show that the model is consistent with the findings of Autor et al. (2020) and Kehrig and Vincent (2021), in that the labor share decline is driven by a reallocation of value added toward large firms. Thus, the model endogenously produces "superstar" firms with large shares of total sales and low labor shares. In line with the findings of Kehrig and Vincent, superstar status is transient, since TFP is mean reverting. While the aggregate share falls by about 1.5 percentage between 1982 and 2012, the median labor share rises by about 3 percentage points. Kehrig and Vincent (2021) show that for the manufacturing sector, the median firm's labor share rose by about 3 percentage points, indicating that the model is consistent with the differential trends in the median and the aggregate labor share, and also indicating that the model's declining labor share is indeed driven by reallocation of value added toward low labor share firms.

Finally, I evaluate the model's implications for concentration of sales. In the model, due to the fixed cost, only relatively large firms adopt new technologies. The reduction in costs that comes from technology adoption allows these firms to charge lower prices and therefore acquire larger sales, implying that the model naturally features a force towards concentration of sales in large firms. In the model, the implied 4-firm concentration ratio for sales, defined as the fraction of total sales accounted for by the top 4 firms, rises from about 40% in 1982 to 47% in 2012. This increase of 7 percentage points is somewhat larger than the 5.3% found by Barkai (2020). One possible reason why the model predicts a larger sales share increase for the largest firms than in the data is because it features constant markups, whereas there is evidence that larger firms enjoy higher markups and therefore more muted increases in their sales shares as costs fall. Extending my model to include a variable elasticity of substitution structure for aggregate demand in order to account for this fact is an important extension I leave to future work.

[31] Post 2010, data on the blue and white collar employment at the establishment level is no longer available in my extract of the data.

6.2 The skill premium and the labor share



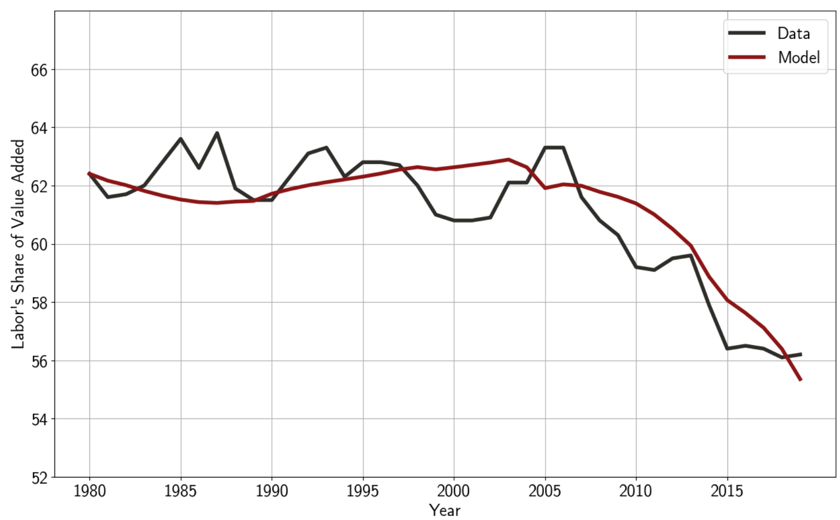


Figure 7: Model's fit for the skill premium (top) and the labor share of value added (bottom). Data on the labor share are for the business sector from the BEA-BLS Integrated National Income accounts. The model's implied share for labor in value added is given by the ratio $\frac{w_s S + w_u (H-S)}{Y}$.

Figure 7 plots the model's performance at matching the behavior of the skill premium and the labor share. The model's fit on both series is excellent, given that the calibration only targets the initial values of the two series. Towards the end of the period in 2019, the model over-predicts the data-implied level of the skill premium by about 3 percentage points (or 6% of the value of the series). The initial increase in the skill premium is driven by a combination of technical change being relatively directed at unskilled labor over this period. As the skill premium rises, firms switch toward displacing relatively skilled workers.

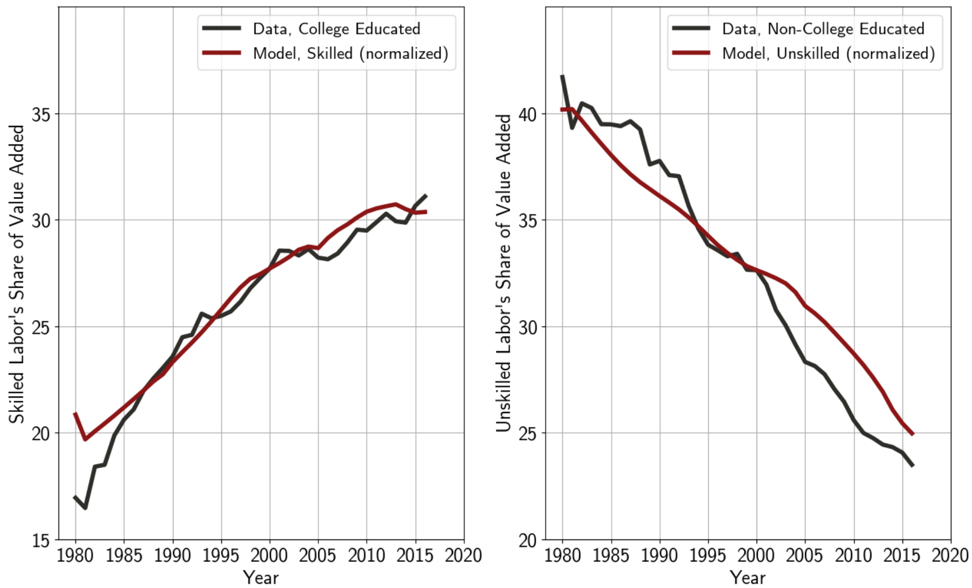
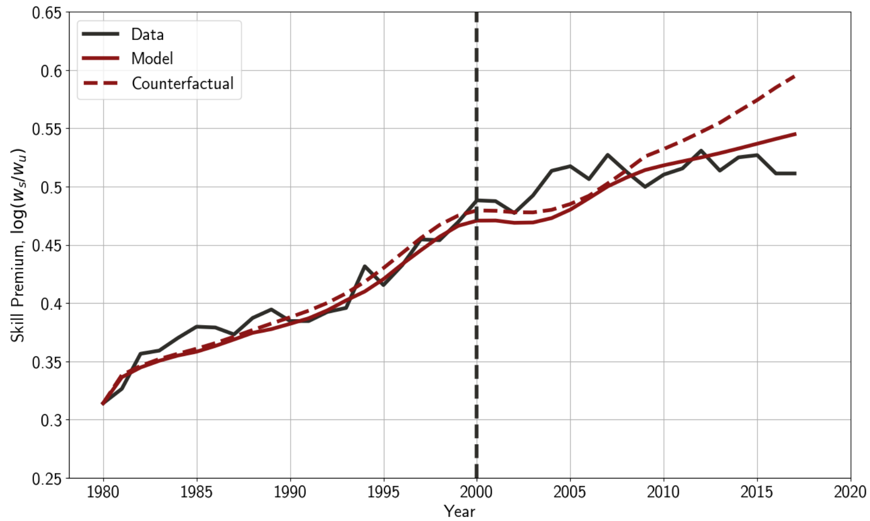


Figure 8: The skilled and unskilled labor share of value added. Data: BEA-BLS Integrated National Accounts, compensation of college/non-college workers in total value added. Model: $\frac{w_{it}\ell_{it}}{Y_t}$, renormalized to equal data's value in 2000.

6.3 Counterfactual behavior without Technology Choices



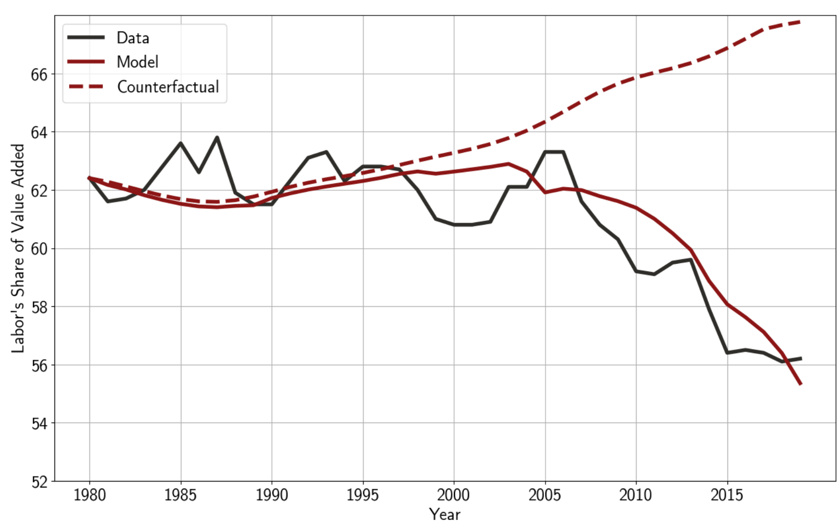


Figure 9: Model's fit for the skill premium (top) and the labor share of value added (bottom), vs a counterfactual with no technical change. Data on the labor share are for the business sector from the BEA-BLS Integrated National Income accounts. The model's implied share for labor in value added is given by the ratio $\frac{w_s S + w_u(H-S)}{Y}$.

Figure 9 plots the model's performance at matching the behavior of the skill premium and the labor share along with the counterfactual, in which firms are forced to use the same technologies over the entire period. Towards the end of the period in 2019, the counterfactual model over-predicts the data-implied level of the skill premium by about 3 percentage points (or 6% of the value of the series), and the decline in the labor share by 12 percentage points.

Figure 10 shows that this decline is driven by a slowdown in the skilled labor share, consistent with the data. I compare the model's predictions with the BEA-BLS integrated national income accounts, which provides a breakdown of the labor share by college/non-college labor. The concept of skilled labor I use in the model considers half the hours worked by workers with only some college education as a part of skilled labor, and so my model predicts a higher labor share for skilled workers in levels than in the data (and correspondingly, a lower unskilled labor share). I therefore translate the model predicted series and the counterfactual predict series so that all three series are equal in 2000, roughly the midpoint of the period I study. The figure shows that the model does an excellent job rationalizing the decline in the share of unskilled labor *and* the share of skilled labor, even though the dynamics of these series were never targeted in the calibration. The counterfactual underpredicts the decline in the unskilled share and fails to get the slowdown in the growth of the skilled labor share.

Figure 10 also illustrates that the slowdown in the demand for skilled workers is crucial to explain the behavior of the labor share. Prior to 2000, the rising share of skilled workers - driven by increases in the relative demand for skilled workers - was sufficient to offset the decline in the share of unskilled workers, leading to a roughly stable labor share. Post 2000, the slowdown in skilled labor's share meant this offsetting force was weaker, manifesting as a decline in the share of skilled workers. That the model can match the dynamics of skilled and unskilled shares separately is a success for the model that to the best of my knowledge has not be achieved by alternative models in the literature.

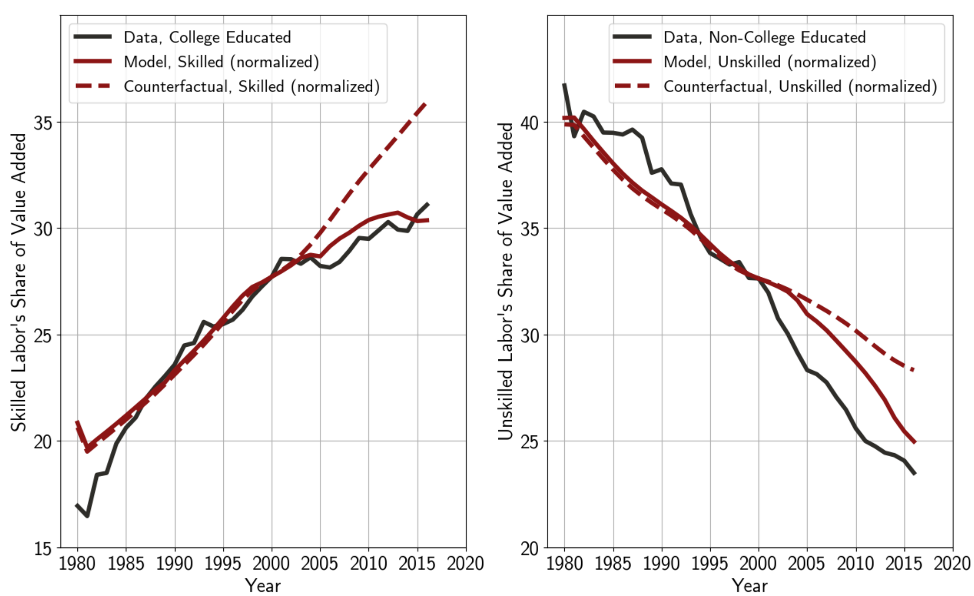


Figure 10: The skilled and unskilled labor share of value added. Data: BEA-BLS Integrated National Accounts, compensation of college/non-college workers in total value added. Model: $\frac{w_{it}\ell_{it}}{Y_t}$, renormalized to equal data's value in 2000.

7 Micro-Evidence on Technology Adoption

In this section, I provide some microeconomic evidence for my key mechanism: that a rising skill premium induces technology adoption that tends to displace relatively skilled workers. First, I conduct a case study using accountants, and show that rising exposure to dedicated accounting software reduces the rate of growth of wages for accountants. Second, I show that over time, newer technologies are increasingly associated with the kinds of tasks that relatively skilled workers perform.

7.1 A case study: Accountants

Accountants work in a prototypically routine cognitive occupation, and are on average highly skilled. And yet, modern commentators regularly worry that accountants are highly exposed to displacement by modern technologies such as AI [32]. These concerns echo the ones that accompanied the widespread deployment of commercial accounting software during the late 1990s, partly in the tax preparation industry[33]. In this section, I show that the deployment of accounting software was indeed associated with slower growth in accountant wages.

7.1.1 Data

Data on the wages of accountants comes from Census 5% samples for 1990, 2000 and 2010 from IPUMS (Ruggles et al., 2022). Accountants are defined as members of the consistent classification code `occ`=23. I residualize these wages on race, sex, age and experience categories within each year following Autor (2019). I then average over all accountants within a commuting zone using labor supply weights, defined as the product of the hours worked by an accountant and the individual demographic weight. More details on the data cleaning procedure can be found in appendix A.3.

Data on the diffusion of accounting software come from the Computer Intelligence Technology Database (CiTDB), maintained by Harte-Hanks and later by Aberdeen. The database contains establishment-level information on whether a given technology has been adopted, along with an address for the establishment and industry identifiers for firms. A technology in the dataset is a combination of a manufacturer and a specific make and model (for instance, "Microsoft Excel" and "Microsoft Word" are distinct products, but some firms report using "Microsoft Office".). In addition, it provides information regarding the use of the technology at the establishment, a description of the technology and product category information for each technology. This richness makes the dataset ideal for the study of technology adoption at a more granular level than has

[32]See, for instance, PwC (2015), Frey and Osborne (2017), Sheedy (2017), Henderson (2018) and Roose (2021).

[33]See, for instance, Zarowin (1994), Knight-Ridder/Tribune (1997), Stipe (1997), Kirkpatrick (1998) and Lee (2000). The overall thrust of these works is a mixture of optimism regarding the automation of routine tasks performed by accountants combined with a dismissal of the possibility that accounting software would displace workers. McCormally (1991) provides a description of some of these software platforms and their relative costs.

previously been explored. The CiTDB dataset has been widely used to track the diffusion of computing technologies and the internet owing to its quality[34], which owes to the fact that the database is used commercially by marketing departments of large producers and suppliers of IT and related products to identify sales opportunities. The extracts of the data I have access to nonetheless require extensive cleaning to correct for errors in the geographic and industrial classification data, the process for which I describe in appendix A.3.

After cleaning, I construct establishment-level weights which ensure the representativeness of the data with respect to the number of establishments by geography and industry as documented by the County Business Practices. That is, letting N_{cit}^{CBP} be the number of establishments in county c, 2-digit industry i and year t in the County Business Practices dataset and letting N_{cit}^{HH} be the same in the Harte-Hanks dataset, I weight all establishments j in county $c(j)$, 2-digit industry $i(j)$ and year t by the weight

$$\omega_{jt} = \frac{N_{c(j)i(j)t}^{CBP}}{N_{c(j)i(j)t}^{HH}}$$

I identify accounting software using a decision rule that uses both Harte-Hanks' own classification of technologies and the metadata that companies provide on the uses of technologies at their companies (see appendix A.3 for details). For each establishment j, I define an indicator for whether that establishment has adopted accounting software at or before date t

$$I_{jt} = \begin{cases} 1 & j \text{ ever used Accounting software at date} t' \leq t \\ 0 & \text{otherwise} \end{cases}$$

I then define the fraction of establishments in commuting zone c which have adopted accounting technologies by date t,

$$FracAdopt_{ct}^{ACCT} = \frac{\sum_{j \in c} I_{jt}\omega_{jt}}{N_{ct}}$$

where ω_{jt} is the weight associated with establishment j.

I define adoption as an indicator variable for whether or not an establishment uses a given technology. Note that this measure fails to capture the intensity with which the software is used at a given establishment: regrettably, my extract of the CiTDB data does not contain information on the "quantity" of software[35] after 2008. This is a cause for concern if firms sell

[34] Papers using subsets of the Harte-Hanks dataset to study the diffusion of computing technologies and ICT include Brynjolfsson and Hitt (2000), Bresnahan et al. (2002), Brynjolfsson and Hitt (2003), Forman et al. (2008), Forman et al. (2012), Bloom et al. (2016) and Hershbein and Kahn (2018). To the best of my knowledge, this is the first time the data have been used in economics to study the impacts of the adoption of an occupation-specific technology on the employment and wages of that occupation.

[35] Ideally, I would have liked to have known the share of PCs at the establishment on which the given software was being used, for instance.

7.1.2 The slowdown in Accountant wage growth

To study the slowdown in the wage growth for accountants, I run variants of the regression

$$\Delta_{t-10\to t}\log w_{i,ct}^{ACCT} = \beta_0 + \beta_1 \Delta_{t-10\to t} FracAdopt_{ct}^{ACCT} + \delta_s + \mathbf{x}'_{ct-10}\beta + \varepsilon_{ct}$$

where $\Delta_{t-10\to t}\log w_{i,ct}^{ACCT} \equiv \log w_{i,ct}^{ACCT} - \log w_{i,ct-10}^{ACCT}$ is the 10-year change in mean log wages of accountants who are skilled ($i = s$) or unskilled ($i = u$) in a given commuting zone and $\Delta_{t-10\to t} FracAdopt_{ct}^{ACCT}$ is the 10-year change in the fraction of establishments in a given commuting zone which have adopted computing technologies. δ_s denotes a state fixed effect. I run this regression using a stacked panel with $t = 2000, 2010$, implying that I study changes over the periods 1990-2000 and 2000-2010. I choose these periods because post 2010, the Harte-Hanks data underwent significant changes including a reclassification of many technologies and a large increase in the sampling frame, which affects the representativeness of the sample. My results thus reflect the impacts of accounting software on accountants, and do not include the impacts that the deployment of more advanced technologies like AI and cloud computing.

Table 3 shows that the effect of rising adoption are consistently negative on the growth rate of wages over the long period. The left panel shows changes in wages of skilled workers and the right panel shows changes in wages of unskilled workers. To interpret magnitudes, accountants in commuting zones at the 25th percentile of exposure growth saw their wages grow about 4.4 percentage points[36] over 10 years *slower* than accountants at the 75th percentile of exposure growth. This result is robust to the inclusion of a wide range of time-varying controls for differences across commuting zones. The effect for unskilled workers is also negative, but completely disappears once I control for average wages in the commuting zone and include controls for industrial composition.

7.2 The rising exposure of skilled occupations to technical change

I use the method developed by Webb (2020) to document that the exposure of occupations to new inventions has shifted toward occupations that are relatively high-skilled. The idea of this method is to identify tasks as verb-noun pairs, where the noun is the direct object of the verb, from unstructured text descriptions of occupational tasks and from the descriptions of technologies. For instance, consider a patent titled "Method for generating a bitmap." My algorithm extracts the verb-noun pair (generating, bitmap). I lemmatize the verb to ensure that different tenses or cases associated with a verb are all mapped to the root of the verb (hence translating "generating" into its verb form "to generate"). I lemmatize the noun as well. Since patent titles use relatively specific nouns ("bitmap") and occupational tasks use relatively general nouns ("image"), I use WordNet, an English Language corpus which groups nouns into hierarchies of concepts. At a given level of this hierarchy, conceptual categories are mutually exclusive, allowing me to translate nouns to a

[36] $-0.0438 = -0.0996 \times (0.57 - 0.308)$, where -0.0996 is the coefficient on adoption growth and the 25th and 75th percentiles of changes in the exposure growth measure are 30.8 percentage points and 57 percentage points respectively.

	Changes in Skilled Wages			Changes in Unskilled Wages		
	$\Delta w_{s,ct}$	$\Delta w_{s,ct}$	$\Delta w_{s,ct}$	$\Delta w_{u,ct}$	$\Delta w_{u,ct}$	$\Delta w_{u,ct}$
$\Delta FracAdopt_{ct}^{ACCT}$	-0.162***	-0.153***	-0.0996***	-0.0391**	-0.0366**	-0.0244
	(0.0359)	(0.0356)	(0.0338)	(0.0166)	(0.0170)	(0.0176)
State FE	Y	Y	Y	Y	Y	Y
Race Comp. Controls	Y	Y	Y	Y	Y	Y
Age Controls	N	Y	Y	N	Y	Y
Income, Industry Controls	N	N	Y	N	N	Y
N	1,386	1,386	1,386	1,386	1,386	1,386

Table 3: An observation is a 10-year change in skilled wage growth and a 10-year change in cumulative adoption rates as defined above. Observations weighted by commuting zone population in initial period. Data on wage growth from ACS 1990, 2000, 2010. Data on rising adoption of Accounting technologies from Computer Intelligence Technology Database (CiTDB). Race controls include indicators for share of Black and share of other racial groups (with whites as the base category). Age controls include indicators for fraction of population below 18 and fraction above 65. Income, Industry controls include per capita income, employment-population ratio, share employed in manufacturing and share employed in services industries. All regressions include state fixed effects. Standard errors clustered at the commuting zone level in parentheses. *, **, *** indicate statistical significance at 0.1, 0.05 and 0.01% respectively.

common level of generality. I follow Webb (2020) and use WordNet level 3, checking that using level 4 does not drastically alter results. Appendix 2 describes this process in more detail.

I start with data on patents from the USPTO's PatentsView database. From the titles of these patents I extract the set of all (noun,verb) pairs associated with a patent. I assume that these bigrams individually identify a task performed by the invention[37]. I next perform a similar exercise with the definitions of occupations in the Department of Labor's O*NET Database, which contains occupation-level information on the tasks performed by workers on that occupation. For each occupation, I compute, year-by-year, the cosine similarity between the tasks identified in the Tasks section of the occupation's O*NET entry and the tasks associated with all inventions in a given year. This exercise provides a measure of exposure for each occupation to technological changes occurring in a given year[38].

Figure 11 shows that the rise in the exposure metric has been highest for occupations in the top quintile of the distribution of skill intensity. My results complement those of Kogan et al. (2021),

[37] I reduce the generality of all nouns using the lexical database WordNet's conceptual hierarchy for each noun. This is important since at the level of generality associated with patent titles or occupational task descriptions, the set of exactly overlapping (noun, verb) pairs is virtually empty. This is because patents can be extremely specific in their titles but refer to technologies with much wider applicability. A similar argument holds for occupational titles.

[38] For validation of my measure, figure 16 shows the exposure measure for occupations of different levels of routineness to validate the exposure measure. It is clear that between 1980 and 2000, technological changes were overwhelmingly directed at performing the kinds of tasks that routine manual employees performed, which is exactly in line with the Routine-biased technological change literature. Over time, as the technologies necessary to automate these tasks mature and eventually are adopted into the workforce, the effects on the labor market show up in the form of job polarization and wage polarization. However, importantly for the purpose of this paper, over the period starting roughly in the mid 1990s, there is a steady increase in the exposure of nonroutine and cognitive occupations, which continues throughout the 2000s. This increase occurs roughly at the same time as the initial slowdown of the skill premium, suggesting that the two may be linked.

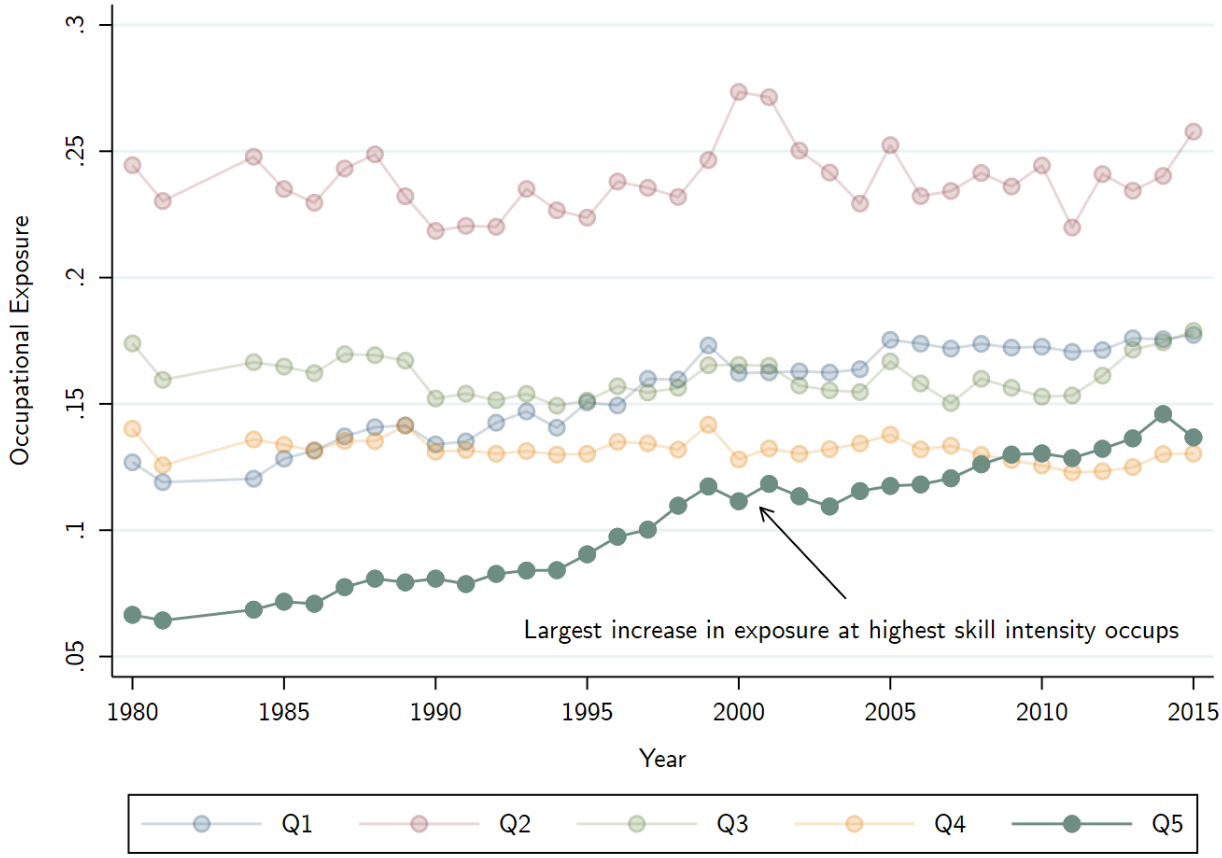


Figure 11: Exposure to technical change of Occupations in different quintiles of the skill intensity distribution (skill intensity measured by mean value of ℓ_s/ℓ_u in 2000, weighted by total hours worked in that occupation.). Exposure is measured by the cosine similarity between overlapping sets of bigrams representing tasks in O*NET Occupational Tasks and USPTO Patent Titles.

who use a different natural language processing approach to quantify the rising exposure to drastic innovations of different groups, and conclude that there has been a large increase in the exposure of relatively well-paid, experienced and skilled workers. My results are also robust to constructing the cosine similarity measure using sets of just verbs, instead of bigrams.

8 Conclusion

Simultaneously rationalizing the behavior of the skill premium and the labor share of income is a challenge for most models explaining skill-driven inequality via the relationship between skills and technological change. I construct a simple model in which endogenous technology adoption in response to an initial rise in the skill premium causes a slowdown in the growth of skilled labor demand. This slowdown endogenously slows growth in the skill premium and growth in skilled labor's share of income, allowing the model to rationalize the fall in the labor share.

There are several substantial extensions possible for this paper. First, including an endogenous supply for skilled labor would allow for counterfactuals particularly relevant to contemporary settings, such as whether subsidizing college is a valuable policy measure. Second, in this paper, there is no possibility for skilled labor to displace for unskilled labor at the tasks it performs - all substitution between skilled labor and unskilled labor occurs through movements along an isoquant. Adding this margin would allow the model to explore the implications of downskilling: as technologies displace skilled labor gradually, skilled workers would start encroaching on tasks that were previously the domain of unskilled labor. Finally, this model features two skill types. An extension to multiple skill types would allow the model to speak to issues such as labor market polarization.

References

Acemoglu, Daron, "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality," *The Quarterly Journal of Economics*, November 1998, *113* (4), 1055–1089.

— , "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature*, March 2002, *40* (1), 7–72.

— , "When Does Labor Scarcity Encourage Innovation?," *Journal of Political Economy*, 2010, *118* (6), 1037–1078.

— **and David Autor**, "Chapter 12 - Skills, Tasks and Technologies: Implications for Employment and Earnings," in David Card and Orley Ashenfelter, eds., *Handbook of Labor Economics*, Vol. 4 of Handbook of Labor Economics, Elsevier, 2011, pp. 1043–1171.

— **and Pascual Restrepo**, "Low-Skill and High-Skill Automation," *Journal of Human Capital*, 2018, *12* (2), 204 – 232.

— **and** — , "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, June 2018, *108* (6), 1488–1542.

— **and** — , "Robots and jobs: Evidence from US labor markets," *Journal of Political Economy*, 2020, *128* (6), 2188–2244.

— **and** — , "Unpacking Skill Bias: Automation and New Tasks," *AEA Papers and Proceedings*, May 2020, *110*, 356–61.

— **and** — , "Tasks, Automation, and the Rise in US Wage Inequality," Working Paper 28920, National Bureau of Economic Research February 2022.

— **, Claire Lelarge, and Pascual Restrepo**, "Competing with Robots: Firm-Level Evidence from France," *AEA Papers and Proceedings*, May 2020, *110*, 383–88.

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad, "The Skill Complementarity of Broadband Internet," *The Quarterly Journal of Economics*, 2015, *130* (4), 1781–1824.

Allen, Robert C, *The British industrial revolution in global perspective*, Cambridge University Press, 2009.

Altonji, Joseph G., Erica Blom, and Costas Meghir, "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers," *Annual Review of Economics*, July 2012, *4* (1), 185–223.

Atalay, Enghin, Phai Phongthiengtham, Sebastian Sotelo, and Daniel Tannenbaum, "New technologies and the labor market," *Journal of Monetary Economics*, 2018, *97* (C), 48–67.

Aum, Sangmin, "The rise of software and skill demand reversal," mimeo, Washington University in St. Louis March 2018.

— **and Yeongseok Shin**, "Is Software Eating the World?," mimeo, Washington University in St. Louis March 2022.

— **and Yongseok Shin**, ""Why Is the Labor Share Declining?"," *Federal Reserve Bank of St. Louis Review*, October 2020, pp. 413–28.

Autor, David, *Comment on Recent Flattening in the Higher Education Wage Premium: Polarization, Skill Downgrading, or Both?*, University of Chicago Press, December

— **, Caroline Chin, Anna M Salomons, and Bryan Seegmiller**, "New Frontiers: The Origins and Content of New Work, 1940-2018," Working Paper 30389, National Bureau of Economic Research August 2022.

— **, David Dorn, Lawrence Katz, Christina Patterson, and John van Reenen**, "The Fall of the Labor Share and the Rise of Superstar Firms*," *The Quarterly Journal of Economics*, 2020, *135* (2), 645–709.

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

— **, — , and Richard Murnane**, "Upstairs, Downstairs: Computers and Skills on Two Floors of a Large Bank," *ILR Review*, 2002, *55* (3), 432–447.

Autor, David H., "The "task approach" to labor markets : an overview," *Journal for Labour Market Research*, 2013, *46* (3), 185–199.

— , "Why Are There Still So Many Jobs? The History and Future of Workplace Automation," *Journal of Economic Perspectives*, September 2015, *29* (3), 3–30.

— , "Work of the Past, Work of the Future," *AEA Papers and Proceedings*, May 2019, *109*, 1–32.

— **and David Dorn**, "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, August 2013, *103* (5), 1553–97.

— **, Lawrence F. Katz, and Alan B. Krueger**, "Computing Inequality: Have Computers Changed the Labor Market?*," *The Quarterly Journal of Economics*, 11 1998, *113* (4), 1169–1213.

— **, — , and Melissa S. Kearney**, "The Polarization of the U.S. Labor Market," *American Economic Review*, May 2006, *96* (2), 189–194.

— **, — , and —** , "Trends in U.S. Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics*, 2008, *90* (2), 300–323. Data and Replication Files.

Barany, Zsofia and Christian Siegel, "Job Polarization and Structural Change," *American Economic Journal: Macroeconomics*, 2018, *10* (1), 57–89.

Barkai, Simcha, "Declining Labor and Capital Shares," *The Journal of Finance*, 2020, *75* (5), 2421–2463.

Bartel, Ann, Casey Ichniowski, and Kathryn Shaw, "How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills," *The Quarterly Journal of Economics*, 11 2007, *122* (4), 1721–1758.

Beaudry, Paul, David A. Green, and Benjamin M. Sand, "The Declining Fortunes of the Young since 2000," *American Economic Review*, May 2014, *104* (5), 381–86.

— **, — , and —** , "The Great Reversal in the Demand for Skill and Cognitive Tasks," *Journal of Labor Economics*, 2016, *34* (S1), S199–S247.

Berlingieri, Giuseppe, Filippo Boeri, Danial Lashkari, and Jonathan Vogel, "Capital-Skill Complementarity in Firms and in the Aggregate Economy," Working Paper, UCLA May 2022.

Berman, Eli, John Bound, and Zvi Griliches, "Changes in the demand for skilled labor within US manufacturing: evidence from the annual survey of manufactures," *The quarterly journal of economics*, 1994, *109* (2), 367–397.

Bloom, Nicholas, Mirko Draca, and John Van Reenen, "Trade-induced Technical Change? The impact of Chinese imports on innovation, IT and productivity," *The Review of Economic Studies*, 2016, *83* (1), 87–117.

Boehm, Michael J., "Has Job Polarization Squeezed the Middle Class? Evidence from the Allocation of Talents," CEP Discussion Papers, Centre for Economic Performance, LSE 2013.

Bound, John and George Johnson, "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations," *The American Economic Review*, 1992, *82* (3), 371–392.

Bresnahan, Timothy, Erik Brynjolfsson, and Lorin M. Hitt, "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence," *The Quarterly Journal of Economics*, 2002, *117* (1), 339–376.

Brynjolfsson, Erik and Lorin M. Hitt, "Beyond Computation: Information Technology, Organizational Transformation and Business Performance," *Journal of Economic Perspectives*, December 2000, *14* (4), 23–48.

— **and Lorin M Hitt**, "Computing productivity: Firm-level evidence," *Review of Economics and Statistics*, 2003, *85* (4), 793–808.

Castex, Gonzalo and Evgenia Dechter, "The Changing Roles of Education and Ability in Wage Determination," *Journal of Labor Economics*, 2014, *32* (4), 685 – 710.

— **, Sang-Wook (Stanley) Cho, and Evgenia Dechter**, "The decline in capital-skill complementarity," *Journal of Economic Dynamics and Control*, 2022, *138*, 104363.

Cheng, Hong, Lukasz A. Drozd, Rahul Giri, Mathieu Taschereau-Dumouchel, and Junjie Xia, "The Future of Labor: Automation and the Labor Share in the Second Machine Age," Working Papers 20-11, Federal Reserve Bank of Philadelphia March 2021.

Deming, David J and Kadeem Noray, "Earnings Dynamics, Changing Job Skills, and STEM Careers*," *The Quarterly Journal of Economics*, 06 2020, *135* (4), 1965–2005.

Demirer, Mert, "Production Function Estimation with Factor-Augmenting Technology: An Application to Markups," Mimeo, MIT January 2020.

DiCecio, Riccardo, "Sticky wages and sectoral labor comovement," *Journal of Economic Dynamics and Control*, 2009, *33* (3), 538–553.

Dillender, Marcus and Eliza Forsythe, "Computerization of White Collar Jobs," Working Paper 29866, National Bureau of Economic Research March 2022.

Dinlersoz, Emin and Zoltàn Wolf, "Automation, Labor Share, and Productivity: Plant-Level Evidence from U.S. Manufacturing," Working Papers, U.S. Census Bureau, Center for Economic Studies 2018.

Dorn, David, "Essays on Inequality, Spatial Interaction, and the Demand for Skills," PhD Dissertation 3613, University of St. Gallen 2009.

Eden, Maya and Paul Gaggl, "Capital Composition and the Declining Labor Share," mimeo November 2019.

Eisfeldt, Andrea L, Antonio Falato, and Mindy Z Xiaolan, "Human Capitalists," Working Paper 28815, National Bureau of Economic Research May 2021.

Elsby, Michael WL, Bart Hobijn, and Ayşegül Şahin, ""The decline of the US labor share"," *"Brookings Papers on Economic Activity"*, 2013, *2013* (2), 1–63.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, and Michael Westberry, "Integrated Public Use Microdata Series, Current Population Survey: Version 9.0," 2021.

Forman, Chris, Avi Goldfarb, and Shane Greenstein, "Understanding the inputs into innovation: Do cities substitute for internal firm resources?," *Journal of Economics & Management Strategy*, 2008, *17* (2), 295–316.

— **,** — **, and** — , "The Internet and Local Wages: A Puzzle," *The American Economic Review*, 2012, *102* (1), 556–575.

Frey, Carl Benedikt and Michael A Osborne, "The future of employment: How susceptible are jobs to computerisation?," *Technological forecasting and social change*, 2017, *114*, 254–280.

Gandhi, Amit, Salvador Navarro, and David A. Rivers, "On the Identification of Gross Output Production Functions," *Journal of Political Economy*, 2020, *128* (8), 2973–3016.

Goldin, Claudia and Lawrence F. Katz, "The Origins of Technology-Skill Complementarity," *The Quarterly Journal of Economics*, 1998, *113* (3), 693–732.

Gollin, Douglas, "Getting Income Shares Right," *Journal of Political Economy*, April 2002, *110* (2), 458–474.

Goos, Maarten, Alan Manning, and Anna Salomons, "Job Polarization in Europe," *American Economic Review*, May 2009, *99* (2), 58–63.

_ , _ , and _ , "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring," *American Economic Review*, August 2014, *104* (8), 2509–26.

_ and _ , "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *The Review of Economics and Statistics*, 2007, *89* (1), 118–133.

Greenwood, Jeremy, Zvi Hercowitz, and Per Krusell, "Long-Run Implications of Investment-Specific Technological Change," *The American Economic Review*, 1997, *87* (3), 342–362.

Grogger, Jeff and Eric Eide, "Changes in college skills and the rise in the college wage premium," *Journal of Human Resources*, 1995, pp. 280–310.

Grossman, Gene M. and Ezra Oberfield, "The Elusive Explanation for the Declining Labor Share," *Annual Review of Economics*, 2022, *14* (1), 93–124.

Gutiérrez, Germán and Sophie Piton, "Revisiting the Global Decline of the (Non-housing) Labor Share," *American Economic Review: Insights*, September 2020, *2* (3), 321–38.

Habakkuk, Hrothgar John, *American and British technology in the nineteenth century: the search for labour saving inventions*, Cambridge University Press, 1962.

Havranek, Tomas, Zuzana Irsova, Lubica Laslopova, and Olesia Zeynalova, "The Elasticity of Substitution between Skilled and Unskilled Labor: A Meta-Analysis," MPRA Paper 102598, University Library of Munich, Germany August 2020.

He, Hui and Zheng Liu, "Investment-specific technological change, skill accumulation, and wage inequality," *Review of Economic Dynamics*, 2008, *11* (2), 314–334.

Hémous, David and Morten Olsen, "The Rise of the Machines: Automation, Horizontal Innovation, and Income Inequality," *American Economic Journal: Macroeconomics*, January 2022, *14* (1), 179–223.

Henderson, Rebecca, *Comment on 'Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics'*, University of Chicago Press, January

Hendricks, Lutz and Oksana Leukhina, "The Return to College: Selection Bias and Dropout Risk," Working Paper Series 4733, CESifo April 2014.

_ and _ , "How risky is college investment?," *Review of Economic Dynamics*, 2017, *26*, 140–163.

Hershbein, Brad and Lisa B. Kahn, "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings," *American Economic Review*, July 2018, *108* (7), 1737–72.

Hicks, J. R., "Marginal Productivity and the Principle of Variation," *Economica*, 1932, (35), 79–88.

Hoffmann, Florian, David S. Lee, and Thomas Lemieux, "Growing Income Inequality in the United States and Other Advanced Economies," *Journal of Economic Perspectives*, November 2020, *34* (4), 52–78.

Holmes, Thomas J. and Matthew F. Mitchell, "A theory of factor allocation and plant size," *The RAND Journal of Economics*, 2008, *39* (2), 329–351.

Hubmer, Joachim and Pascual Restrepo, "Not a Typical Firm The Joint Dynamics of Firms, Labor Shares, and Capital-Labor Substitution," Working Paper 28579, National Bureau of Economic Research March 2021.

Hulten, Charles R., "Growth Accounting When Technical Change is Embodied in Capital," *The American Economic Review*, 1992, *82* (4), 964–980.

Humlum, Anders, "Robot Adoption and Labor Market Dynamics," Mimeo, University of Chicago November 2021.

Jiang, Wei, Yuehua Tang, Rachel (Jiqiu) Xiao, and Vincent Yao, "Surviving the Fintech Disruption," Working Paper 28668, National Bureau of Economic Research April 2021.

Johnson, George E., "Changes in Earnings Inequality: The Role of Demand Shifts," *Journal of Economic Perspectives*, June 1997, *11* (2), 41–54.

Juhn, Chinhui, Kevin M Murphy, and Brooks Pierce, "Wage inequality and the rise in returns to skill," *Journal of Political Economy*, 1993, *101* (3), 410–442.

Kaiser, Ulrich, "New Technologies And The Demand For Heterogeneous Labor: Firm-Level Evidence For The German Business-Related Service Sector," *Economics of Innovation and New Technology*, 2000, *9* (5), 465–486.

Karabarbounis, Loukas and Brent Neiman, "The Global Decline of the Labor Share*," *The Quarterly Journal of Economics*, 10 2013, *129* (1), 61–103.

Katz, Lawrence F. and David H. Autor, "Chapter 26 - Changes in the Wage Structure and Earnings Inequality," in Orley C. Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3 of Handbook of Labor Economics, Elsevier, 1999, pp. 1463–1555.

_ **and Kevin M. Murphy**, "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 1992, *107* (1), 35–78.

Kehrig, Matthias and Nicolas Vincent, "The Micro-Level Anatomy of the Labor Share Decline," *The Quarterly Journal of Economics*, 03 2021, *136* (2), 1031–1087.

Kirkpatrick, John, "Legal, accounting software give public a leg up Professional praise some applications but warn that experts can't be totally replaced with computers: [HOME FINAL Edition]," Jun 30 1998. Copyright - Copyright DALLAS MORNING NEWS Jun 30, 1998; Last updated - 2013-09-20.

Knight-Ridder/Tribune, "Tax-return software hard on accountants' businesses: [Final Edition]," Mar 09 1997. Copyright - Copyright The Herald Mar 9, 1997; Last updated - 2013-08-22.

Knoblach, Michael, Martin Roessler, and Patrick Zwerschke, ""The Elasticity of Substitution Between Capital and Labour in the US Economy: A Meta-Regression Analysis"," *Oxford Bulletin of Economics and Statistics*, 2020, *82* (1), 62–82.

Kogan, Leonid, Dimitris Papanikolaou, Lawrence D. W Schmidt, and Bryan Seegmiller, "Technology-Skill Complementarity and Labor Displacement: Evidence from Linking Two Centuries of Patents with Occupations," Working Paper 29552, National Bureau of Economic Research December 2021.

Koh, Dongya, RaÃŒl SantaeulÃ lia-Llopis, and Yu Zheng, "Labor Share Decline and Intellectual Property Products Capital," *Econometrica*, 2020, *88* (6), 2609–2628.

Kopytov, Alexandr, Nikolai Roussanov, and Mathieu Taschereau-Dumouchel, "Short-run pain, long-run gain? Recessions and technological transformation," *Journal of Monetary Economics*, 2018, *97*, 29–44.

Korinek, Anton, Martin Schindler, and Joseph Stiglitz, "Technological Progress, Artificial Intelligence, and Inclusive Growth," IMF Working Paper, Institute for Capacity Development, International Monetary Fund June 2021.

Krusell, Per, Lee E. Ohanian, José-VÃctor Ríos-Rull, and Giovanni L. Violante, "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, 2000, *68* (5), 1029–1053.

Lee, Berton, "The computer strikes back - at accountants," *Accounting Today*, Sep 2000, *14* (16), 7–7,35. Copyright - Copyright Faulkner & Gray, Inc. Sep 4-Sep 24, 2000; Last updated - 2021-09-10; SubjectsTermNotLitGenreText - United States–US.

Lee, Yoonsoo and Toshihiko Mukoyama, "Productivity and employment dynamics of US manufacturing plants," *Economics Letters*, 2015, *136*, 190–193.

Lemieux, Thomas, "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?," *American Economic Review*, June 2006, *96* (3), 461–498.

Leukhina, Oksana and Joseph McGillicuddy, "What's behind Rising Returns to High-Quality College Education?," Regional Economist, Federal Reserve Bank of St. Louis October 2019.

Lindenlaub, Ilse, "Sorting Multidimensional Types: Theory and Application," *The Review of Economic Studies*, 01 2017, *84* (2), 718–789.

Loecker, Jan De, Jan Eeckhout, and Gabriel Unger, "The rise of market power and the macroeconomic implications," *The Quarterly Journal of Economics*, 2020, *135* (2), 561–644.

Lohr, Steve, "Economists Pin More Blame on Tech for Rising Inequality," *The New York Times*, January 2022.

Maliar, Lilia, Serguei Maliar, and Inna Tsener, "Capital-Skill Complementarity and Inequality: Twenty Years After," DP 15228, Centre for Economic Policy Research August 2020.

Markoff, John, "Armies of Expensive Lawyers, Replaced by Cheaper Software," *The New York Times*, March 2011.

McCormally, Kevin, "Taxes: Software to the Rescue," *Changing Times (1986-1991)*, 02 1991, *45* (2), 56. Copyright - Copyright Kiplinger Washington Editors Feb 1991; Last updated - 2021-09-11.

Michaels, Guy, Ashwini Natraj, and John Van Reenen, "Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years," *The Review of Economics and Statistics*, March 2020, *96* (1), 60–77.

Milgrom, Paul and Chris Shannon, "Monotone Comparative Statics," *Econometrica*, 1994, *62* (1), 157–80.

Mishel, Lawrence, Heidi Shierholz, and John Schmitt, "Don't Blame the Robots: Assessing the Job Polarization Explanation of Growing Wage Inequality," Working Paper, Economic Policy Institute and Center for Economic and Policy Research November 2013.

Moll, Benjamin, Lukasz Rachel, and Pascual Restrepo, "Uneven Growth: Automation's Impact on Income and Wealth Inequality," *Econometrica*, Forthcoming.

Oberfield, Ezra and Devesh Raval, "Micro Data and Macro Technology," *Econometrica*, 2021, *89* (2), 703–732.

Ohanian, Lee E, Musa Orak, and Shihan Shen, "Revisiting Capital-Skill Complementarity, Inequality, and Labor Share," Working Paper 28747, National Bureau of Economic Research October 2022.

PwC, "A Smart Move: Future-proofing Australia's workforce by growing skills in science, technology, engineering and maths (STEM)," Report, PwC April 2015.

Qureshi, Zia, "Technology and the future of growth: Challenges of change," *Brookings Institute Report, Feb*, 2020, *25*, 2020.

Rognlie, Matthew, "Deciphering the Fall and Rise in the Net Capital Share: Accumulation or Scarcity?," *Brookings Papers on Economic Activity*, 2015, *46* (1 (Spring)), 1–69.

Roose, Kevin, "The Robots are coming for Phil in Accounting," *The New York Times*, March 2021.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Megan Schouweiler, and Matthew Sobek, "Integrated Public Use Microdata Series USA: Version 12.0," 2022.

Sheedy, Chris, "What CPAs need to do to survive the automation revolution," *Journal of Accountancy*, June 2017.

Spence, Michael, "Job Market Signaling," *The Quarterly Journal of Economics*, 1973, *87* (3), 355–374.

Spitz-Oener, Alexandra, "Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure," *Journal of Labor Economics*, 2006, *24* (2), 235–270.

Stipe, Suzanne E., "Accountants survive by adapting their skills Competing against accounting software, accountants now are more than bookkeepers," *Baltimore Business Journal*, Oct 07 1997, *15* (20), 17. Copyright - Copyright American City Business Journals Oct 7, 1997; Last updated - 2021-09-13.

Valletta, Robert G., "Recent Flattening in the Higher Education Wage Premium: Polarization, Skill Downgrading, or Both?," in "Education, Skills, and Technical Change: Implications for Future US GDP Growth," University of Chicago Press, January 2018, pp. 313–342.

Violante, Gianluca, "Capital-Skill Complementarity and Inequality," August 2022.

Webb, Michael, "The Impact of Artificial Intelligence on the Labor Market," 2020. mimeo.

Weitzman, Martin L., "On the Welfare Significance of National Product in a Dynamic Economy*," *The Quarterly Journal of Economics*, 02 1976, *90* (1), 156–162.

Zarowin, Stanley, "CPA 2000: What's ahead for accounting software," *Journal of Accountancy*, 03 1994, *177* (3), 54. Copyright - Copyright American Institute of Certified Public Accountants Mar 1994; Last updated - 2021-09-09; CODEN - JACYAD; SubjectsTermNotLitGenreText - United States–US.

Zeira, Joseph, "Workers, Machines, and Economic Growth," *The Quarterly Journal of Economics*, 1998, *113* (4), 1091–1117.

Zolas, Nikolas, Zachary Kroff, Erik Brynjolfsson, Kristina McElheran, David N Beede, Cathy Buffington, Nathan Goldschlag, Lucia Foster, and Emin Dinlersoz, "Advanced Technologies Adoption and Use by U.S. Firms: Evidence from the Annual Business Survey," Working Paper 28290, National Bureau of Economic Research December 2020.

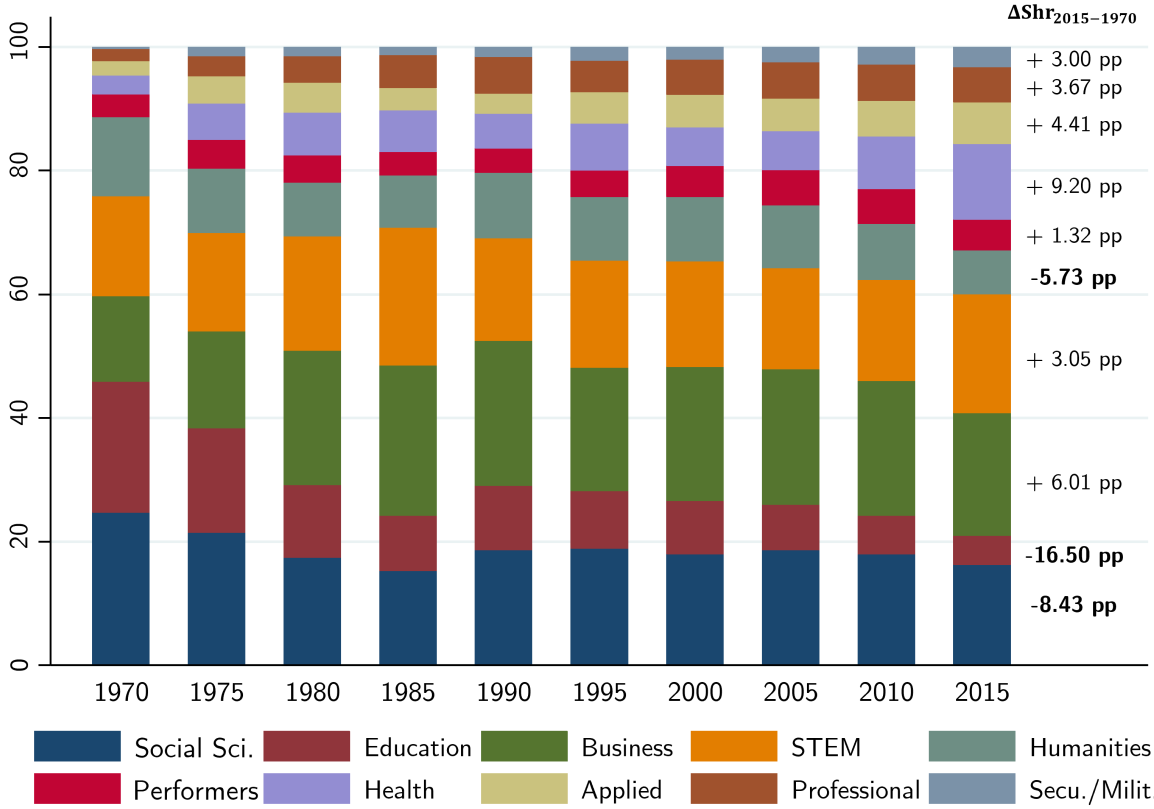


Figure 12: Share of a given group of majors awarded out of all degrees awarded in the academic year beginning in the year indicated (so the column labelled 1970 contains shares of majors in the total number of degrees awarded in academic year 1970-1971). Data from the National Center for Education Statistics (various years), Undergraduate Retention and Graduation Rates. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. The original data can be accessed here. I aggregate some degree categories for legibility.

A Appendix

A.1 Alternative Explanations for the slowdown in the skill premium

I consider three prominent alternative hypotheses that might account for the slowdown in the skill premium other than technology driven displacement: changes in the distribution of degrees earned by students, selection into who attends college and composition effects across occupations and industries.

A.1.1 Changes in the distribution of degrees earned by students

It is possible that new students entering colleges are pursuing less marketable majors and thus acquiring less valuable human capital in college. In figure 12, I show that to the contrary, the distribution of degrees awarded has remained reasonably stable over the last 45 years. The largest changes are the declines in education, social sciences and humanities and the increases in business

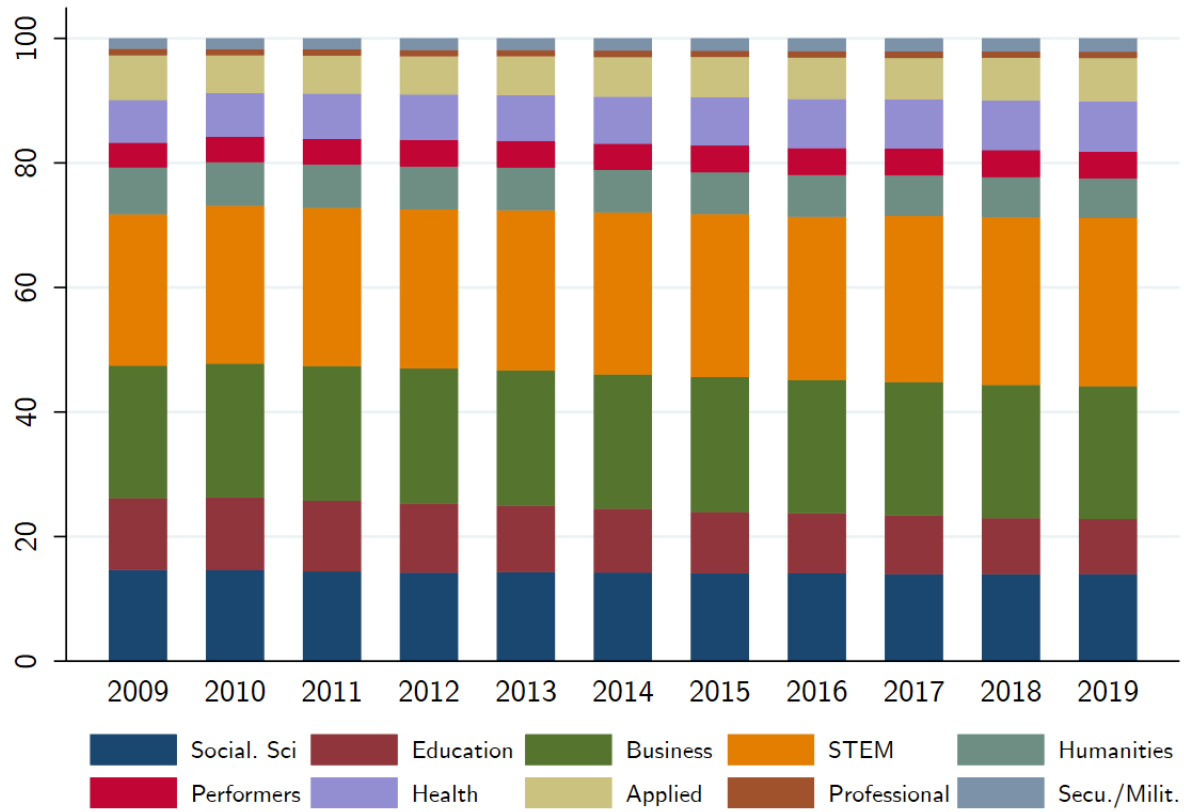


Figure 13: Distribution of degrees held by the college-graduate population in each given year. Data from ACS 2009-2019, aggregated using demographic weights.

and healthcare occupation degrees[39]. This pattern of changes suggests, if anything, that students are more likely to pursue relatively *more* marketable majors today[40]. Data on the distribution of degrees held in the population is scarce before 2009, when the ACS started collecting this information. Between 2009 and 2019, an important period for the slowdown in the skill premium, there was essentially no change in this distribution (see figure 13).

A.1.2 Selection on ability in the composition of the skilled population

If attendance at college is expensive and there is a unidimensional attribute a which raises the private benefits of attendance, only individuals with sufficiently high values of a choose to go to college. A rising college population then implies that the threshold value of a above which attendance is optimal is declining. If wages are also increasing in this attribute[41], then it is possible

[39] My results are consistent with patterns documented by Altonji et al. (2012), who show the declining trend in education majors occurs for both genders.

[40] Grogger and Eide (1995) use data from the National Longitudinal Study of the High School Class of 1972 to show that this trend of college graduates choosing degrees with higher premia can account for about 25% of the increase in the skill premium for men in the 1980s.

[41] This could include a measure of ability which affects private costs of college attendance, in line with signaling models of higher education (Spence, 1973), or directly based on heterogeneity in learning ability that interacts with the

	(1) Clerical/ Retail	(2) Mgr., Prof., Tech.	(3) Operatives	(4) Prodn. Workers	(5) Service Workers	(6) Transp., Mech., Crafts.
Time Trend	0.00625***	0.00882***	0.00131	0.00119	0.00378***	0.00292***
	(0.000702)	(0.00137)	(0.00182)	(0.00106)	(0.000864)	(0.000808)
$\log \frac{\ell_s}{\ell_u}$	-0.0679	-0.160*	0.0563	0.0911	0.0210	0.0165
	(0.0620)	(0.0917)	(0.0603)	(0.0580)	(0.0431)	(0.0407)
N	41	41	41	41	41	41
p-value	0.000	0.000	0.003	0.005	0.000	0.000

Table 4: Regression results from running Katz-Murphy regressions separately by industry. Robust standard errors in parentheses. The final row shows the p-value from a Wald test of a structural break in the data in 2000.

that the skill premium could decline via a fall in the average value of the attribute.

I argue this is unlikely to contribute substantially to the skill premium's slowdown for three reasons.

First, note that this form of selection would also reduce the average wage for the unskilled population. This is because the worker at the margin of acquiring skills is simultaneously less skilled than the mean skilled worker and more skilled than the mean unskilled worker. The net effect of selection is thus ambiguous. With multidimensional attributes governing selection, the impacts of selection on the average wages of the two groups is much more difficult to characterize (see, eg Lindenlaub (2017)).

Second, there has been a substantial increase in college costs over the time period considered. In a model of one-dimensional selection, this would (all else equal) raise the threshold level of a. An increase in the supply of skilled labor would then have to be rationalized by a *rising* value of the skill premium, rather than a slowdown.

Finally, there is direct evidence that suggests that the average ability of college graduates has risen relative to undergraduate students. Structural models of college attendance (Hendricks and Leukhina, 2014, 2017) show that students who attend college are, more often than not, well-informed about their own higher ability, which contributes to their ability to complete college and attain the benefits of the skill premium.

A.1.3 Compositional shifts across occupations and industries

Growth in the skill premium can decline if there is a reallocation of workers towards occupations or industries in which the skill premium is lower. The US has seen substantial shifts in its occupational and industrial structure. I argue this force by itself is unlikely to explain a large fraction of the change in the slowdown in the skill premium. To see this, I show in figures 14 and 15 that this slowdown is visible *within* industries and occupation groups. Table 4 formally runs Katz-Murphy regressions occupation-by-occupation and reports the p-value of a Wald test for a trend break in the year 2000.

ability of workers to earn income (Leukhina and McGillicuddy, 2019).

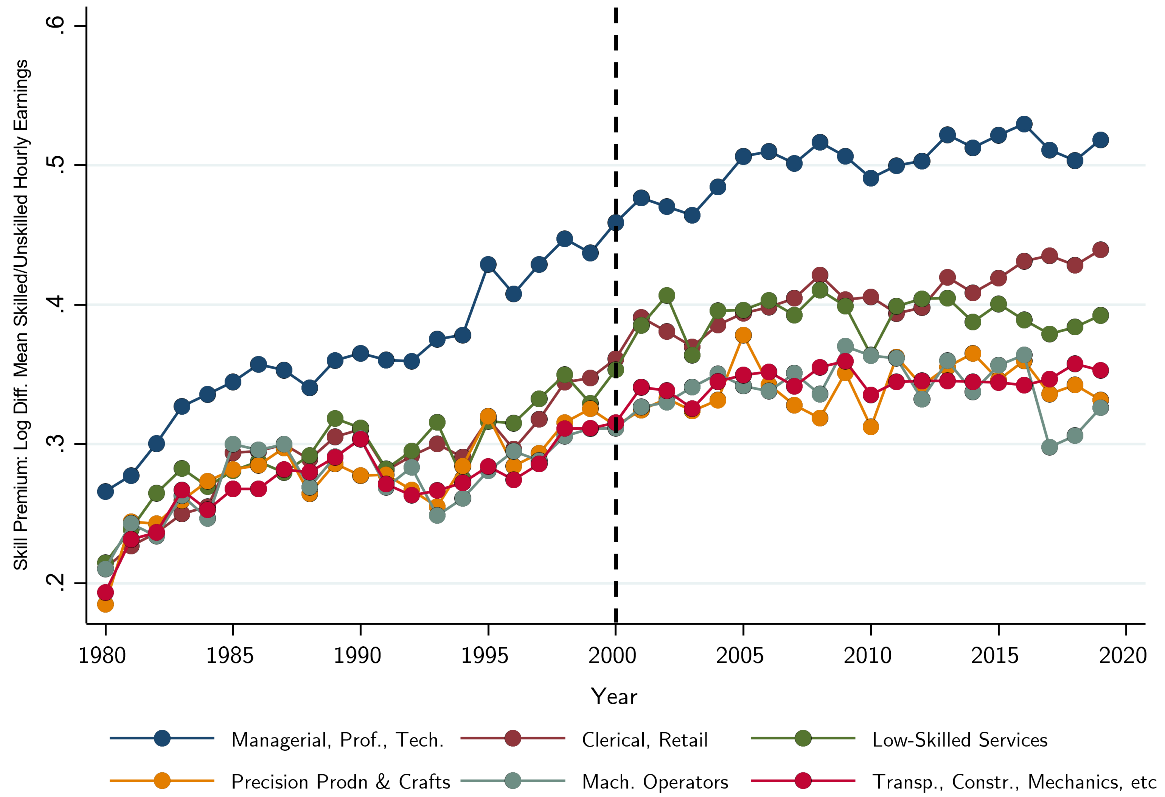


Figure 14: CPS ASEC 1980-2019, Males 16-64. The skill premium is the difference in the (log of) composition-adjusted residual mean hourly earnings of skilled to unskilled. In calculating this mean, averages are taken over composition-adjusted average wages (residualized on race, age and experience categories) within categories of workers by education, experience and race, each group weighted using labor supply weights. Skilled workers are defined as workers with a college degree or more, plus 1/2 the workers with some college. Occupations are defined by consistent `occ1990dd` codes (Autor and Dorn, 2013).

A.2 Rising Exposure of Skilled Occupations to Technical Change and Technology Adoption

I provide a self-contained description of the process by which I document rising exposure of relatively skilled occupations to technical change. The technique follows Webb (2020) very closely and exploits the power of Python's `SpaCy` package, which provides extensive natural language processing functionalities and pre-trained models. I will rely throughout on the trained `en_ core_ web_ trf` model.

First, I define a task as a (verb,noun) pair where the noun is the direct object, or a conjunct of the direct object, of the verb. I extract tasks from descriptions of occupations or technologies as follows.

- Using SpaCy's Dependency Parsing algorithm, I convert descriptions of occupations into a hierarchal sentence tree, which associates with each word (token) in the description a part of speech and a head which is its antecedent in the sentence tree.
- For all words identified by the part of speech tagger as a verb, I identify the first direct object which lies on the verb's subtree as defined by the head relationship. That is, I consider the

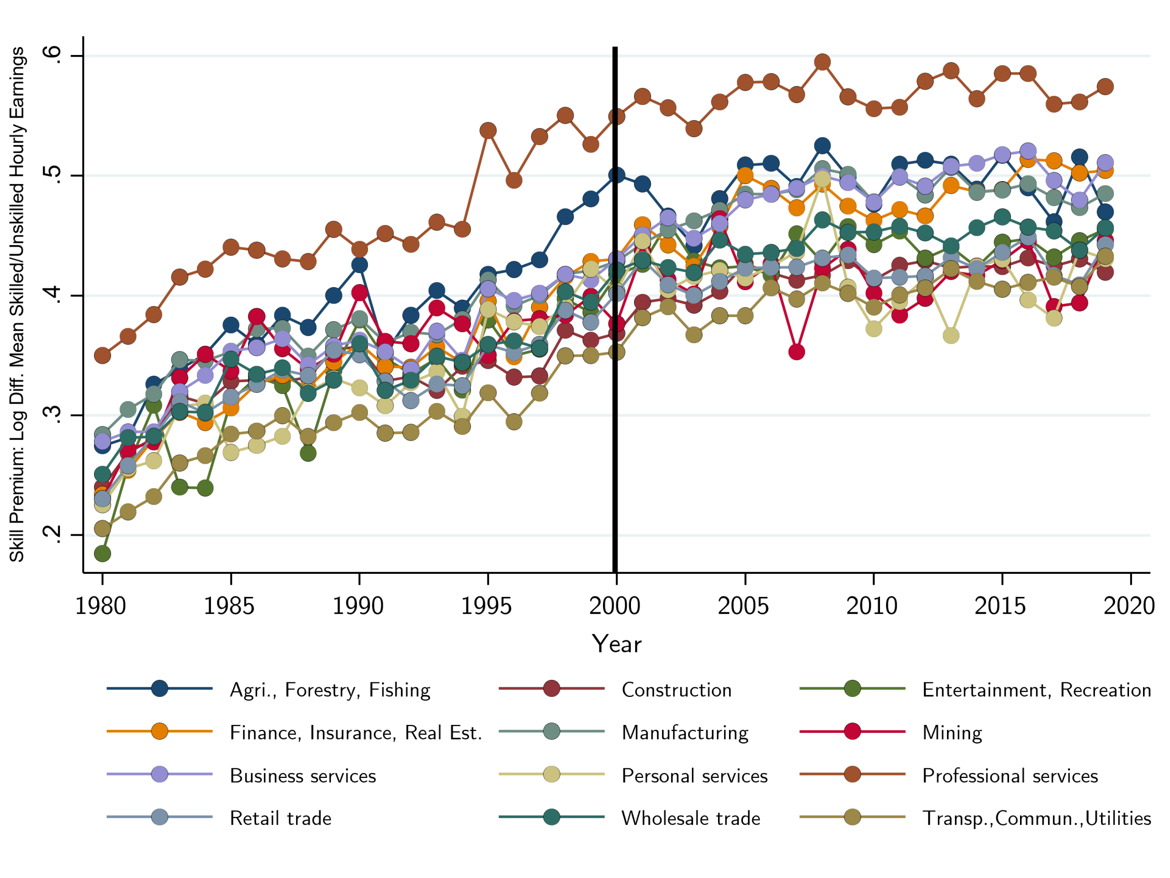


Figure 15: CPS ASEC 1980-2019, Males 16-64. The skill premium is the difference in the (log of) composition-adjusted residual mean hourly earnings of skilled to unskilled. In calculating this mean, averages are taken over composition-adjusted average wages (residualized on race, age and experience categories) within categories of workers by education, experience and race, each group weighted using labor supply weights. Skilled workers are defined as workers with a college degree or more, plus 1/2 the workers with some college. Industries defined by consistent Census `ind90ly` codes assigned by IPUMS, aggregated to highest level.

verb's head. If the head is a noun and is a direct object, then I store the verb and noun. If not, I identify the head of the head, and check this new token. I repeat until a head is identified or until the algorithm reaches the root of the sentence. At this point, I reject the noun if no root is found[42].

- I lemmatize verbs using the lemmatizer that accompanies the `en_ core_ web_ trf` model.
- I translate nouns to a higher level of generality using WordNet 3.1's hierarchical structure for nouns, which associates nouns with their conceptual antecedents. This procedure accounts for the fact that patent descriptions tend to use specific nouns while occupation descriptions tend to be more general. I use WordNet level 4 as the common level to which I convert all nouns.
- I drop (verb, noun) pairs associated with attribution (for instance, containing the verb "to be" or "to have"), and drop patents associated with the invention of specific chemical compounds,

[42]For instance, among the tasks that CEOs perform is "participate in meetings of committees." In this task description, the verb "to participate" has no direct object.

which leads to a large number of false positives due to the trained model's inability to recognize the text associated with them as a noun.

This task extraction procedure described above is used to extract tasks from two datasets. The first is from the titles[43] of all patents awarded in a given year. The second is the task descriptions file from O*NET 26.3, which contains a description of a large number of tasks actually performed on the job by workers of a given occupation. To describe the application of the procedure above, I now introduce some notation. Let $\mathcal{P}_t$ be the set of all patents awarded in a given year. The notation $p \in \mathcal{P}_t$ will indicate that patent p was filed in year t. Let $\mathcal{O}$ be the set of all occupations listed in O*NET 26.3, and let $o \in \mathcal{O}$ index an occupation.

First, for each patent $p \in \mathcal{P}_t$, I obtain the set of tasks $\tau^p(p) = \{\tau \mid \tau \text{ extracted from the title of } p\}$. Let $\mathcal{T}_t^{patents} = \bigcup_{p \in \mathcal{P}_t} \tau^p(p)$ be the set of all tasks that are mentioned at least once in a patent filed at date t. I construct weights for all patents based on the citations they obtain from year t to $t+5$, normalized to add up to 1. Letting $\omega(p)$ be the weight assigned to patent p, I construct the ordered set of weights associated with each task ever mentioned in a patent at year t,

$$\tilde{v}_t^{PATENTS} = \left\{ \sum_{p \in \mathcal{P}_t} \omega(p) \mathbf{1}\left(\tau \in \tau(p)\right) \middle| \tau \in \mathcal{T}_t^{patents} \right\}$$

Note that an element of this vector is the weight associated with a given task, with the weight of a task equal to the sum of the weights of all patents containing it. This weights tasks associated with particularly influential patents by more, making v more representative of the task capabilities of innovations made in year t. I normalize the vector $\tilde{v}_t^{PATENTS}$ to sum to 1.

Second, for each occupation $o \in \mathcal{O}$, O*NET provides a set of task descriptions $\hat{\tau} \in \mathcal{TD}_o^{onet}$. I extract tasks $\tau_o(\hat{\tau}) = \{\tau \mid \tau \text{ extracted from task description } \hat{\tau} \text{ of occupation } o\}$. Let $\mathcal{T}_o^{onet} = \bigcup_{\hat{\tau} \in \mathcal{TD}_o^{onet}} \tau^o(\hat{\tau})$. O*NET provides importance and relevance scale measures for each O*NET task description $\hat{\tau}$, and I use the importance measure to construct the analogous ordered set of weights associated with each task ever mentioned in a task description for a given occupation,

$$\tilde{v}_o^{ONET} = \left\{ \sum_{\hat{\tau} \in \mathcal{TD}_o^{onet}} \omega^{IM}(\hat{\tau}) \mathbf{1}\left(\tau \in \hat{\tau}\right) \mid \tau \in \mathcal{T}_o^{onet} \right\}$$

As above, I normalize the vector $\tilde{v}_o^{ONET}$ to sum to 1.

Next, for each occupation o and year t, I identify the set of tasks which appear in both the description of the occupation and in the set of patent tasks for year t, $\mathcal{T}_{ot} = \mathcal{T}_t^{patents} \bigcap \mathcal{T}_o^{onet}$. I define the restricted vectors

$$v_t^{PATENTS} = \left\{ \sum_{p \in \mathcal{P}_t} \omega(p) \mathbf{1}\left(\tau \in \tau(p)\right) \middle| \tau \in \mathcal{T}_{ot} \right\}$$

and

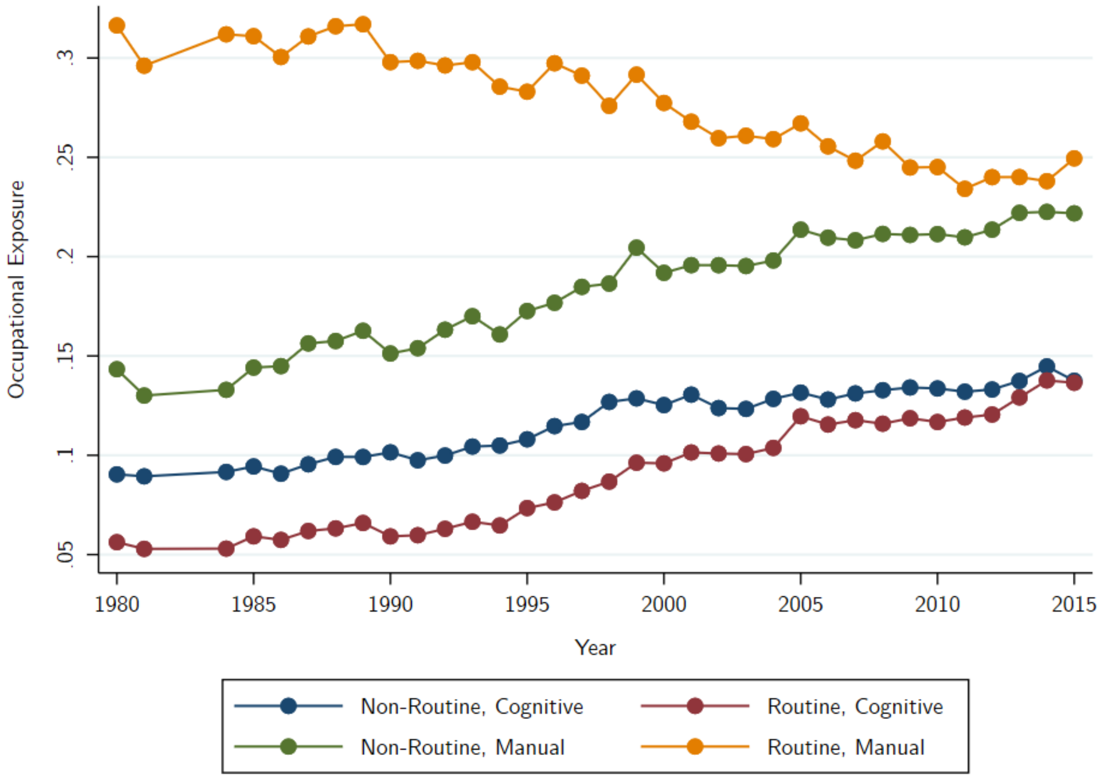
$$v_o^{ONET} = \left\{ \sum_{\hat{\tau} \in \mathcal{TD}_o^{onet}} \omega^{IM}(\hat{\tau}) \mathbf{1}\left(\tau \in \hat{\tau}\right) \middle| \tau \in \mathcal{T}_{ot} \right\}$$

[43] I follow Webb (2020) in using only patent titles, since patent abstracts and titles contain descriptions of the previous art and extra information which raises the noise to signal ratio when extracting tasks. An interesting extension of this approach could be to use patent claims information, available via the USPTO's PatentsView database.

which restricts the weight vectors to only the common set of tasks. I compute the Cosine similarity between these vectors, which provides a measure of the exposure of occupation o to new technologies created in year t.

$$CosineSimilarity(o,t) = \frac{v_t^{PATENTS} \cdot v_o^{ONET}}{||v_t^{PATENTS}|| \, ||v_o^{ONET}||}$$

Figure 16 shows the rising exposure of different groups of occupations by their classification into routine/non-routine and cognitive/manual. While the exposure of routine and manual occupations remains high throughout the sample period, it is clear that there has been a large increase in the exposure of routine/cognitive occupations since 1980. Figure 17 shows that changes in the exposure metric have concentrated on occupations that weren't necessarily the most exposed and is another validation of the metric.



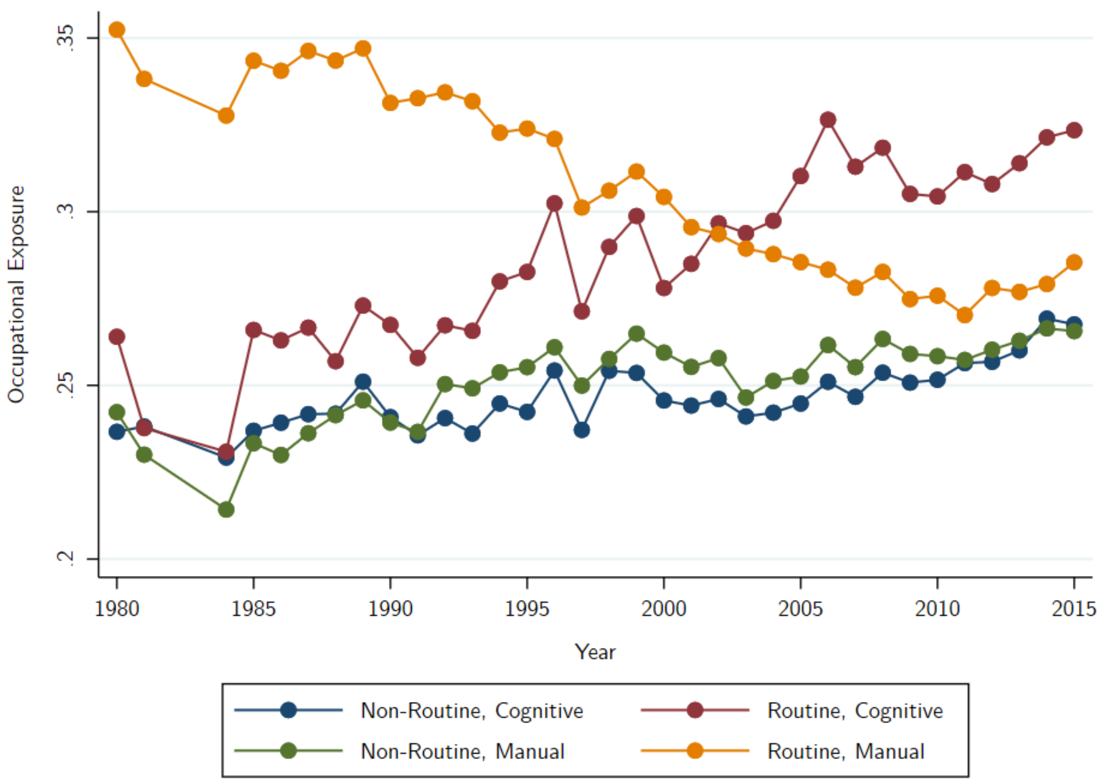


Figure 16: Exposure of different occupational groups to technological change in each year, measured by the cosine similarity in the set of Bigrams (LEFT) or Verbs only (RIGHT). Data from O*NET and USPTO PatentsView.

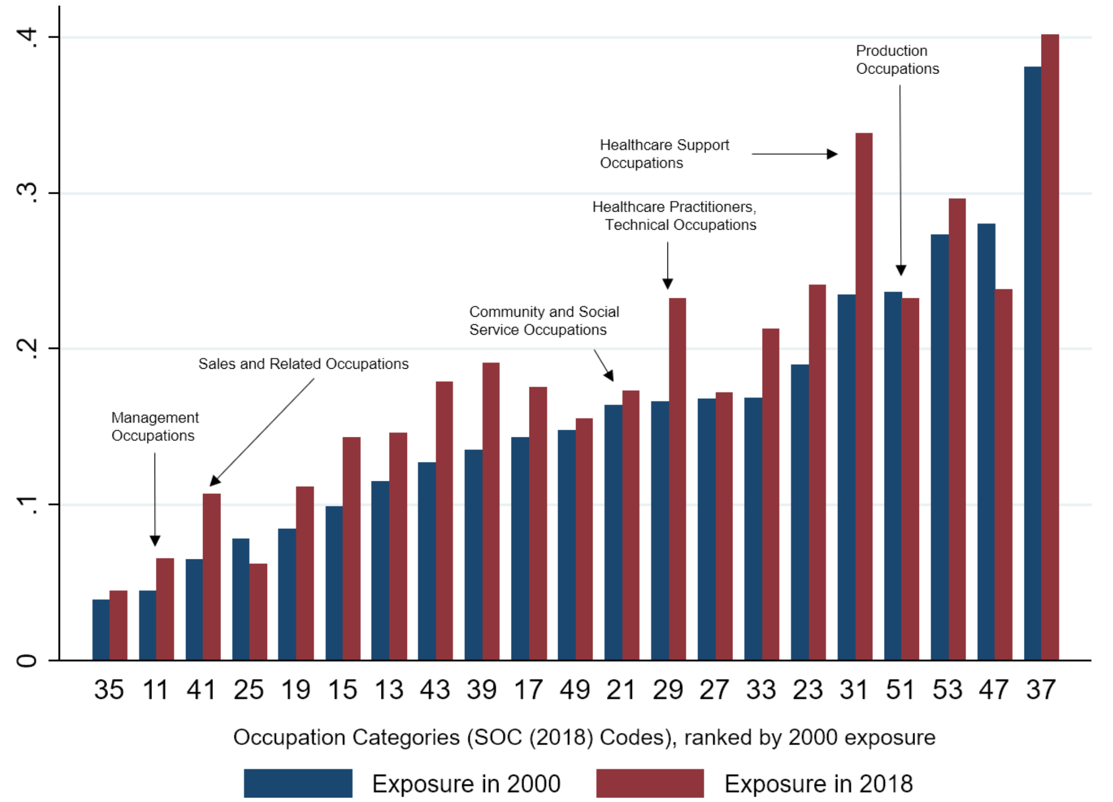


Figure 17: Exposure of different occupational groups to technological change, measured by the cosine similarity in the set of Bigrams. Data from O*NET and USPTO PatentsView.

A.3 Datasets and Data Cleaning Procedures

A.3.1 Census/ACS Data

I obtain data from the Integrated Public-Use Microdata System Flood et al. (2021), which provides harmonized microdata from the US Census bureau. I use the following files in all analyses:

- The 1980 and 1990 5% state samples
- The 2000 5% sample
- The 2009, 2010, 2011, 2017, 2018 and 2019 ACS samples. I pool the 2009-2011 data for 2010 and the 2017-2019 data for 2018.

I follow the following sample selection and data cleaning procedure across all samples.

- I drop unpaid family workers (`classwkrd`=29).
- I retain only regular households and exclude individuals who live in group quarters, such as in institutional settings (i.e. keep households such that `gqtyped`<100 or `gqtyped`>499 or `gqtyped` is missing).
- I retain only civilians. To do this, I do the following.
  - drop workers if their detailed employment status (`empstatd`) lies between 13 and 15 ("armed forces" workers)
  - drop workers if their reported occupation is a military occupation, (`occ90ly`=905), and if their employment status does not explicitly reflect private sector employment (i.e. if `empstatd`$\neq$ 10).
  - drop workers if their reported occupation is a military occupation, (`occ90ly`=905, and their reported industry is a military one (`ind90ly` lies between 940 and 960).
- I retain only working-age workers (aged 18-65).
- I drop any observations for which industry, occupation information and county are all simultaneously missing.

The resulting sample constitutes the base sample on which I estimate wages. To do so, I proceed as follows, largely following Hoffmann et al. (2020).

- I set income variables, hours worked and weeks worked to missing if these variables are allocated. If the allocation flag for these variables indicates that an observation was not allocated and the value is missing, I set the value to 0. If any of these variables is ever negative, I set the corresponding value to 0.
- For some of the years, the variable for weeks worked (`wkswork1`) is not available, and we only have a measure of the interval for weeks worked (`wkswork2`). In these years, I impute `wkswork1` using the midpoint of the interval defined by `wkswork2`. I similarly impute usual hours worked, `uhrswork`, using `hrswork2` where required.

- I construct annual hours as the product of weeks worked (`wkswork2`) and usual hours worked (`uhrswork`). I define workers to be full-time-full-year employed if their usual hours worked per week exceeds 30 and they worked at least 48 weeks last year.

- I recode years of education by sex, race and highest attained educational classification, which is an important variable both to classify workers by skill status and to construct their years of experience. I do this following the imputation procedure in Acemoglu and Autor (2011), using codes available on their website. I construct potential experience using the formula $Exp_{it}^{Pot} = \max\{0, Age_{it} - YearsEduc_{it} - 6\}$.

- I construct consistent occupation codes using IPUMS' `occ1990` codes to construct the coding `occ1990dd`, following the work of Autor and Dorn (2013) and Autor (2015). I construct the mapping between 2010-11 ACS Codes, 2017-2018 ACS Codes and 2019 ACS codes and the variable `occ1990dd` from scratch, following the principles defined by the appendix of Dorn (2009). I do the same for harmonized `ind1990` codes.

- I construct labor supply weights for workers in the data, where the weight for worker i in year t is given by
$$\omega_{it}^{LS} = \omega_{it}^{demog} \times \frac{wkswork_{it}}{52} \times \frac{uhrswork_{it}}{35}$$
Intuitively, the labor supply weight multiplies the demographic weight by the number of full-time-equivalent labor supply hours supplied by a worker in a given year. For workers for whom any one of weeks worked per year or usual hours worked is missing (and yet the worker is classified as employed), I impute the labor supply weight using a regression of the labor supply weights on dummies for sex, race, age categories, education categories and occupation categories.

- I convert all nominal variables to real variables in 2018 dollars, using the GDP Deflator (FRED database series `GDPDEF`).

- I deal with topcoded values as follows.
  - For the 1980 data, I follow the literature (see eg Autor (2015)) and multiply wage and labor income by an adjustment factor of 1.4 if these are equal to their topcode values ($75,000 in current 1980 dollars).
  - For the 1990 data, values of income variables above the threshold were imputed using the state-level *median* value above the threshold (so for instance, wage incomes above $140,000 in Arkansas were replaced by the median wage income for all individuals earning above $140,000 in Arkansas). I use the following Pareto Imputation algorithm for these values. Let $\underline{y}$ be the threshold value for incomes, and let $y^{(50)}$ be the median of values of income above $\underline{y}$. I assume that incomes for individuals within the state above $\bar{y}$ follow a Pareto distribution with tail coefficient α. This implies that
$$\begin{aligned} \Pr\left(y \geq y^{(50)}\right) &= \frac{1}{2} \\ \implies \left(\frac{\underline{y}}{y^{(50)}}\right)^{\alpha} &= \frac{1}{2} \\ \implies \alpha &= \frac{\log 2}{\log\left(y^{(50)}/\underline{y}\right)} \end{aligned}$$

which I use to calculate α state-by-state. I then use the formula for the mean of a Pareto distribution to assign the mean value above the threshold to each household, using the formula for the mean of a Pareto distribution, $\frac{\alpha}{\alpha-1}\underline{y}$.

  - For the years 2000 onward, IPUMS replaced the topcoded values by the mean value above the topcode state-by-state. I do not adjust these values.

- Following Hoffmann et al. (2020), I compute wages as the total market income, which is wage and salary income plus any income from operating one's own business, divided by hours worked. Results based on using salary income alone are virtually identical. Following the labor literature, I drop all observations with real wages below the 1980 minimum wage and drop the top 1% of observations.

- I perform a composition adjustment, following Autor (2019) closely. I group all agents into 2 sex and 3 race (white, black and other) categories. I run a regression of log hourly wages saturated in these categories interacted with a 4th-order polynomial in potential experience, and use the residuals from this regression in further analyses.

A.3.2 CPS Data

I use data from the Annual Socio-Economic Complement of the CPS and clean the data closely following Hoffmann et al. (2020).

- I correct the demographic weights for the ASEC to ensure that they sum to 1 correctly. To do this, I first divide all demographic weights by 10,000, then drop all observations with negative weights, and modify the weights for 2014 by multiplying them either by 3/8 or 5/8 as required by the variable `hflag`.

- I retain only individuals aged 18-65 and recode years of education, accounting for the fact that in 1992 a change in the CPS meant that the years of schooling completed was no longer reported. I impute years of schooling for all households by race, sex and reported highest grade completed following Autor et al. (2008) and Autor (2019). I construct potential experience using the formula $Exp_{it}^{Pot} = \max\{0, Age_{it} - YearsEduc_{it} - 7\}$. Finally, I reduce years of education to 5 categories (less than high school, high school diploma, some college, college degree and some post-college education).

- I impute weeks worked where required following the same procedure as for the ACS, and do the same for usual hours worked.

- For observations starting after 1976, I follow Hoffmann et al. (2020)'s methodology to deal with the fact that topcoded income variables were swapped across neighboring observations. I also use their methodology to construct labor income and wage income for households and use their adjustment procedure.

- I construct composition-adjusted real wages following[44] Autor (2019). I run a regression of individual wages on a set of race, region and experience categories, separately for each year and separately by gender. For each education and experience category, I construct the predicted wages for a white individual with 10 years of experience averaged across all regions

[44]Performing the analysis using the method in Lemieux (2006) produces qualitatively similar results.

proportional to their share of hours worked as the relevant wage variable for analysis. The wage measure thus constructed is composition-adjusted since it is not affected by changes in the gender, race, and region.

A.3.3 Harte-Hanks/Aberdeen Data

Data on IT capital and plant-level stocks of skilled and unskilled workers come from the Computer Intelligence Technology Database (CiTDB), which was produced by Harte-Hanks Intelligence until 2015 and by the Aberdeen Group, a spin-off from Harte-Hanks, thereafter. The database has gone through multiple re-designs over time. My extract of the data consists of three iterations:

- the "MITYE" files: a set of raw data files obtained for the years 1987-1995 and 2000-2001, corresponding to a combination of the variables in the site description and site computing summary tables.
- the "DOMS" files: a set of files for 1992-2002 at 2-year intervals obtained from Mark Doms' data archive, corresponding to a combination of the variables in the site description and site computing summary tables. These data also include the stocks of product-specific IT capital at each plant.
- the "CiTDB" files: a set of files obtained from Harte-Hanks for academic use for the years 1996, 1999, 2003, 2004, 2006, 2008-10, 2012-2016.

The three sets of files contain the same data, but in different formats with different variable names and with different sets of value labels. Industry and geographic data are rich but incomplete and inconsistent across plants over time and even across industries (with a significant number of NAICS codes and SIC codes mutually inconsistent with each other). To fix this and combine the datasets, I employ a number of steps as follows.

1. **Cleaning up industry codes:** I clean up industry codes year-by-year, looking within the appropriate source file for the observation.
    - The "MITYE" files: For each observation, we have the 4-digit SIC Code. I impute SIC Codes for observations with missing codes using the corporation identifiers `corpcode` and `icorp`, which identify the ultimate and immediate corporate parents of the plant. When different plants within the same firm have different SIC Codes, I use the modal SIC Code across plants within each firm for the imputation. When there are multiple modal SIC Codes, I pick the lowest one. The sample consists of 243,596 plants, with an average of 27,066 plants per year.
    - The "DOMS" files: I use the same process as above. The sample contains 1,210,580 observations, with an average of 172,940 plants per year.
    - The "CiTDB" files: I use the same process as above. The sample contains 21,454,593 observations, with an average of 401,451 plants per year prior to 2010 and 3,107,406 plants per year between 2010-2016. This large jump in coverage in 2010 is driven by the large increase in relatively small establishments covered by Harte-Hanks.
2. **Cleaning up geography codes:** Since geographic codes are expected to be stable over time within a site ID, I iteratively clean the data to ensure that each site in my dataset maps to a unique stable geography. I do this as follows.

- I append all three samples together to obtain one large panel. In order to do this, I reconstruct the site ID to be consistent across years (the MITYE files use an 8-digit ID while the DOMS and CiTDB files use a 9-digit ID). In doing so, I only retain observations in the mainland USA (i.e. I drop establishments in Alaska and Hawaii).
- I pre-process the geographic variables I will use, which include the 5-digit zipcode of an establishment, the county FIPS code, the state, metropolitan statistical area (MSA) and major metropolitan area. I correct some obvious errors (inconsistent names and codes for MSAs or zipcodes/towns) directly at this stage to reduce the computational burden of the steps that follow.
- I construct crosswalks which uniquely link the codes associated with each of these variables to the names of the respective geographies. This is important because for a large number of observations, the name of a geography, say the name for a metropolitan statistical area, can be inconsistent with the MSA code reported. I use these crosswalks to harmonize names and codes for counties, metro areas and states.
- I impute the county, zipcode and metro area for all sites which have a unique county associated with all the non-missing observations for these sites.
- I use the GeoNames database to obtain a list of valid ZIP codes and a list of geographic features, retaining cities and towns. If an observation has a valid zipcode, I replace the remaining variables (MSA, county and state) to be consistent with this zipcode. The idea is that the zipcode, being the component of the address that one most frequently enters into surveys and forms, is likely to be more salient and therefore less error prone than the reported county or MSA. In practice, this step never changes the state associated with an observation.
- I choose the county associated with the greatest number of observations for each zipcode. I replace the county variable by this chosen county in any observation with that zipcode and for which the geonames-reported county agrees with the chosen county. I fix some faulty observations by hand.
- I retain observations for which zipcodes do not have a good match to GeoNames. I construct the outer product of this dataset with the GeoNames zipcode dataset and obtain the best matching true zipcode, which is defined as the minimizer of the quadratic loss function

  $$Q_{01} = [Zip_0 - Zip_1]^2 + \left[Levenshtein\left(Zip_0 - Zip_1\right)^2\right]$$

  The idea here is to minimize the distance both in the numeric space in which zipcodes live, but also in terms of their string similarity, which is more relevant for errors associated with data entry or inattention. I replace bad zipcodes by their best matching true zipcode.
- At this stage, I replace all bad observations on geography (i.e. ones with an inconsistent zipcode or county-zipcode match) by good ones associated with the same site ID.
- Finally, for any bad observations that still survive, I impute the zipcode using the modal zipcode using the city-state combination, and thus assign a county based on this modal zipcode. I drop any observations for which this final step is impossible to perform.

3. **Constructing Weights:** I use the County Business Practices dataset to construct the number of establishments at the two-digit industry and county level. I construct weights for

establishments in county c and industry i at date t using

$$w_{cit} = \frac{N_{cit}^{CBP}}{N_{cit}^{HH}}$$

where N_{cit}^{CBP}, N_{cit}^{HH} are the number of establishments in county c and industry i at date t in the County Business Practices and the Harte-Hanks datasets. There are some county-industry-year combinations for which the Harte-Hanks count exceeds the CBP count, which indicates the presence of measurement error in Harte-Hanks. I choose to retain these observations nonetheless, with weights less than 1.

4. **Identifying Accounting Software:** To be classified as an accounting software technology, a technology must meet all of the following criteria.

    (a) The technology must belong to technology class PRG, indicating software.

    (b) Either the model group or the model series must contain one or more of the terms "Accounting", "A/P" (Accounts Payable), "A/R" (Accounts Receivable) or "G/L" (General Ledger).

    (c) For any remaining technologies, the technology's description or definition must include one of the terms above.

    When I perform this classification, I require that a technology ever classified as an accounting technology by an establishment is classified as an accounting technology at all dates for that establishment. For example, suppose an establishment f classifies the technology "Microsoft Excel" as an accounting technology at date t but not at date $t+1$. I consider the establishment f to be using "Microsoft Excel" as an accounting technology at date $t+1$ as well.

A.4 Derivation of the Ideal Price Index Condition from the Retailer's Problem

Recall that the final goods retailer solves the following profit maximization problem.

$$\max_{Y_t, y_t(s)} P_t Y_t - \int p_t y_t dM_t \text{ subject to } [\lambda_t^F] : Y_t = \left[\int y_t^{\frac{\alpha-1}{\alpha}}(s) dM_t(s) \right]^{\frac{\alpha}{\alpha-1}}$$

The focs of this problem give

$$\begin{aligned} [y_t] &: \quad p_t = \lambda_t^F \left(\frac{y_t}{Y_t} \right)^{-1/\alpha} \\ [Y_t] &: \quad P_t = \lambda_t^F \end{aligned}$$

Eliminate the λ_t^F to obtain the demand curves for each firm,

$$y_t(s) = \left(\frac{p_t(s)}{P_t} \right)^{-\alpha} Y_t$$

Next, substitute for y_t into the production function to get

$$\begin{aligned} Y_t &= \left[\int y_t^{\frac{\alpha-1}{\alpha}}(s) dM_t \right]^{\frac{\alpha}{\alpha-1}} \\ \implies Y_t^{\frac{\alpha-1}{\alpha}} &= \int \left(\frac{p_t(s)}{P_t} \right)^{1-\alpha} Y_t^{\frac{\alpha-1}{\alpha}} dM_t \\ \implies P_t^{1-\alpha} &= \int p_t^{1-\alpha}(s) dM_t \\ P_t &= \left[\int p_t^{1-\alpha}(s) dM_t \right]^{\frac{1}{1-\alpha}} \end{aligned}$$

For the final good to be the numeraire we must have $P_t = 1$ which immediately yields the condition 5.

A.5 Static Profit Maximization by the Firm

I prove both lemmas 1 and 2 simultaneously by solving for the period profit function, which requires solving for the allocation of factors across tasks as a first step. Recall that the period profit function $\pi_t(\lambda_s, \lambda_u, z; \boldsymbol{s}_t)$ solves

$$\begin{aligned} \pi_t(\lambda_s, \lambda_u, z; \boldsymbol{s}_t) = \max_{\left\{G_i, \{\mathcal{Y}_{Gi}(x), \ell_i(x), k_i(x)\}_{x=0}^1\right\}_{i=u,s}, G, p, y} & \quad py \\ & - \int_0^1 \left(r_{kt} k_u(x) + w_{ut}\ell_u(x)\right) dx \\ & - \int_0^1 \left(r_{kt} k_s(x) + w_{st}\ell_s(x)\right) dx \end{aligned}$$

subject to, for $i = s, u$,

$$\begin{aligned} [\Lambda^{y,D}] &: \quad y \geq p^{-\alpha} Y_t \\ [\Lambda^{y,F}] &: \quad y \leq z\left[\mu G_u(\ell_u, k_u)^{\frac{\sigma-1}{\sigma}} + (1-\mu) G_s(\ell_s, k_s)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \\ [\Lambda^G_{yi}(x_i)] &: \quad \mathcal{Y}_i(x_i) = \begin{cases} \psi_i(x_i)\,\ell_i(x_i) + k_i(x_i) & x_i \leq \lambda_i \\ \psi_i(x_i)\,\ell_i(x_i) & x_i > \lambda_i \end{cases} \\ [\Lambda^G_{k0i}(x_i)] &: \quad k_i(x_i) \geq 0 \\ [\Lambda^G_{\ell 0i}(x_i)] &: \quad \ell_i(x_i) \geq 0 \\ [\Lambda^G_i] &: \quad G_i \leq \left[\int \mathcal{Y}_i(x_i)^{\frac{\rho-1}{\rho}} dx_i\right]^{\frac{\rho}{\rho-1}} \end{aligned}$$

For ease of notation, define $\mu_u = \mu$ and $\mu_s = 1 - \mu$ and drop time subscripts since this problem is static. The first order conditions of this problem are

$$\begin{aligned} [G_i] &: \quad \Lambda^G_i = z\left(\frac{y}{z}\right)^{\frac{1}{\sigma}} \mu_i G_i^{-\frac{1}{\sigma}} \Lambda^{y,F} \\ [\mathcal{Y}_{Gi}(x_i)] &: \quad \Lambda^G_{yi}(x_i) = \Lambda^G_i \left(\frac{G_i}{\mathcal{Y}_{Gi}(x_i)}\right)^{\frac{1}{\rho}} \\ [k_i(x_i)] &: \quad r_k = \begin{cases} \Lambda^G_{k0i}(x_i) + \Lambda^G_{yi}(x_i) & x_i \leq \lambda \\ \Lambda^G_{k0i}(x_i) & x_i > \lambda \end{cases} \\ [\ell_i(x_i)] &: \quad w_i = \Lambda^G_{\ell 0i}(x_i) + \Lambda^G_{yi}(x_i)\,\psi_i(x_i) \\ [p] &: \quad 0 = y - \Lambda^{y,D} \alpha p^{-\alpha-1} Y \\ [y] &: \quad \Lambda^{y,F} + \Lambda^{y,D} = p \end{aligned}$$

along with appropriate complementary slackness conditions on all the constraints. We can solve this system as follows.

- **Static profit maximization.** Since there are no fixed costs in the static problem and the production function is constant returns to scale, it is clearly never optimal to produce $y = 0$. This immediately implies that $p > 0$ from the foc for p. But the final focs then imply $\Lambda^{y,D}, \Lambda^{y,F} > 0$, so the demand curve and the production function constraints hold with equality. From the foc for p, we get

$$\Lambda^{y,D} = \frac{p^{-\alpha}Y}{\alpha p^{-\alpha-1}Y} = \frac{p}{\alpha} \implies \Lambda^{y,F} = \frac{\alpha-1}{\alpha}p > 0$$

Since $\Lambda^{y,F}$ is the multiplier on the production function constraint, it is also the marginal cost of production by the Envelope Theorem. Thus, this equation just says that

$$p = \frac{\alpha}{\alpha-1}MC$$

which is the standard constant markup formula that emerges from the CES demand system combined with constant returns to scale production.

- **Cost minimization, part 1: All inputs are essential.** We argued above that $y > 0$. Both μ_i values are strictly positive in the calibration. Thus, it must be that $\Lambda_i^G G_i^{\frac{1}{\sigma}} > 0$ which is only possible if $\Lambda_i^G > 0$ and $G_i > 0$. From the foc for $\mathcal{Y}_{Gi}$, this implies that $\Lambda_{yi}^G(x_i)\,\mathcal{Y}_{Gi}(x_i)^{\frac{1}{\rho}} > 0$ which in turn implies that $\mathcal{Y}_{Gi}(x_i) > 0$. This implies that either $k_i(x_i) > 0$ or $\ell_i(x_i) > 0$. Therefore, at most one of $\Lambda_{\ell 0i}^G(x_i)$ and $\Lambda_{k0i}^G(x_i)$ is positive.

- **Cost minimization, part 2: Factor allocation.** Consider the foc for $k_i(x_i)$.
  - For $x_i > \lambda$ we have $\Lambda_{k0i}^G(x_i) = r_k > 0$, which implies $k_i(x_i) = 0$. Since $\mathcal{Y}_{Gi}(x_i) > 0$ nonetheless, this task must be produced entirely by labor, so we have $\ell_i(x_i) = \frac{\mathcal{Y}_{Gi}(x_i)}{\psi_i(x_i)}$.
  - For $x_i < \lambda$ combining the two first order conditions for $k_i(x_i)$ and $\ell_i(x_i)$ yields

    $$\frac{w_i}{\psi_i(x_i)} - r_k = \frac{\Lambda_{\ell 0i}^G(x_i)}{\psi_i(x_i)} - \Lambda_{k0i}^G(x_i) \tag{19}$$

    Define $\hat{\lambda}_i(w_i, r_k)$ as the value of x such that

    $$\frac{w_i}{\psi_i\left(\hat{\lambda}(w_i, r_k)\right)} = r_k$$

    Note that $\hat{\lambda}$ only depends on the aggregate state of the economy.
  - Recall that $\psi_i' > 0$, so the left hand side of 19 is decreasing in x_i (the left hand side is the marginal savings from switching the marginal task x_i from labor to capital, which is decreasing since for high x labor has a stronger comparative advantage.)
    * Consider any task $x_i < \lambda$ which also satisfies $x_i < \hat{\lambda}$. For any such task, the left side of this equation must therefore be positive. This is only possible if $\Lambda_{\ell 0i}^G(x_i) > 0 = \Lambda_{k0i}^G(x_i)$ where the latter equality follows from the fact that at most one of $\Lambda_{\ell 0i}^G(x_i)$ and $\Lambda_{k0i}^G(x_i)$ is positive. This task is therefore performed solely by capital, and we have $k_i(x_i) = \mathcal{Y}_{Gi}(x_i)$.
    * For any tasks $x_i > \hat{\lambda}$ the right side of this equation must be negative. This is only possible if $\Lambda_{\ell 0i}^G(x_i) = 0 < \Lambda_{k0i}^G(x_i)$ where the former equality follows from the fact that at most one of $\Lambda_{\ell 0i}^G(x_i)$ and $\Lambda_{k0i}^G(x_i)$ is positive. This task is therefore performed solely by labor, and we have $\ell_i(x_i) = \frac{\mathcal{Y}_{Gi}(x_i)}{\psi_i(x_i)}$.

– We can now characterize the entire allocation of tasks to capital and labor of each type. Define

$$\lambda_i^*\left(\lambda_i, w_i, r_k\right) = \min\left(\lambda, \hat{\lambda}\left(w_i, r_k\right)\right)$$

For each type i, we have the following configurations.

* A firm will said to be **unconstrained** with respect to labor type $i = s, u$ if $\hat{\lambda}_i\left(w_i, r_k\right) \leq \lambda_i \implies \lambda_i^*\left(\lambda_i, w_i, r_k\right) = \hat{\lambda}\left(\cdot\right)$. Such a firm has labor i and capital i demands that are independent of their technology level λ. In the dynamic problem of the firm, any firm which expects to be unconstrained with respect to labor type i at date $t+1$ given its technology at date t will optimally choose to not invest in technological upgrading.
* A firm will said to be **constrained** with respect to labor type $i = s, u$ if $\hat{\lambda}_i\left(w_i, r_k\right) > \lambda_i \implies \lambda_i^*\left(\lambda_i, w_i, r_k\right) = \lambda$. Such a firm has labor i and capital i demands that are constrained by their technology state λ_i. Such firms will invest in technological upgrading.
* Regardless of constraint status, a firm's demands for labor and capital are characterized by

$$\begin{aligned} \ell_i\left(x_i\right) &= \begin{cases} 0 & x_i \leq \lambda_i^*\left(\lambda_i, w_i, r_k\right) \\ \frac{\mathcal{Y}_G\left(x_i\right)}{\psi_i\left(x_i\right)} & x_i > \lambda_i^*\left(\lambda_i, w_i, r_k\right) \end{cases} \\ k_i\left(x_i\right) &= \begin{cases} \mathcal{Y}_G\left(x_i\right) & x_i \leq \lambda_i^*\left(\lambda_i, w_i, r_k\right) \\ 0 & x_i > \lambda_i^*\left(\lambda_i, w_i, r_k\right) \end{cases} \end{aligned}$$

* In what follows, I'll drop the arguments in $\hat{\lambda}$ and λ^* for clarity of exposition, but it should be understood that these are functions of the firm's state and the aggregate state.

- **Cost Minimization, part 3: Factor demands by a firm**

  – Let's consider a firm $s = \left(\lambda_s, \lambda_u, z\right)$.

  – Given the allocation rules characterized above, we can show from the focs for $\ell_i\left(x_i\right)$ and $k_i\left(x_i\right)$ that

$$\Lambda_{yi}^G\left(x_i\right) = \begin{cases} r_k & x_i \leq \lambda_i^* \\ \frac{w_i}{\psi_i\left(x_i\right)} & x_i > \lambda_i^* \end{cases}$$

which is intuitive, since $\Lambda_{yi}^G\left(x_i\right)$ being the multiplier on the production function for each task is also the marginal cost of each task, by the Envelope Theorem.

  – From the foc for $\mathcal{Y}_{Gi}\left(x_i\right)$ we get $\Lambda_{yi}^G\left(x_i\right) = \Lambda_i^G\left(\frac{G_i}{\mathcal{Y}_{Gi}}\right)^{\frac{1}{\rho}} \implies \mathcal{Y}_{Gi}\left(x_i\right) = G_i\left(\frac{\Lambda_{yi}^G\left(x_i\right)}{\Lambda_i^G\left(x_i\right)}\right)^{-\rho}$.

Raise both sides to the power $\frac{\rho-1}{\rho}$ and integrate over the task index x_i, to get

$$G_i^{\frac{\rho-1}{\rho}} = G_i^{\frac{\rho-1}{\rho}} \int_0^1 \left(\frac{\Lambda_{yi}^G\left(x_i\right)}{\Lambda_i^G\left(x_i\right)}\right)^{1-\rho} dx_i$$

where the left side follows from the definition of the production function for each task aggregate. This gives

$$\Lambda_i^G = \left[\int_0^1 \Lambda_{yi}^G (x_i)^{1-\rho}\, dx_i\right]^{\frac{1}{1-\rho}}$$

Partition the region of integration $[0,1] = [0,\lambda_i^*] \cup [\lambda_i^*, 1]$ to get

$$\Lambda_i^G = \left[\int_0^{\lambda_i^*} r_k^{1-\rho} dx_i + \int_{\lambda_i^*}^1 w_i^{1-\rho} \psi_i (x_i)^{\rho-1}\, dx_i\right]^{\frac{1}{1-\rho}}$$

But Λ_i^G is the lagrange multiplier on the technology constraint for the production of G_i and is therefore the marginal cost of a unit of G_i to the firm, by the Envelope Theorem. Thus,

$$p_{Gi} (\lambda_i, w_i, r_k) = \left[r_k^{1-\rho} \lambda_i^* + w_i^{1-\rho} \int_{\lambda_i^*}^1 \psi_i (x_i)^{\rho-1}\, dx_i\right]^{\frac{1}{1-\rho}}$$

Conditional on a value for λ_i^*, this is isomorphic to the cost function for a CES production function. The key difference here is that the share parameters are now endogenous. This makes the model a microfoundation for papers like Dinlersoz and Wolf (2018) which assume that firms directly choose their share parameters.

– Finally, from the foc for G_i, we know that[45]

$$\begin{aligned}
\Lambda_i^G &= z\left(\frac{y}{z}\right)^{\frac{1}{\sigma}} \mu_i G_i^{-\frac{1}{\sigma}} \Lambda^{y,F} \\
\implies G_i &= \frac{y}{z}\left(\frac{\Lambda_i^G}{\mu_i z \Lambda^{y,F}}\right)^{-\sigma} \\
\implies \mu_i G_i^{\frac{\sigma-1}{\sigma}} &= \mu_i^{\sigma}\left(\frac{y}{z}\right)^{\frac{\sigma-1}{\sigma}} z^{\sigma-1}\left(\frac{p_{Gi}(\cdot)}{\Lambda^{y,F}}\right)^{1-\sigma}
\end{aligned}$$

From the production function,

$$\begin{aligned}
y &= z\left[\mu G_u (\cdot)^{\frac{\sigma-1}{\sigma}} + (1-\mu) G_s (\cdot)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \\
\implies \left(\frac{y}{z}\right)^{\frac{\sigma-1}{\sigma}} &= \left[\mu^{\sigma} z^{\sigma-1}\left(\frac{p_{Gu}(\cdot)}{\Lambda^{y,F}}\right)^{1-\sigma}\left(\frac{y}{z}\right)^{\frac{\sigma-1}{\sigma}} + (1-\mu)^{\sigma} z^{\sigma-1}\left(\frac{p_{Gs}(\cdot)}{\Lambda^{y,F}}\right)^{1-\sigma}\left(\frac{y}{z}\right)^{\frac{\sigma-1}{\sigma}}\right] \\
\implies \Lambda^{y,F} &= \frac{1}{z}\left[\mu^{\sigma} p_{Gu}(\cdot)^{1-\sigma} + (1-\mu)^{\sigma} p_{Gs}(\cdot)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}
\end{aligned}$$

Since $\Lambda^{y,F}$ is the Lagrange multiplier on the production function of the firm, it is also the ultimate marginal cost of production.

[45] We showed above that $\Lambda^{y,F} > 0$ so the production function equation holds with equality, i.e. firms choose not to make use of their "free disposal" option, which is intuitive.

A.6 Existence of a Steady State

I show that a steady state using a series of claims. Throughout, I will use the notation $\tilde{\pi}(\lambda) \equiv \pi(\lambda, 1)$, i.e. $\tilde{\pi}(\lambda)$ is the profit function for a firm with technology feasibility cutoffs λ and unit TFP $(z = 1)$. Analogously, recall that $\tilde{C}_F(\lambda)$ is the unit cost function for a firm with unit productivity. Clearly, we have $\tilde{\pi}(\lambda) = \frac{Y}{\alpha^\alpha}\left(\frac{\tilde{C}_F(\lambda)}{\alpha-1}\right)^{1-\alpha}$. Note that

$$\pi(\lambda, z) = \frac{Y}{\alpha^\alpha}\left(\frac{C_F(\lambda, z)}{\alpha - 1}\right)^{1-\alpha} = z^{\alpha-1}\frac{Y}{\alpha^\alpha}\left(\frac{\tilde{C}_F(\lambda, z)}{\alpha - 1}\right)^{1-\alpha} = z^{\alpha-1}\tilde{\pi}(\lambda)$$

Claim 5. If $\bar{r} > 0$ and $\kappa_i > 0$ then the policy functions for firms are weakly increasing in z. That is, letting $g_i(\lambda, z)$ be the policy function for λ_i', we have $g_i(\lambda, z') \geq g_i(\lambda, z)$ for $z' \geq z$.

Proof. Consider a firm with current state (λ, z) at date t. The choice of $g(\lambda, z) = (g_s(\lambda, z), g_u(\lambda, z))$ is characterized by

$$g(\lambda, z) = \arg\max_{\lambda' \geq \lambda} - \sum_i \kappa_i(\lambda_i' - \lambda_i) + \frac{1 - p_E}{1 + r}\mathbb{E}\left[V_{t+1}(\lambda', z') \mid z\right] \tag{20}$$

Define

$$\Omega_{it+1}(\lambda', z) = \frac{\partial \mathbb{E}(V_{t+1}(\lambda', z') \mid z)}{\partial \lambda_i}$$

which is the expected marginal benefit of raising the chosen cutoff parameter λ_i for a firm that has chosen the vector λ'.

The Envelope Theorem implies that

$$\Omega_{it}(\lambda, z) = \frac{\partial \tilde{\pi}_t(\lambda)}{\partial \lambda_i}\mathbb{E}\left((z')^{\alpha-1} \mid z\right) + (1 - p_E)\min\left[\kappa_i Y_{t+1}, \frac{\Omega_{it+1}(\lambda, z')}{1 + \bar{r}}\middle| z\right]$$

where the minimum operator accounts for the fact that for some states of the world, the constraint $\lambda_i' \geq \lambda_i$ will bind, in which case the firm will continue with the same value of the parameter as it entered the period with.

For every (t, λ) define the sequence

$$\begin{aligned}
\Omega_{it}^{(1)}(\lambda, z) &= \frac{\partial \tilde{\pi}_t(\lambda)}{\partial \lambda_i}\mathbb{E}\left((z')^{\alpha-1} \mid z\right) \\
\Omega_{it}^{(n+1)}(\lambda, z) &= \frac{\partial \tilde{\pi}_t(\lambda)}{\partial \lambda_i}\mathbb{E}\left((z')^{\alpha-1} \mid z\right) + (1 - p_E)\,\mathbb{E}\left\{\min\left[\kappa_i Y_{t+1}, \frac{\Omega_{it+1}^{(n)}(\lambda, z')}{1 + \bar{r}}\right]\middle| z\right\}
\end{aligned}$$

I now show that for all (t, λ) that each term of this sequence is weakly increasing in z. The proof is by induction on n. The base case $n = 1$ follows from the fact that $\mathbb{E}\left((z')^{\alpha-1} \mid z\right)$ is increasing in z as long as the persistence of the process for z, $\rho_z > 0$, and since $\frac{\partial \tilde{\pi}(\lambda)}{\partial \lambda_i} \geq 0$ since an increase in λ_i always weakly reduces costs. Suppose the statement is true for case n, i.e. that $\Omega_{it}^{(n)}(\lambda, z)$ is weakly increasing in z for any t, λ and $n \leq N$. By definition,

$$\Omega_{it}^{(N+1)}(\lambda, z) = \frac{\partial \tilde{\pi}_t(\lambda)}{\partial \lambda_i} \mathbb{E}\left(\left(z'\right)^{\alpha-1} \mid z\right) + (1 - p_E)\, \mathbb{E}\left\{\min\left[\kappa_i Y_{t+1}, \frac{\Omega_{it+1}^{(N)}(\lambda, z')}{1+\bar{r}}\right] \Big| z\right\}$$

The first term is once again weakly increasing in z. The term $\Omega_{it+1}^{(N)}(\lambda, z')$ is increasing in z' and the distribution of z' for higher z first-order-stochastically-dominates the distribution of z' for a lower value. Thus, the expectation of the minimum is itself weakly increasing in z. This completes the inductive step.

Finally, note that $\Omega_t(\lambda, z) = \lim_{n\to\infty} \Omega_t^{(n)}(\lambda, z)$. Since the set of weakly increasing functions is closed, $\Omega_t(\lambda, z)$ must itself be increasing in z. Thus, we have shown that $\mathbb{E}\left(V_{t+1}(\lambda', z') \mid z\right)$, and hence the maximand in equation 20, satisfies increasing differences in (λ_s', z) and (λ_u', z). I now show that the maximand is supermodular in the choice variables λ'. To do this, it is sufficient to show that the cross partial derivative of the maximand is positive, which in turn is guaranteed as long as $\frac{\partial^2 \tilde{\pi}(\lambda)}{\partial \lambda_s \partial \lambda_u} \geq 0$. But since

$$\frac{\partial^2 C_{Ft}}{\partial \lambda_s \partial \lambda_u} \geq 0$$

with strict positivity when $\lambda_s < \hat{\lambda}_{st}, \lambda_u < \hat{\lambda}_{ut}$ and $\alpha > 1$, we know that

$$\begin{aligned}
\frac{\partial^2 \tilde{\pi}(\lambda)}{\partial \lambda_s \partial \lambda_u} &= \frac{\partial}{\partial \lambda_s}\left(\frac{Y}{\alpha^\alpha} \frac{(1-\alpha)}{(\alpha-1)^{1-\alpha}} \tilde{C}_F(\lambda, z)^{-\alpha} \frac{\partial \tilde{C}_F(\lambda, z)}{\partial \lambda_u}\right) \\
&= \frac{Y}{\alpha^\alpha} \frac{(\alpha-1)\alpha}{(\alpha-1)^{1-\alpha}} \tilde{C}_F(\lambda, z)^{-\alpha-1} \frac{\partial^2 \tilde{C}_F(\lambda, z)}{\partial \lambda_s \partial \lambda_u} \geq 0
\end{aligned}$$

which implies that the function $\Omega_t(\lambda, z)$ is supermodular in λ by theorem 6 in Milgrom and Shannon (1994). The choice of λ' lies in the set $[\lambda_s, 1] \times [\lambda_u, 1] \subset \mathbb{R}^2$, which is a rectangle and hence a lattice under the ordering $<$ on the real numbers. We have now verified all the conditions required for application of theorem 5 in Milgrom and Shannon (1994), which guarantees that the functions $g_i(\lambda, z)$ are nondecreasing in z. □

Claim 6. Suppose that $g_Y < \bar{r}$. The policy functions are bounded above by the optimality thresholds, so that $g_i(\lambda, z) \leq \hat{\lambda}_i\left(\frac{w_i'}{r_k'}\right)$.

Proof. Consider a firm which chooses $\lambda_i' > \lambda_i$. The first order condition for such a firm is

$$\begin{aligned}
\kappa_i Y_t &= \frac{1-p_E}{1+\bar{r}}\left[\frac{\partial \tilde{\pi}_t(\lambda')}{\partial \lambda_i} \mathbb{E}\left(\left(z'\right)^{\alpha-1} \mid z\right) + \kappa_i Y_{t+1}\right] \\
\implies \kappa_i\left(\frac{Y_t}{Y_{t+1}} - \frac{1-p_E}{1+\bar{r}}\right) &= \frac{1-p_E}{1+\bar{r}} \frac{1}{Y_{t+1}}\left(\frac{\partial \tilde{\pi}_t(\lambda')}{\partial \lambda_i} \mathbb{E}\left(\left(z'\right)^{\alpha-1} \mid z\right)\right)
\end{aligned}$$

Under the assumption $g_Y < \bar{r}$, the left side is a finite positive constant independent of λ. Now consider the right side. Recall that the support of z is unbounded above. We have,

$$\begin{aligned}\log z' &= \rho_z \log z + \varepsilon \\ \implies \left(z'\right)^{\alpha-1} &= \left(z^{\rho_z}\right)^{\alpha-1} \exp\left(\varepsilon\left(\alpha-1\right)\right) \\ \implies \mathbb{E}\left[\left(z'\right)^{\alpha-1}\right] &= \left(z^{\rho_z}\right)^{\alpha-1} \int_{-\infty}^{\infty} \left\{\exp\left(\varepsilon\right)\right\}^{\alpha-1} d\Phi\left(\varepsilon\right)\end{aligned}$$

The term $\int_{-\infty}^{\infty} \left\{\exp\left(\varepsilon\right)\right\}^{\alpha-1} d\Phi\left(\varepsilon\right)$ is the $\alpha-1$th moment of a lognormal random variable, which we know is finite. It is clear that $\lim_{z\to\infty} \mathbb{E}\left[\left(z'\right)^{\alpha-1}\right] = \infty$ since $\alpha > 1$ and $\rho_z > 0$. The only way this first order condition can hold for firms with arbitrarily high z is thus if $\frac{\partial \tilde{\pi}_t(\lambda')}{\partial \lambda_i} \to 0$, which requires that $\lambda_i \to \hat{\lambda}_i$. We have thus shown that the policy function for firms is bounded above by the optimality threshold. □

Claim 7. If $\bar{r} > 0, \kappa_i > 0$ and $q_{kt} = q_k, S_t = S$, then a steady state exists in which output Y and factor prices w_s, w_u, r_k are constant, all firms operate technologies with capital feasibility cutoffs $\lambda_i' \geq \hat{\lambda}_i\left(\frac{w_i}{r_k}\right)$ and there is no investment in further technology adoption.

Proof. Suppose all firms are in a steady state of the form described.

First, define the objects $\bar{\lambda}_{it} = \int \lambda_{it} dM(s)/\bar{M}$, so that $\bar{\lambda}_{it}$ is the mean value of λ across firms operating at date t. I argue that $\bar{\lambda}_{it} \geq \hat{\lambda}_i\left(\frac{w_i}{r_k}\right)$. To see this, first observe that $\bar{\lambda}_{it}$ is bounded above by 1 and below[46] by some positive number, which means that it eventually lies in an ergodic set with infimum $\underline{\lambda}_i^{\infty}$. I claim that these infima must lie above the optimality cutoffs, i.e. $\underline{\lambda}_i^{\infty} \geq \hat{\lambda}_i\left(\frac{w_i}{r_k}\right)$.

Suppose not, and that $\underline{\lambda}_i^{\infty} < \hat{\lambda}_i\left(\frac{w_i}{r_k}\right)$. For each period t, entrants who entered at date t enter with cutoff parameters $\bar{\lambda}_{it} \geq \hat{\lambda}_i\left(\frac{w_i}{r_k}\right)$. By the contradiction hypothesis, for sufficiently large t, some entrants will enter with $\underline{\lambda}_i^{\infty} < \bar{\lambda}_{it} < \hat{\lambda}_i\left(\frac{w_i}{r_k}\right)$. Note that entrants' cutoff parameters are monotone increasing, and by the unboundedness of TFP z, some firms in the cohort entering at date t will always draw a high-enough z shock such that they will choose a strictly higher $\lambda_{it+1} > \bar{\lambda}_{it}$. The distribution of λ_{it+1} for a cohort entering at date t thus first order stochastically dominates the (degenerate) distribution of λ_i on entry. This implies that the average λ_i for firms entering at a given date t is increasing between t and $t+1$. But this then implies that the average $\bar{\lambda}_{it+1} > \bar{\lambda}_{it}$ almost surely, which contradicts the fact that $\underline{\lambda}_i^{\infty}$ is the infimum of the ergodic set of $\bar{\lambda}_{it}$.

Since the infimum of $\bar{\lambda}_{it} \geq \hat{\lambda}_i$, new entering firms enter with levels of adoption exceeding the optimality threshold and never choose to adopt new technologies since the marginal benefit of doing so is 0. A firm with a value of λ_i exceeding the optimality threshold chooses to produce all type-i tasks in $[0, \hat{\lambda}_i]$ using capital, irrespective of its value of λ_i. The allocations of labor of each type and capital across all firms are therefore equivalent to those of an economy in which all firms operate exactly the same technology, parametrized by $\left(\hat{\lambda}_s, \hat{\lambda}_u\right)$ and whose labor demands are just scaled by their TFP. □

[46]If this were not so and $\bar{\lambda}_{it} = 0$, then there would be no employment of capital in tasks of type i, which is inconsistent with equilibrium since the cost of performing the marginal task would be infinite and firms would always prefer to pay the marginal cost κ_i to automate the marginal task.

Note that the fact that entrants enter at the mean level of technology in the economy is an important assumption for this conclusion.

A.7 The Comparative Advantage Schedule

I assume, following Hubmer and Restrepo (2021), that $\psi_u(z) = B_u \left[z^{\frac{1-\rho-\gamma_u}{\gamma_u}} - 1\right]^{\frac{1}{1-\rho-\gamma_u}}$ and $\psi_s(z) = B_s \left[z^{\frac{1-\rho-\gamma_s}{\gamma_s}} - 1\right]^{\frac{1}{1-\rho-\gamma_s}}$ where $0 < \gamma_i < 1$ and $\bar{\rho}_i = \rho + \gamma_i > 1$. Define $\Psi_k(\lambda^*) = \lambda^*$.
We have,

$$
\begin{aligned}
\Psi_i(\lambda^*) = \int_{\lambda^*}^1 \psi_i(z)^{\rho-1} dz &= B_i^{\rho-1} \int_{\lambda^*}^1 \left(z^{\frac{1-\rho-\gamma_i}{\gamma_i}} - 1\right)^{\frac{\rho-1}{1-\rho-\gamma_i}} dz \\
&= B_i^{\rho-1} \left[1 - (\lambda^*)^{a_i}\right]^{1/a_i}
\end{aligned}
$$

where $a_i = \frac{\rho+\gamma_i-1}{\gamma_i}$. Consider the cost of production of the task intermediate. We have,

$$
P_{Gi}(\lambda, w_i, r_k) = \left[r_k^{1-\rho} \Psi_k(\lambda_i^*(\cdot)) + w_i^{1-\rho} \Psi_i(\lambda_i^*(\cdot))\right]^{\frac{1}{1-\rho}}
$$

Suppose the constraint on technology doesn't bind (this is the case in the steady states I will consider), so that $\lambda_i^* = \hat{\lambda}_i$. Then,

$$
\begin{aligned}
\hat{P}_{Gi}(w_i, r_k) &= r_k \Psi_k(\hat{\lambda}_i)^{\frac{1}{1-\rho}} \left[1 + \left(\frac{w_i}{r_k}\right)^{1-\rho} \frac{\Psi_s(\hat{\lambda}_i)}{\Psi_k(\hat{\lambda}_i)}\right]^{\frac{1}{1-\rho}} \\
\implies \hat{P}_{Gi}(w_i, r_k) &= r_k \hat{\lambda}_i^{\frac{1}{1-\rho}} \left[1 + \left(\frac{w_i/B_i}{r_k}\right)^{1-\rho} \left(\hat{\lambda}_i^{-a_i} - 1\right)^{1/a_i}\right]^{\frac{1}{1-\rho}}
\end{aligned}
$$

Using that $\hat{\lambda}_i^{-a_i} = \hat{\lambda}_i^{\frac{1-\bar{\rho}_i}{\gamma_i}} = \left(\frac{w_i/B_i}{r_k}\right)^{1-\bar{\rho}_i} + 1$, we know that $\frac{\Psi_i(\hat{\lambda})}{\hat{\lambda}} = B_i^{\rho-1}\left[(\lambda^*)^{-a_i} - 1\right]^{1/a_i} = B_i^{\rho-1}\left(\frac{w_i/B_i}{r_k}\right)^{-\gamma}$. Thus,

$$
\begin{aligned}
\hat{P}_{Gi}(w_i, r_k) &= r_k \hat{\lambda}_i^{\frac{1}{1-\rho}} \left[1 + \left(\frac{w_i/B_i}{r_k}\right)^{1-\rho} \left(\hat{\lambda}_i^{-a_i} - 1\right)^{1/a_i}\right]^{\frac{1}{1-\rho}} \\
&= r_k \hat{\lambda}_i^{\frac{1}{1-\rho}} \left[1 + \left(\frac{w_i/B_i}{r_k}\right)^{1-\rho-\gamma_i}\right]^{\frac{1}{1-\rho}} = r_k \hat{\lambda}^{\frac{1}{\gamma_i}}
\end{aligned}
$$

Now we can solve for factor demands. Recall that

$$
\begin{aligned}
k(s) &= \frac{y(s)\tilde{C}_F(s)^\sigma}{z}\left[\frac{P_{Gu}(s)^{\rho-\sigma}}{r_k^\rho}\mu^\sigma\lambda_u + \frac{P_{Gs}(s)^{\rho-\sigma}}{r_k^\rho}(1-\mu)^\sigma\lambda_s\right] \\
\ell_s(s) &= \frac{y(s)}{z}\left(\frac{\tilde{C}_F(s)}{P_{Gs}(s)}\right)^\sigma\left(\frac{P_{Gs}(s)}{w_s}\right)^\rho(1-\mu)^\sigma\Psi_s(\lambda_s) \\
\ell_u(s) &= \frac{y(s)}{z}\left(\frac{\tilde{C}_F(s)}{P_{Gu}(s)}\right)^\sigma\left(\frac{P_{Gu}(s)}{w_u}\right)^\rho\mu^\sigma\Psi_u(\lambda_u)
\end{aligned}
$$

which implies that in our steady state with $\lambda_i = \hat{\lambda}_i$, $i = u, s$, we must have

$$
\begin{aligned}
\frac{\ell_s(s)}{k(s)} &= \frac{\frac{y(s)}{z}\left(\frac{\tilde{C}_F(s)}{P_{Gs}(s)}\right)^\sigma\left(\frac{P_{Gs}(s)}{w_s}\right)^\rho(1-\mu)^\sigma\Psi_s\left(\hat{\lambda}_s\right)}{\frac{y(s)\tilde{C}_F(s)^\sigma}{z}\left[\mu^\sigma\left(\hat{\lambda}_u\right)^{\frac{\bar{\rho}_u-\sigma}{\gamma_u}} + (1-\mu)^\sigma\left(\hat{\lambda}_s\right)^{\frac{\bar{\rho}_s-\sigma}{\gamma_s}}\right]r_k^{-\sigma}} \\
&= \left(\frac{1}{B_s}\right)\left[\frac{(1-\mu)^\sigma\left(\hat{\lambda}_s\right)^{\frac{\bar{\rho}_s-\sigma}{\gamma_s}}}{\mu^\sigma\left(\hat{\lambda}_u\right)^{\frac{\bar{\rho}_u-\sigma}{\gamma_u}} + (1-\mu)^\sigma\left(\hat{\lambda}_s\right)^{\frac{\bar{\rho}_s-\sigma}{\gamma_s}}}\right]\left(\frac{w_s/B_s}{r_k}\right)^{-\bar{\rho}_s}
\end{aligned}
$$

By an exact analogy,

$$
\frac{\ell_u(s)}{k(s)} = \left(\frac{1}{B_u}\right)\left[\frac{\mu^\sigma\left(\hat{\lambda}_u\right)^{\frac{\bar{\rho}_u-\sigma}{\gamma_u}}}{\mu^\sigma\left(\hat{\lambda}_u\right)^{\frac{\bar{\rho}_u-\sigma}{\gamma_u}} + (1-\mu)^\sigma\left(\hat{\lambda}_s\right)^{\frac{\bar{\rho}_s-\sigma}{\gamma_s}}}\right]\left(\frac{w_u/B_u}{r_k}\right)^{-\bar{\rho}_u}
$$

In this model, it is no longer possible to interpret the γ_i as the extra induced elasticity of substitution, because the values of $\hat{\lambda}_u$ and $\hat{\lambda}_s$ will change in response to any change in r_k.

Next, consider the capital equation. We have,

$$
\begin{aligned}
k(s) &= \frac{y(s)\tilde{C}_F(s)^\sigma}{z}\left[\mu^\sigma\left(\hat{\lambda}_u\right)^{\frac{\bar{\rho}_u-\sigma}{\gamma_u}} + (1-\mu)^\sigma\left(\hat{\lambda}_s\right)^{\frac{\bar{\rho}_s-\sigma}{\gamma_s}}\right]r_k^{-\sigma} \\
&= Y\left(\frac{\alpha}{\alpha-1}\right)^{-\alpha}\frac{\tilde{C}_F^{\sigma-\alpha}}{z^{1-\alpha}}\left[\mu^\sigma\left(\hat{\lambda}_u\right)^{\frac{\bar{\rho}_u-\sigma}{\gamma_u}} + (1-\mu)^\sigma\left(\hat{\lambda}_s\right)^{\frac{\bar{\rho}_s-\sigma}{\gamma_s}}\right]r_k^{-\sigma}
\end{aligned}
$$

Use the fact that $\tilde{C}_F = r_k\left[(1-\mu)^\sigma\hat{\lambda}^{\frac{1-\sigma}{\gamma_s}} + \mu^\sigma\hat{\lambda}^{\frac{1-\sigma}{\gamma_u}}\right]^{\frac{1}{1-\sigma}}$ and integrate over the distribution to get

$$
\begin{aligned}
K &= \int k(s)dM_{SS} \\
&= Y\left(\frac{\alpha}{\alpha-1}\right)^{-\alpha}\tilde{C}_F^{\sigma-\alpha}\left[\mu^\sigma\left(\hat{\lambda}_u\right)^{\frac{\bar{\rho}_u-\sigma}{\gamma_u}} + (1-\mu)^\sigma\left(\hat{\lambda}_s\right)^{\frac{\bar{\rho}_s-\sigma}{\gamma_s}}\right]r_k^{-\sigma}\int z^{\alpha-1}dM_{SS} \\
&= Y\left(\frac{\alpha}{\alpha-1}r_k\right)^{-\alpha}\left[\frac{\mu^\sigma\left(\hat{\lambda}_u\right)^{\frac{\bar{\rho}_u-\sigma}{\gamma_u}} + (1-\mu)^\sigma\left(\hat{\lambda}_s\right)^{\frac{\bar{\rho}_s-\sigma}{\gamma_s}}}{\left(\mu^\sigma\left(\hat{\lambda}_u\right)^{\frac{1-\sigma}{\gamma_u}} + (1-\mu)^\sigma\left(\hat{\lambda}_s\right)^{\frac{1-\sigma}{\gamma_s}}\right)^{\frac{\alpha-\sigma}{1-\sigma}}}\right]\int z^{\alpha-1}dM_{SS}
\end{aligned}
$$

A.8 Solution Algorithm

A.8.1 The Steady State

Given values for exogenous variables q_k, S, U and $\bar{r}$, The steady state is characterized by the following system of nonlinear equations.

$$
\begin{aligned}
r_k &= q_k\left(\bar{r}+\delta\right) \\
r_k &= \frac{w_i}{\psi_i\left(\hat{\lambda}_i\right)} \quad , \qquad i=s,u \\
\Psi_i &= \int_{\hat{\lambda}_i}^{1} \psi_i(x)^{\rho-1}dx \quad , \qquad i=s,u \\
P_{Gi} &= \left[r_k^{1-\rho}\hat{\lambda}_i + w_i^{1-\rho}\Psi_i\right]^{\frac{1}{1-\rho}} \quad , \qquad i=s,u \\
\tilde{C}_F &= \left[\mu^{\sigma}P_{Gu}^{1-\sigma} + (1-\mu)^{\sigma}P_{Gs}^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \\
M_{SS}(s) &= \begin{cases} \bar{M}\phi^{stat}(z) & \lambda=\lambda_{ss} \\ 0 & \text{otherwise} \end{cases} \\
\frac{S}{K} &= \frac{(1-\mu)^{\sigma}\frac{P_{Gs}^{\rho-\sigma}}{w_s^{\rho}}\Psi_s\left(\hat{\lambda}_s\right)}{\left[\mu^{\sigma}\frac{P_{Gu}^{\rho-\sigma}}{r_k^{\rho}}\hat{\lambda}_u + (1-\mu)^{\sigma}\frac{P_{Gs}^{\rho-\sigma}}{r_k^{\rho}}\hat{\lambda}_s\right]} \\
\frac{U}{K} &= \frac{\mu^{\sigma}\frac{P_{Gu}^{\rho-\sigma}}{w_u^{\rho}}\Psi_u\left(\hat{\lambda}_u\right)}{\left[\mu^{\sigma}\frac{P_{Gu}^{\rho-\sigma}}{r_k^{\rho}}\hat{\lambda}_u + (1-\mu)^{\sigma}\frac{P_{Gs}^{\rho-\sigma}}{r_k^{\rho}}\hat{\lambda}_s\right]} \\
\frac{K}{Y} &= \left(\frac{\alpha}{\alpha-1}r_k\right)^{-\alpha}\left[\frac{\mu^{\sigma}\frac{P_{Gu}^{\rho-\sigma}}{r_k^{\rho}}\hat{\lambda}_u + (1-\mu)^{\sigma}\frac{P_{Gs}^{\rho-\sigma}}{r_k^{\rho}}\hat{\lambda}_s}{\tilde{C}_F^{\alpha-\sigma}}\right]\int z^{\alpha-1}dM_{SS} \\
1 &= \left(\frac{\alpha}{\alpha-1}\tilde{C}_F\right)\left[\int z^{\alpha-1}dM_{SS}(s)\right]^{\frac{1}{1-\alpha}}
\end{aligned}
$$

where $\phi^{stat}(z)$ is the stationary probability distribution of z resulting from the Markov process 7. Let there be n_s points in the idiosyncratic state space $(\lambda_s, \lambda_u, z)$. This is a system of $12+n_s$ equations in the 12 aggregates

$$
r_k, w_s, w_u, \hat{\lambda}_s, \hat{\lambda}_s, \Psi_s, \Psi_u, P_{Gs}, P_{Gu}, \tilde{C}_F, Y, K
$$

and n_s points at which the distribution function $M(\cdot)$ is computed. After computing these aggregates, it is easy to solve for the profit function $\pi(\cdot)$ using equation 14. Given the profit function is stationary in a steady state, I iterate backward on the Bellman equation 8 to obtain the value function and the stationary policy functions associated with the steady state.

A.8.2 Dynamics

I obtain data on the price of capital q_k from the St. Louis Fed's series for the relative price of investment goods (PIRIC) and use the share of skilled workers S to be the share of hours worked by skilled workers, which I calculate from the CPS. I smooth both time series using an HP Filter with parameter 6.25, since my data is annual. I assume that the absolute value of the growth rate of both time series declines linearly to zero over the subsequent 20 years, following the literature, and therefore obtain a long-run steady state level for both exogenous variables. I choose a large number of periods T such that the economy settles to a new equilibrium at date T. For all dates after the growth rate of q_{kt} and S_t has settled to zero, I assume both of these variables are constant at their terminal values.

To solve for the equilibrium paths for the aggregates $w_{st}, w_{ut}, r_{kt}, Y_t$ I proceed as follows.

1. Given a constant $\bar{r}$ and the path for q_{kt} I solve for the path for r_{kt} using the household's no arbitrage equation 2.
2. I solve for an initial and a final steady state using the initial values and the terminal values of q_{kt} and S_t.
3. I guess paths for $\{w_{st}, w_{ut}, Y_t\}_{t=0}^{T}$ with $t = 0$ corresponding to 1980.
   (a) Given these paths, I solve backward for the value functions and optimal capital feasibility cutoff choices made by firms at each date, with the terminal value function at date T corresponding to the value function for a firm in the long-run steady state.
   (b) This gives me the policy functions $g_{\lambda t}(s, \mathbf{s}_t)$ at each date. I use these policy functions in equation 15 to iterate forward the distribution of firms from date 1 to T with the initial distribution corresponding to one where all firms use technologies with capital feasibility cutoffs at the same initial steady state values.
   (c) I check three equilibrium conditions, given by equations 16, 17 and 18. If these equations hold (up to a specified tolerance), I terminate the iterations.
4. Given paths for aggregate prices, it is easy to solve for the paths of the skill premium w_{st}/w_{ut} and the labor share in value added, $(w_{st}S_t + w_{ut}U_t)/Y_t$.

A.9 Additional Figures

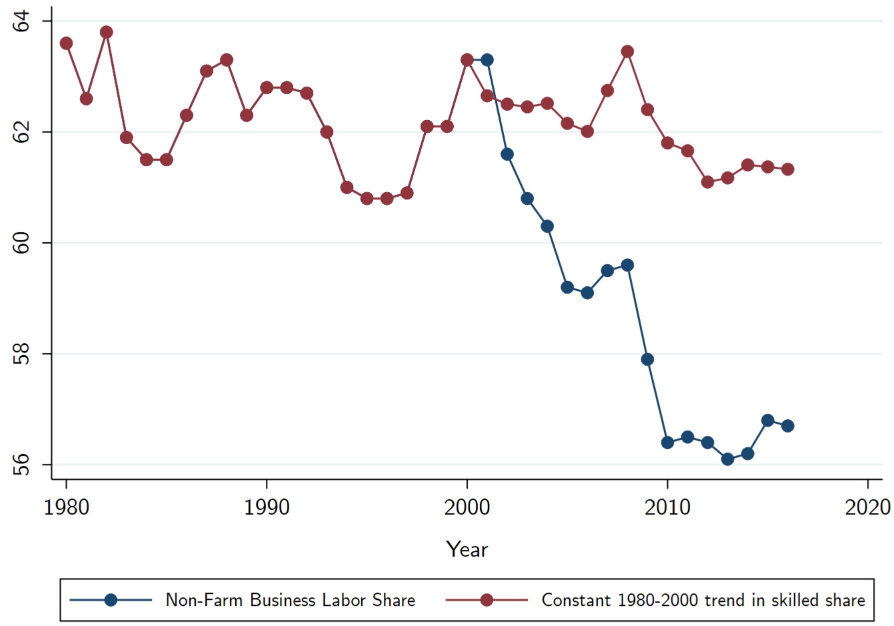


Figure 18: Non-farm Business Sector Labor shares in the data and the counterfactual labor share had the pre-2000 trend in the skilled labor share continued (holding the data on the unskilled labor share unchanged) post 2000. The counterfactual values for years 2001 and onward are given by the predicted values from a regression of the skilled labor share on a time trend from 1980 to 2000. Data from the BEA-BLS integrated national income accounts. The labor share decline in the data between 2000 and 2016 is 6.6 percentage points, while the decline in the counterfactual series is 1.97 percentage points.

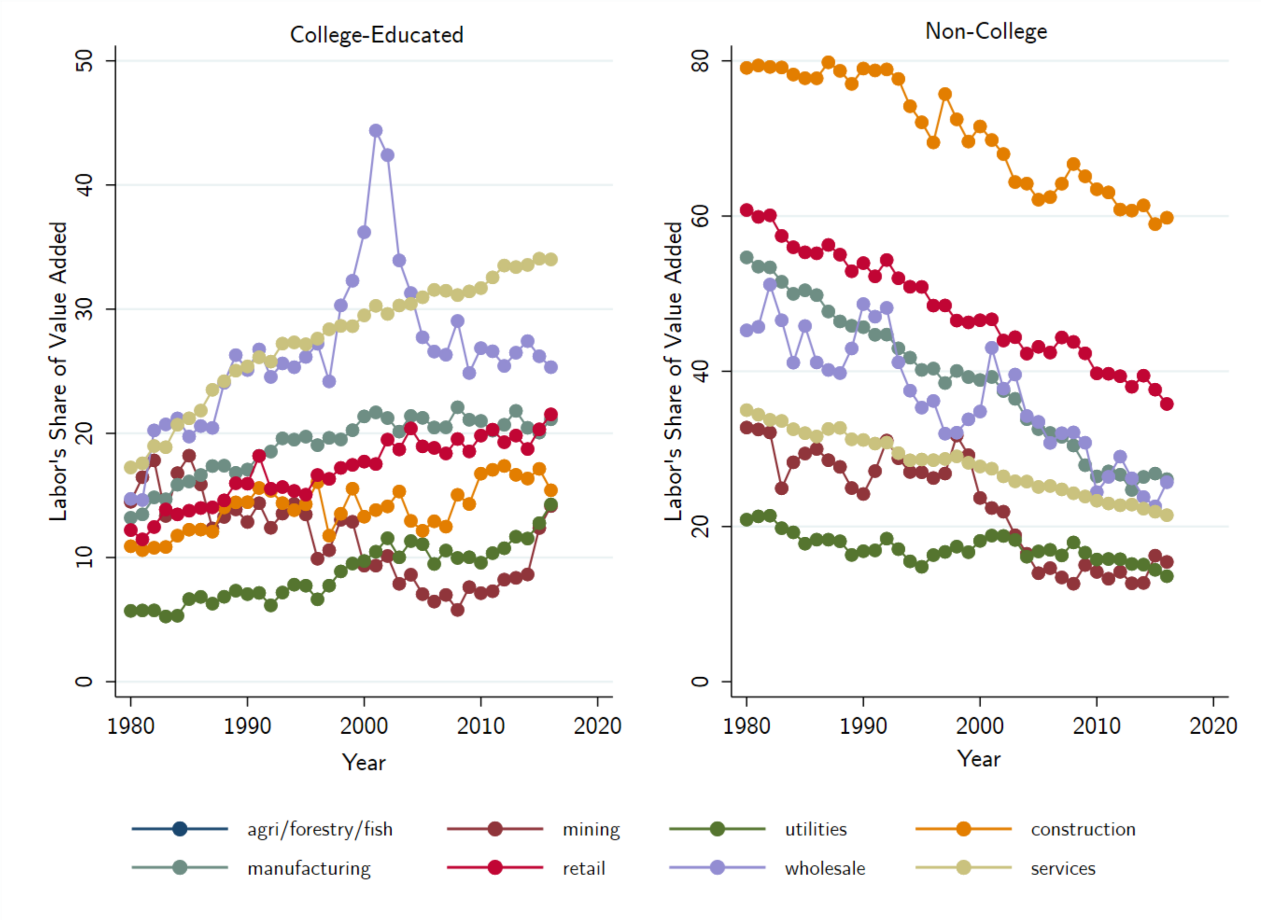


Figure 19: The share of skilled and unskilled workers by industry over time. All data from BLS-BEA integrated national income accounts. The labor share at the industry level is calculated as the ratio of nominal inputs of college educated plus nominal inputs of non-college educated workers divided by value added, which is calculated as nominal gross output less the nominal value of intermediate inputs.

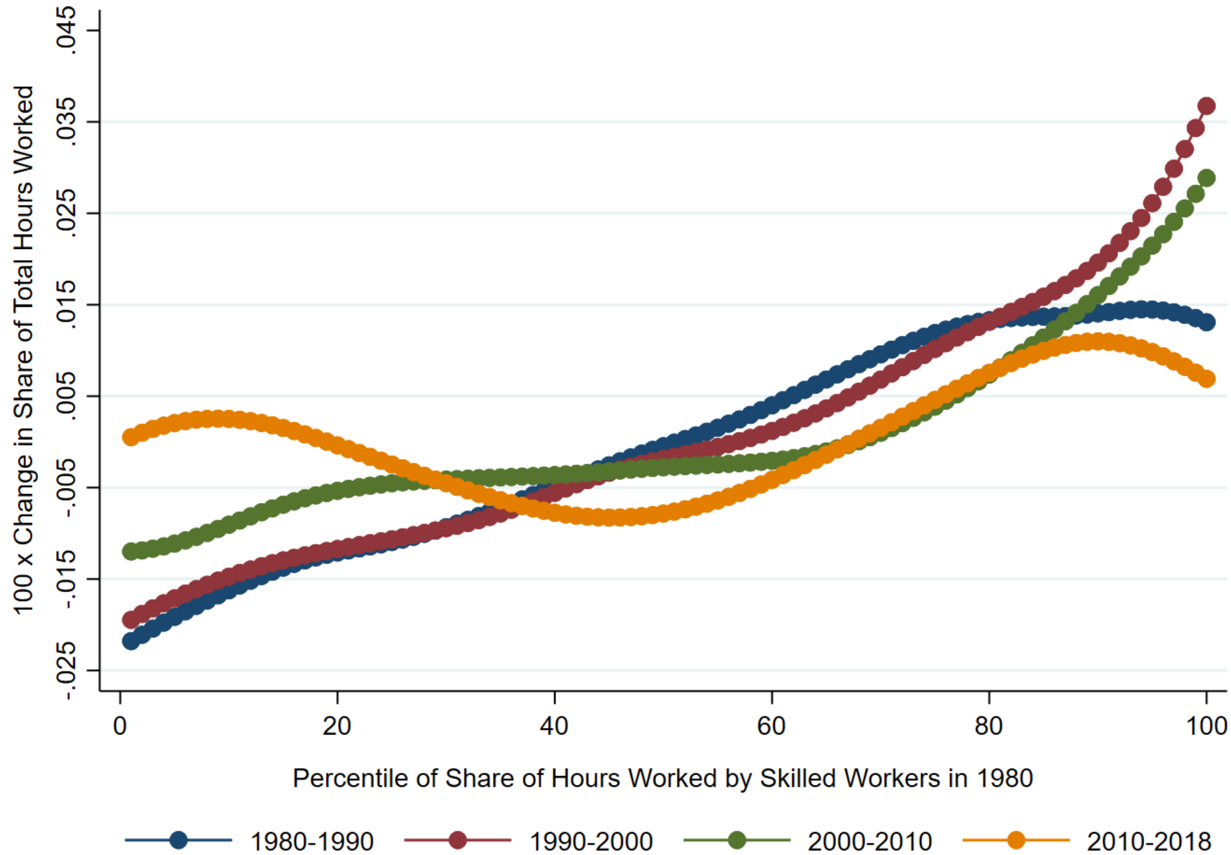


Figure 20: This figure plots smoothed ten-year changes in the share of total hours worked in an occupation. Occupations are ranked in increasing order of skill intensity, measured as the share of hours worked by skilled workers. They are then grouped into percentiles, with each occupation weighted by hours worked in that occupation. Between 1980 and 1990 and 1990 and 2000, growth in employment was concentrated at the top of the skill distribution, which was reflected in a broad increase in the relative demand for skilled workers. Post 2000, growth in the employment shares in the most skilled occupations has slowed dramatically.