The Rise of the Machines: Automation, Horizontal Innovation and Income Inequality

David Hémous and Morten Olsen *

March 2018 (first draft: September 2013)

Abstract

We construct an endogenous growth model with automation and horizontal innovation in an economy with low- and high-skill workers. Automation enables the replacement of low-skill workers with machines, increasing the skill premium and possibly decreasing low-skill wages. Horizontal innovation increases both wages. Higher low-skill wages increase incentives to automate so that automation plays a bigger role as an economy develops. Our model is consistent with a permanently increasing skill premium, a temporary drop in low-skill wages and a drop in the labor share. We calibrate it and show that taxing automation innovation reduces low-skill wages in the long run.

JEL: O41, O31, O33, E23, E25.

KEYWORDS: Endogenous growth, automation, horizontal innovation, directed technical change, income inequality.

*David Hémous, University of Zurich and CEPR, david.hemous@econ.uzh.ch, Morten Olsen, University of Copenhagen, mgo@econ.ku.dk. Morten Olsen gratefully acknowledges the financial support of the European Commission under the Marie Curie Research Fellowship program (Grant Agreement PCIG11-GA-2012-321693) and the Spanish Ministry of Economy and Competitiveness (Project ref: ECO2012-38134). We thank Daron Acemoglu, Philippe Aghion, Ufuk Akcigit, Pol Antràs, Tobias Broer, Steve Cicala, Per Krusell, Brent Neiman, Jennifer Page, Andrei Shleifer, Che-Lin Su, Fabrizio Zilibotti and Joachim Voth among others for helpful comments and suggestions. We also thank seminar and conference participants at IIES, University of Copenhagen, Warwick, UCSD, UCLA Anderson, USC Marshall, Barcelona GSE Summer Forum, the 6th Joint Macro Workshop at Banque de France, Chicago Harris, the 2014 SED meeting, the NBER Summer Institute, the 2014 EEA meeting, Ecole Polytechnique, the University of Zurich, NUS, London School of Economics, the 2015 World Congress of the Econometric Society, ECARES, Columbia University, EIEF, Brown University, Boston University, Yale University, Collège de France and Washington University. We thank Ria Ivandic and Marton Varga for excellent research assistance.

1 Introduction

How does the automation of production drive economic growth and affect the distribution of income? Conversely, how do wages shape technological progress? Developed economies have seen dramatic changes in the income distribution which are often attributed to skill-biased technical change, notably automation. By allowing for the use of machines in some tasks, automation increases economic output, but also reduces the demand for labor in those tasks. As the range of tasks performed by machines is expanding, the general public increasingly worries about the negative consequences of technological progress. Yet, economists often argue that technological development also creates new products and tasks, which boost the demand for labor: certainly many of today's jobs did not exist just a few decades ago.[1] Surprisingly, the economics literature lacks a dynamic framework to analyze the interaction between automation and the creation of new products or tasks. This paper provides the first model to do so.

We build our model to be consistent with three stylized facts of the evolution of the income distribution in the United States over the past 50 years (shown in Figure 1). Since 1963, the college premium (a proxy for the skill premium) has increased by 31% even though the relative skill supply increased substantially. Wages at the bottom of the income distribution have stagnated with accumulated growth between 1963 and 2007 of 13% for non-college educated workers. Finally, over the same period the labor share has declined. As a result, our model does not feature balanced growth: as the economy develops, labor income inequality increases and the labor share declines.

Of course, a large literature exists that relates exogenous technical change to the income distribution (e.g. Goldin and Katz, 2008, and Krusell, Ohanian, Ríos-Rull and Violante, 2002). Previous attempts at endogenizing the direction of technical change rely on factor-augmentation and exogenous shocks to the skill supply (Acemoglu, 1998) and feature balanced growth. The novelty of our approach is that we present an endogenous growth version of a task framework in the vein of Autor, Levy and Murnane (2003) and Acemoglu and Autor (2011), in which the direction of innovation evolves endogenously. Endogenizing technological change matters, not only because it is consistent with empirical evidence, but also, as we show, because the response of an economy with endogenous technological change to policy interventions (such as taxes on automation or on machines) differs from that of an economy with exogenous technological progress.

[1] For instance, the introduction of the telephone led to the creation of new jobs. In 1970 there were 421 000 switchboard operators in the United States. This occupation has largely been automated today.

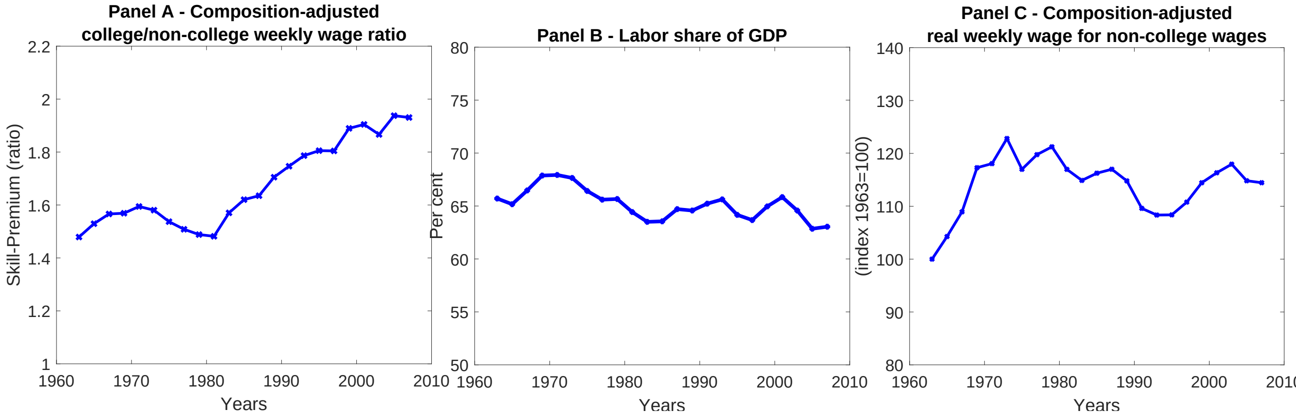

Figure 1: The US skill-premium, labor share and real wage growth for low-skill workers. Panel A and C are taken from Acemoglu and Autor (2011), college educated workers correspond to those with a college degree and half of those with some college, non-college educated workers are the rest. Panel B is from Koh, Santaeulàlia-Llopis and Zheng (2016). See further details in Section 4.

Formally, we consider an expanding variety growth model with low-skill and high-skill workers. Horizontal innovation, modeled as in Romer (1990), increases the demand for both low- and high-skill workers. Automation allows for the replacement of low-skill workers with machines in production and takes the form of a secondary innovation in existing product lines.[2] Within a firm, automation increases the demand for high-skill workers but reduces the demand for low-skill workers. "Non-automated" products only use low-skill and high-skill labor. Once invented, a specific machine is produced with the same technology as a consumption good.

We initially take technical progress as given and study the effect of technological change on wages. An increase in the number of products increases all wages, while an increase in automation both increases the overall productivity of the economy and allows for the substitution away from low-skill workers, resulting in an ambiguous net effect on low-skill wages. Yet, we show that for very general processes of horizontal and automation innovation, the asymptotic growth rate of low-skill wages must be positive, albeit strictly lower than that of high-skill wages. Therefore, the introduction of new (non-automated) products is not sufficient to guarantee balanced growth.

We then endogenize innovation. Low-skill wages play a key role in determining technological change: as they increase the cost advantage of an automated over a non-automated firm increases and the incentive to automate is not constant. Instead, an

[2]Secondary innovations in a growth model were introduced by Aghion and Howitt (1996) who study the interplay between applied and fundamental research.

economy with an initially low level of technology first goes through a phase where growth is mostly generated by horizontal innovation and the skill premium and the labor share are constant. Only when low-skill wages are sufficiently high, do firms invest in automation. During this second phase, the share of automated products increases, the skill premium rises, the labor share drops and low-skill wages might temporarily decrease—replicating the stylized facts of Figure 1. Finally, the economy moves towards its asymptotic steady state. The share of automated products stabilizes as the entry of new, non-automated products compensates for the automation of existing ones. The skill-premium keeps rising but more slowly. The economy will then have endogenously shifted from a Cobb-Douglas aggregate production function to a nested CES.

Finally, we extend our baseline model to include a capital stock and calibrate it to match the evolution of the skill premium, the labor share, productivity and the equipment to GDP ratio the 1960s. As is common in the literature and only for this exercise, we identify skill groups with education groups, such that high-skill workers correspond to college-educated workers. Our model is able to reproduce the trends in the data quantitatively. In particular, the impact of automation on low-skill wages and the skill-premium decreases in the 90s and 2000s even though expenditures on automation remain nearly constant; a response to a point of critique of the skill-biased technological change hypothesis put forward by Card and DiNardo (2002). We use these parameters to discipline our model for policy experiments. We find that a tax on machine has a positive effect on low-skill wages even more so due to the endogeneity of technology. In addition, a tax on automation innovation initially increases low-skill wages but ends up having a negative impact after a few years.

Our modeling of automation as high-skill-biased is motivated by a large empirical literature. Autor, Katz and Krueger (1998) and Autor, Levy and Murnane (2003) demonstrate that computerization is associated with relative shifts in demand for college-educated workers. Bartel, Ichniowski and Shaw (2007) present similar evidence at the firm level. Autor, Katz, and Kearney (2006, 2008) and Autor and Dorn (2013) show that the more recent phenomenon of wage and job polarization, the relative decline of wages and employment in the middle of the income distribution, can be explained by the computer-driven automation of routine tasks often performed by middle-skill workers (storing, processing and retrieving information).[3] Graetz and Michaels (2018) and Acemoglu and Restrepo (2017b) find that the introduction of industrial robots leads to

[3]See also Spitz-Oener (2006) and Goos, Manning and Salomons (2009) for job polarization in Europe, and Feng and Graetz (2016) for a model of why automation targets middle-skill workers.

a reduction in the demand for (mostly) low-skill workers. For our purpose, we will not distinguish between low- and middle-skill workers since both have often performed tasks which have later on been automated.[4]

The idea that high wages might incentivize technological progress in the form of automation dates back to Habakkuk (1962). The associated empirical literature is smaller, but Lewis (2001) finds that low-skill immigration slows down the adoption of automation technology and Hornbeck and Naidu (2014) find that the emigration of black workers from the American South favored the adoption of modern agricultural production techniques there. A large literature shows that the direction of innovation is endogenous in other contexts (see Newell, Jaffe and Stavins, 1999, or Aghion, Dechezleprêtre, Hémous, Martin and Van Reenen, 2016).

There is a small theoretical literature on labor-replacing technology. In Zeira (1998), exogenous increases in TFP raise wages and encourage the adoption of a capital-intensive technology analogous to automation in this paper. Acemoglu (2010) shows that labor scarcity induces labor-saving innovation. Neither paper analyzes labor-saving innovation in a fully dynamic model nor focuses on income inequality. Peretto and Seater (2013) build a dynamic model where innovation allows firms to replace labor with capital. Since wages are constant over time, so is the incentive to automate. In addition, they do not focus on income inequality. In subsequent work, Acemoglu and Restrepo (2017a) also develop a growth model where technical change involves automation and the creation of new tasks. Automation plays a similar role in both papers (although in their baseline version, there is only one type of labor). Yet, while in our model all tasks are symmetric (except for whether they are automated), in theirs, new tasks are exogenously born with a higher labor productivity. As a result, their model features a balanced growth path and their focus is on the self-correcting elements of the economy after a technological shock. In contrast, our model does not feature a balanced growth path and we focus on accounting for secular trends such as the rise in the skill premium.[5]

A large literature argues that skill-biased technical change can explain the increase in the relative demand for skilled workers since the 1970's. This literature can be divided into three strands. The first emphasizes Nelson and Phelps (1966)'s hypothesis that more skilled workers are better able to adapt to technological change (see Lloyd-Ellis,

[4] A previous version of our paper, Hémous and Olsen (2016) presented an extension of our model separating low- and middle- skill workers.

[5] Benzell, Kotlikoff, LaGarda and Sachs (2017), following Sachs and Kotlikoff (2012) build a model where a code-capital stock can substitute for labor, and show that a technological shock which favors the accumulation of code-capital can lead to lower long-run GDP.

1999, Caselli, 1999, Galor and Moav, 2000,, Aghion, Howitt and Violante, 2002, and Beaudry and Green, 2005).[6] While such theories mostly explain transitory increases in inequality, our model features permanent and widening inequality. Yet, we borrow the idea of a shift in production technology spreading through the economy.

A second strand sees the complementarity between capital and skill as the source for the increase in the skill premium. Krusell, Ohanian, Ríos-Rull and Violante (2000) find that the observed increase in the stock of capital equipment can account for most of the variation in the skill premium. Our model also features capital-skill complementarity but our focus is different since we seek to explain *why* innovation has been directed towards automation and analyze the interactions between automation and horizontal innovation.

Finally, a third branch of the literature, building on Katz and Murphy (1992), considers technology to be either high-skill or low-skill labor augmenting. Using this framework, Goldin and Katz (2008) find that technical change has been skill-biased throughout the 20^{th} century in the United States (Katz and Margo, 2014, argue that the relative demand for white-collar workers has been increasing since 1820). Further, the directed technical change literature (Acemoglu, 1998, 2002 and 2007) endogenizes the bias of technology. Such models of factor-augmenting technical change deliver important insights about inequality and technical change, but have no role for labor-replacing technology and therefore cannot generate declines in low-skill wages (a point emphasized in Acemoglu and Autor, 2011). Our model is also a directed technical change framework but deviates from the assumption of factor-augmenting technologies. It explicitly allows for labor-replacing automation, generating the possibility for (temporary) absolute losses for low-skill workers, and permanently increasing income inequality.

Section 2 describes the baseline model with exogenous technology and shows the effects of technological change on wages. Section 3 endogenizes the path of technology and describes the evolution of the economy. Section 4 calibrates an extended version of the model and conducts policy exercises. Section 5 concludes. Appendix 6 presents the main extensions of our model and discusses the identification of our parameters. Appendix 7 presents proofs, other extensions and details the calibration exercise.

[6] Relatedly, Beaudry, Green and Sand (2016) model the IT revolution as an exogenous increase in the demand for organizational capital, which is built by cognitive labor. Once the capital stock reaches its steady-state, the demand for cognitive tasks decreases. We reproduce a somewhat similar pattern but in the growth rate of the skill premium instead of the level of demand for cognitive tasks and with an endogenous increase in the benefits from automation.

2 A Baseline Model with Exogenous Innovation

This section presents a baseline model with exogenous technology to study the consequences of automation and horizontal innovation on factor prices. Section 2.3 derives comparative statics results and section 2.4 the asymptotic behavior of wage for general paths of technology. We discuss some of our modeling assumptions in section 2.5.

2.1 Preferences and production

We consider a continuous time infinite-horizon economy populated by H high-skill and L low-skill workers. Both types of workers supply labor inelastically and have identical preferences over a single final good of:

$$U_{k,t} = \int_t^\infty e^{-\rho(\tau - t)} \frac{C_{k,\tau}^{1-\theta}}{1-\theta} d\tau,$$

where ρ is the discount rate, $\theta \geq 1$ is the inverse elasticity of intertemporal substitution and $C_{k,t}$ is consumption of the final good at time t by group $k \in \{H, L\}$. The final good is produced by a competitive industry combining a set of intermediate inputs, $i \in [0, N_t]$ using a CES aggregator:

$$Y_t = \left(\int_0^{N_t} y_t(i)^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}},$$

where $\sigma > 1$ is the elasticity of substitution between these inputs and $y_t(i)$ is the use of intermediate input i at time t. As in Romer (1990), an increase in N_t represents a source of technological progress. Throughout the paper, we use interchangeably the terms “intermediate input” and “product”.

We normalize the price of Y_t to 1 at all points in time and drop time subscripts when there is no ambiguity. The demand for each intermediate input i is:

$$y(i) = p(i)^{-\sigma} Y, \tag{1}$$

where $p(i)$ is the price of intermediate input i and the normalization implies that the ideal price index, $[\int_0^N p(i)^{1-\sigma} di]^{1/(1-\sigma)}$ equals 1.

Each intermediate input is produced by a monopolist who owns the perpetual rights of production. She can produce the intermediate input by combining low-skill labor, $l(i)$, high-skill labor, $h(i)$, and, possibly, type-i machines, $x(i)$, using the production function:

$$y(i) = \left[l(i)^{\frac{\epsilon-1}{\epsilon}} + \alpha(i) \left(\tilde{\varphi} x\left(i\right) \right)^{\frac{\epsilon-1}{\epsilon}} \right]^{\frac{\epsilon\beta}{\epsilon-1}} h(i)^{1-\beta}, \tag{2}$$

where $\alpha(i) \in \{0, 1\}$ is an indicator function for whether or not the monopolist has access to an automation technology which allows for the use of machines. If the product is not automated ($\alpha(i) = 0$), production takes place using a Cobb-Douglas production function with only low-skill and high-skill labor and a low-skill factor share of β. If the product is automated ($\alpha(i) = 1$) machines can be used in the production process. We allow for perfect substitutability, in which case $\epsilon = \infty$ and the production function is $y(i) = \left[l(i) + \alpha(i)\tilde{\varphi}x\left(i\right)\right]^{\beta} h(i)^{1-\beta}$. The parameter $\tilde{\varphi}$ is the relative productivity advantage of machines over low-skill workers and G denotes the share of automated products.

Since each input is produced by a single firm, we identify each input with its firm and refer to a firm which uses an automated production process as an automated firm. We refer to the specific labor inputs provided by high-skill and low-skill workers in the production of different inputs as "different tasks" performed by these workers, so that each product comes with its own tasks. It is because $\alpha(i)$ is not fixed, but can change over time, that our model captures the notion that machines can replace low-skill labor in new tasks. A model with a fixed $\alpha(i)$ for each product would only allow for machines to be used more intensively in production, but always for the same tasks. Although, we will refer to x as "machines", our interpretation also includes any form of computer inputs, algorithms, the services of cloud-providers, etc.

For now, machines are an intermediate input. Once invented, machines of type i are produced competitively one for one with the final good, such that the price of an existing machine for an automated firm is always equal to 1 and technological progress in machine production follows that in the rest of the economy. Yet, our model can capture the notion of a decline in the real cost of equipment, as automation for firm i can equivalently be interpreted as a decline of the price of machines i from infinity to 1.

2.2 Equilibrium wages

In this section we derive how wages are determined in equilibrium, taking as given the technological levels N (the number of products), G (the share of automated products) and the employment of high-skill workers in production, $H^P \equiv \int_0^N h(i)di$ (we let $H^P \leq H$ to accommodate later sections where high-skill labor is used to innovate).

First, note that all automated firms are symmetric and therefore behave in the same way. Similarly all non-automated firms are symmetric. This gives aggregate output of:

$$Y = N^{\frac{1}{\sigma-1}} \times \left((1-G)^{\frac{1}{\sigma}} \Big(\underbrace{(L^{NA})^{\beta}(H^{P,NA})^{1-\beta}}_{T_1} \Big)^{\frac{\sigma-1}{\sigma}} + G^{\frac{1}{\sigma}} \Big(\underbrace{[(L^{A})^{\frac{\epsilon-1}{\epsilon}} + (\tilde{\varphi}X)^{\frac{\epsilon-1}{\epsilon}}]^{\frac{\epsilon\beta}{\epsilon-1}}(H^{P,A})^{1-\beta}}_{T_2} \Big)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \tag{3}$$

where L^A (respectively L^{NA}) is the total mass of low-skill workers in automated (respectively non-automated) firms, $H^{P,A}$ (respectively $H^{P,NA}$) is the total mass of high-skill workers hired in production in automated (respectively non-automated) firms and $X = \int_0^N x(i)di$ is total use of machines. The aggregate production function takes the form of a nested CES between two sub-production functions. The first term T_1 captures the classic case where production takes place with constant shares between factors (low-skill and high-skill labor), while the second term T_2 represents the factors used within automated products and features the substitutability between low-skill labor and machines. G is the share parameter of the "automated" products nest and therefore an increase in G is T_2-biased (as $\sigma > 1$). $N^{\frac{1}{\sigma-1}}$ is a TFP parameter. Besides the functional form the aggregate production function (3) differs from the often assumed aggregate CES production function in two ways: First, instead of the often-used factor-augmenting technical in an aggregate production function, we explicitly model automation of tasks and derive the aggregate production function from the cost function of individual firms. Second, with an endogenous G we will be able to capture effects that the usual focus on an exogenous aggregate production function cannot.[7]

The unit cost of intermediate input i is given by:

$$c(w_L, w_H, \alpha(i)) = \beta^{-\beta}(1-\beta)^{-(1-\beta)}(w_L^{1-\epsilon} + \varphi\alpha(i))^{\frac{\beta}{1-\epsilon}} w_H^{1-\beta}, \tag{4}$$

where $\varphi \equiv \tilde{\varphi}^{\epsilon}$, w_L denotes low-skill wages and w_H high-skill wages. When $\epsilon < \infty$, $c(\cdot)$ is strictly increasing in both w_L and w_H and $c(w_L, w_H, 1) < c(w_L, w_H, 0)$ for all $w_L, w_H > 0$ (automation reduces costs). Price is set as a markup over costs: $p(i) = \sigma/(\sigma-1) \cdot c(w_L, w_H, \alpha(i))$. Using Shepard's lemma and equations (1) and (4) delivers

[7] At the firm level, this model features an elasticity of substitution between high-skill labor and machines equal to that between high-skill and low-skill labor. This, however, does not hold at the aggregate level, consistent with Krusell et al. (2000), who argue that the aggregate elasticity of substitution between high-skill and low-skill labor is greater than that between high-skill labor and machines.

the demand for low-skill labor of a single firm.

$$l(w_L, w_H, \alpha(i)) = \beta \frac{w_L^{-\epsilon}}{w_L^{1-\epsilon} + \varphi\alpha(i)} \left(\frac{\sigma - 1}{\sigma}\right)^{\sigma} c(w_L, w_H, \alpha(i))^{1-\sigma} Y, \tag{5}$$

which decreases in w_L and w_H. The effect of automation on demand for low-skill labor in a firm is generally ambiguous. This is due to the combination of a negative *substitution* effect (automation allows for substitution between machines and low-skill workers) and a positive *scale* effect (automation decreases costs, lowers prices and increases production). As we focus on labor-substituting innovation, we impose the condition $\epsilon > 1 + \beta(\sigma - 1)$ *throughout* the paper which is necessary and sufficient for the substitution effect to dominate and ensures $l(w_L, w_H, 1) < l(w_L, w_H, 0)$ for all $w_L, w_H > 0$.

Let $x(w_L, w_H)$ denote the use of machines by an automated firm. The relative use of machines and low-skill labor for such a firm is then:

$$x(w_L, w_H)/l(w_L, w_H, 1) = \varphi w_L^{\epsilon}, \tag{6}$$

which increases in w_L as the wage is also the price of low-skill labor relative to machines.

The iso-elastic demand (1), coupled with constant mark-up $\sigma/(\sigma - 1)$, implies that revenues are given by $R(w_L, w_H, \alpha(i)) = ((\sigma - 1)/\sigma)^{\sigma-1} c(w_L, w_H, \alpha(i))^{1-\sigma} Y$ and profits are a fixed share of revenue: $\pi(w_L, w_H, \alpha(i)) = R(w_L, w_H, \alpha(i))/\sigma$. We define $\mu \equiv \beta(\sigma - 1)/(\epsilon - 1) < 1$ (by our assumption on ϵ). Using (4), the relative revenues (and profits) of non-automated and automated firms are given by:

$$R(w_L, w_H, 0)/R(w_L, w_H, 1) = \pi(w_L, w_H, 0)/\pi(w_L, w_H, 1) = (1 + \varphi w_L^{\epsilon-1})^{-\mu}, \tag{7}$$

which is a decreasing function of w. As non-automated firms rely more heavily on low-skill labor, their relative market share drops with higher low-skill wages.

Since firms' profits are a constant share of firms' revenues, aggregate profits are a constant share $1/\sigma$ of output Y. Similarly, the share of firms' revenues accruing to high-skill labor in production is the same for all firms and given by $\nu_h = (1 - \beta)(\sigma - 1)/\sigma$. Therefore payment to high-skill labor in production is a constant share of output:

$$w_H H = (1 - \beta)\frac{\sigma - 1}{\sigma} N \left[G R(w_L, w_H, 1) + (1 - G) R(w_L, w_H, 0)\right] = (1 - \beta)\frac{\sigma - 1}{\sigma} Y. \tag{8}$$

Using factor demand functions, the share of revenues accruing to low-skill labor is given by $\nu_l(w_L, w_H, \alpha(i)) = \frac{\sigma-1}{\sigma}\beta\left(1 + \varphi w_L^{\epsilon-1}\alpha(i)\right)^{-1}$, and decreases with automation. Using labor market clearing ($\int_0^N l(i)di = L$), we obtain total wages of low-skill workers as:

$$w_L L = N\left[GR(w_L, w_H, 1)\nu_l(w_L, w_H, 1) + (1-G)R(w_L, w_H, 0)\nu_l(w_L, w_H, 0)\right]. \tag{9}$$

Equations (7), (8) and (9) give the high-skill to low-skill labor share in production as:[8]

$$\frac{w_H H^P}{w_L L} = \frac{1-\beta}{\beta}\frac{G + (1-G)(1+\varphi w_L^{\epsilon-1})^{-\mu}}{G\left(1+\varphi w_L^{\epsilon-1}\right)^{-1} + (1-G)(1+\varphi w_L^{\epsilon-1})^{-\mu}}. \tag{10}$$

This expression gives the *relative demand curve* for high-skill and low-skill labor. We represent this relationship for given technology levels and factor supply ratio L/H^P by a curve in the (w_L, w_H) space in Figure 2. For $G = 0$, the curve is a straight line, with slope $(1-\beta)L/(\beta H^P)$, reflecting the constant factor shares in a Cobb-Douglas economy. For $G > 0$, the right-hand side of (10) increases in w, so that the relative demand curve is non-homothetic and rotates counter-clockwise as w_L grows. Intuitively, higher low-skill wages both induce more substitution towards machines in automated firms (as reflected by the term $(1+\varphi w_L^{\epsilon-1})^{-1}$ in equation (10)) and improve the cost-advantage and therefore the market share of automated firms (term $(1+\varphi w_L^{\epsilon-1})^{-\mu}$) and increases the ratio of high-skill to low-skill labor share in production. This immediately implies that in the long-run low-skill wages and high-skill wages cannot grow at the same rate, i.e. we cannot have positive balanced growth.

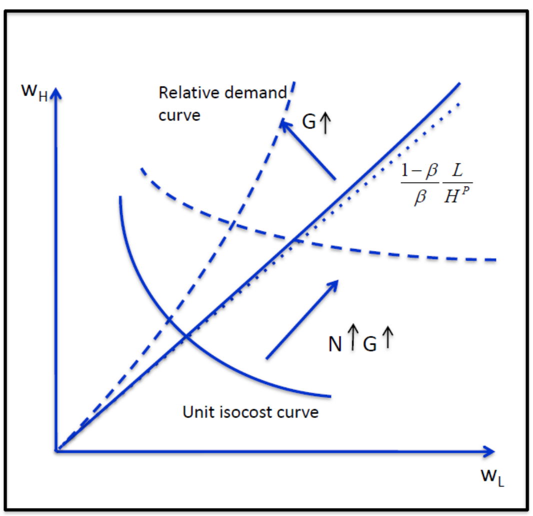

Figure 2: Relative demand curve and isocost curve for different values of N and G.

[8]When $\epsilon = \infty$, the skill premium is given by $\frac{w_H}{w_L} = \frac{1-\beta}{\beta}\frac{L}{H^P}$ if $w_L < \widetilde{\varphi}^{-1}$ such that no firm uses machines, and $\frac{w_H}{w_L} = \frac{1-\beta}{\beta}\frac{L}{H^P}\frac{G+(1-G)(\widetilde{\varphi}w_L)^{-1}}{(1-G)(\widetilde{\varphi}w_L)^{-1}}$ if $w_L > \widetilde{\varphi}^{-1}$.

With constant mark-ups, the cost equation (4) and the price normalization give:

$$\frac{\sigma}{\sigma-1}\frac{N^{\frac{1}{1-\sigma}}}{\beta^{\beta}\left(1-\beta\right)^{1-\beta}}\left(G\left(\varphi+w_L^{1-\epsilon}\right)^{\mu}+(1-G)w_L^{\beta(1-\sigma)}\right)^{\frac{1}{1-\sigma}}w_H^{1-\beta}=1. \tag{11}$$

This relationship defines the *unit isocost curve* in figure 2. It shows the positive relationship between real wages and the level of technology given by N, the number of intermediate inputs, and G the share of automated firms. Together (10) and (11) determine real wages uniquely as a function of N, G and H^P.

Given the amount of resources devoted to production (L, H^P), the static equilibrium is closed by the final good market clearing condition:

$$Y = C + X \tag{12}$$

where $C = C_L + C_H$ is total consumption. GDP includes the payment to labor and aggregate profits so GDP and the total labor share become

$$GDP \equiv \frac{1}{\sigma}Y + w_L L + w_H H, \qquad LS = 1 - \frac{1}{1+(\sigma-1)(1-\beta)\left(\frac{w_L L}{w_H H^P}+\frac{H}{H^P}\right)}, \tag{13}$$

where the second equality uses (8). Therefore, the labor share decreases in the skill premium for a given mass of high-skill workers in production H^P.

2.3 Technological change and wages

The consequences of technological changes on the level of wages are most easily seen with the help of Figure 2. An increase in the number of products, N, pushes out the isocost curve and increases both low-skill and high-skill wages. When $G = 0$, both types of wages grow at the same rate as the relative demand curve is a straight line, but for $G > 0$, the demand curve is non-homothetic and the skill premium grows. Therefore, an increase in N at *constant* G is high-skill biased.

An increase in the share of automated products G has a positive effect on high-skill wages and the skill premium but an ambiguous effect on low-skill wages: Higher automation increases the productive capability of the economy and pushes out the isocost curve (an *aggregate productivity* effect), which increases low-skill wages, but it also allows for easier substitution away from low-skill labor which pivots the relative demand curve counter-clockwise (an *aggregate substitution* effect), decreasing low-skill wages.

Therefore automation is always high-skill labor biased (w_H/w_L increases) but low-skill labor saving (w_L decreases) if and only if the aggregate substitution effect dominates the aggregate productivity effect. Formally, one can show (proof in Appendix 7.1):[9]

Proposition 1. *Consider the equilibrium (w_L, w_H) determined by equations (10) and (11). Assume that $\epsilon < \infty$, it holds that*

A) An increase in the number of products N (keeping G and H^P constant) leads to an increase in both high-skill (w_H) and low-skill wages (w_L). Provided that $G > 0$, an increase in N also increases the skill premium w_H/w_L and decreases the labor share.

B) An increase in the share of automated products G (keeping N and H^P constant) increases the high-skill wages w_H, the skill premium w_H/w_L and decreases the labor share. Its impact on low-skill wages is generally ambiguous, but low-skill wages are decreasing in G if i) $1 \leq (\sigma - 1)(1 - \beta)$ or if ii) N and G are high enough.

Automation is low-skill labor saving if $(1 - \beta)(\sigma - 1) \geq 1$. The aggregate substitution effect is larger than the scale effect when *i*) σ is large (as then the increased demand for the newly automated product does not lead to a large increase in demand for other products). *ii*) when the cost share of the low-skill labor-machines aggregate β is small (as the cost-saving effect of automation is smaller in that case). *iii*) when G is high as most of the aggregate productivity gains are realized for low G and for high G the automation of one more firm hurts low-skill workers more, as there are fewer non-automated firms left. *iv*) when N is large and the higher wages make the substitution effect stronger.

One could also consider the effect of an increase in the number of non-automated products (that is an increase in N keeping GN constant), which corresponds to the "horizontal innovation" introduced in section 3. Such technological change pushes out the iso-cost curve but also makes the relative demand curve rotate clockwise. This increases both low-skill and high-skill wages (see proof in Appendix 7.1). In addition, for N large enough or $\epsilon < \sigma$, the rotation of the relative demand curve is sufficiently strong for this form of technical change to be low-skill labor biased.[10]

In section 3, when we specify the innovation process, we will show that as the number of products N increases, the share of automated products endogenously increases from

[9] In the perfect substitute case, $\epsilon = \infty$, w_H increases in N and weakly increases in G, w_H/w_L weakly increases in N and G and w_L weakly increases in N and weakly decreases in G provided that $1/(1-\beta) \leq \sigma - 1$ or G is large enough. When $\epsilon = \infty$ and $G = 1$, the isocost curve has a horizontal arm and the relative demand curve a vertical one.

[10] In the perfect substitute case, an increase in the number of non-automated products increases w_H, weakly increases w_L and, if $G < 1$ and N is large enough, decreases the skill premium.

an initial level close to 0. As a result, growth will progressively become unbalanced with a rising skill premium (and accordingly a decline in the labor share), and for some parameter values, low-skill wages will temporarily decline.

2.4 Asymptotics for general technological processes

We study the asymptotic behavior of the model for given paths of technologies and mass of high-skill workers in production. For any variable a_t (such as N_t), we let $g_t^a \equiv \dot{a}_t/a_t$ denote its growth rate and $g_\infty^a = lim_{t\to\infty} g_t^a$ if it exists. In Appendix 7.2.1 we derive:

Proposition 2. *Consider three processes $[N_t]_{t=0}^{\infty}$, $[G_t]_{t=0}^{\infty}$ and $[H_t^P]_{t=0}^{\infty}$ where $(N_t, G_t, H_t^P) \in (0,\infty) \times [0,1] \times (0,H]$ for all t. Assume that G_t, g_t^N and H_t^P all admit limits G_∞, g_∞^N and H_∞^P with $g_\infty^N > 0$ and $H_\infty^P > 0$.*

A) If $G_\infty \in (0,1)$, the asymptotic growth of high-skill wages w_{Ht} and output Y_t are:

$$g_\infty^{w_H} = g_\infty^Y = g_\infty^N / \left((1-\beta)(\sigma-1)\right), \tag{14}$$

and the asymptotic growth rate of w_{Lt} is given by

$$g_\infty^{w_L} = g_\infty^Y / \left(1 + \beta(\sigma-1)\right). \tag{15}$$

B). If $G_\infty = 1$, the asymptotic growth rates of w_{Ht} and Y_t also obey (14). If G_t converges sufficiently fast (so that $\lim_{t\to\infty} (1-G_t) N_t^{\psi(1-\mu)\frac{\epsilon-1}{\epsilon}}$ exists and is finite) then :

-i) If $\epsilon < \infty$ the asymptotic growth of w_{Lt} is positive at :

$$g_\infty^{w_L} = g_\infty^Y / \epsilon. \tag{16}$$

-ii) If low-skill workers and machines are perfect substitute then $\lim_{t\to\infty} w_{Lt}$ is finite and weakly greater than $\tilde{\varphi}^{-1}$ (equal to $\tilde{\varphi}^{-1}$ when $\lim_{t\to\infty} (1-G_t) N_t^{\psi} = 0$).

C) If $G_\infty = 0$ and G_t converges sufficiently fast (so that $\lim_{t\to\infty} G_t N_t^{\beta}$ exists and is finite), then the asymptotic growth rates of w_{Lt}, w_{Ht} and Y_t obey:

$$g_\infty^{w_L} = g_\infty^{w_H} = g_\infty^Y = g_\infty^N / (\sigma - 1). \tag{17}$$

This proposition first relates the growth rate of output and high-skill wages to the growth rate of the number of products. Without automation, that is if G_t converges to 0 sufficiently fast, Y_t is proportional to $N_t^{1/(\sigma-1)}$ as in a standard expanding-variety model.

Automation introduces machines as an additional reproducible input such that a higher level of productivity leads to a higher supply of machines further increasing output when $G_\infty > 0$. This multiplier effect is increasing in the asymptotic share of machines, β.

Second, with positive growth in N_t, mild assumptions are sufficient for asymptotic positive growth in low-skill wages. w_{Lt} only remains bounded when there is economy-wide perfect substitution, i.e. low-skill workers and machines are perfect substitutes, $\epsilon = \infty$, and all products are automated asymptotically (G_t converges to 1 sufficiently fast). Even then low-skill wages are bounded below by $\tilde{\varphi}^{-1}$, as a lower wage would imply that no firm would use machines. In general, the processes of N_t and G_t depend on the rate at which new products are introduced, the extent to which they are initially automated, and the rate of automation. As long as new non-automated products are continuously introduced, and the intensity at which non-automated firms are automated is bounded, the share of non-automated products is always positive, i.e. $G_\infty < 1$ (see proof in Appendix 7.2.2). This ensures that there is no economy-wide perfect substitution between low-skill workers and machines.

With aggregate imperfect substitution (because $G_\infty < 1$ or $\epsilon < \infty$), a growing stock of machines and a fixed supply of low-skill labor imply that the relative price of a worker (w_{Lt}) to a machine (p_t^x) must grow at a positive rate. Since machines are produced with the same technology as the consumption good, $p_t^x = p_t^C$, where p_t^C is the price of the consumption good (1 with our normalization), the real wage $w_{Lt} = w_{Lt}/p_t^C = (w_{Lt}/p_t^x)(p_t^x/p_t^C)$ must also grow at a positive rate.

Third, the proposition shows that as long as $G_\infty > 0$, low-skill wages cannot grow at the same rate as output. This is easily seen from the aggregate production function (3), which asymptotically is a nested CES with constant share parameters and where technological change in the form of an increase in N_t is not labor-augmenting (unless $G_\infty = 0$). Therefore, following Uzawa's theorem, balanced growth is impossible. Within our framework, this holds when automation intensity is bounded away from 0 which ensures that the asymptotic share of automated products is positive and rules out case C (see Appendix 7.2.2 for a proof).

If $\epsilon < \infty$ and $G_\infty = 1$ (sufficiently fast), low-skill workers derive their income asymptotically from automated firms and the asymptotic growth rate depends on the elasticity of substitution between machines and low-skill workers in automated firms, ϵ.

In contrast, when $G_\infty \in (0, 1)$, the demand for low-skill labor increasingly comes from the non-automated firms (as automation is labor-saving at the firm level). With growing wages, the relative market share of non-automated firms decreases in proportion with $(1+\varphi w_{Lt}^{\epsilon-1})^{-\mu} \sim \varphi^{-\mu} w_{Lt}^{-\beta(\sigma-1)}$, while most of the demand for high-skill labor comes from automated firms. Then, the growth rate of low-skill wages is a fraction of the growth rate of high-skill wages given by (15). The ratio between high-skill and low-skill wage growth rates increases with a higher importance of low-skill workers (a higher β) or a higher substitutability between automated and non-automated products (a higher σ) since both imply a faster loss of competitiveness of the non-automated firms. Yet, it is independent of the elasticity of substitution between machines and low-skill workers, ϵ or of the exact asymptotic share of automated products G_∞. In this case, non-automated products provide employment opportunities for low-skill workers which limits the relative losses of low-skill workers compared to high-skill workers (their wages grow according to (15) instead of (16) and $\epsilon > 1 + \beta(\sigma - 1)$).

2.5 Discussion

Proposition 2 establishes general conditions under which low-skill wages asymptotically grow but slower than high-skill wages. We now discuss the robustness of this result. First, our assumption that machines are an intermediate input is innocuous: Section 4 relaxes this assumption and lets machines take the form of capital with no qualitative change of result. Second, Appendix 6.4 relaxes the assumption of an exogenous stock of labor and considers a Roy model where workers are heterogeneous in the quantity of high-skill labor they can supply. Proposition 2 generalizes to this case, although the relative growth rate of low-skill wages is higher and asymptotically all workers supply high-skill labor. Third, Appendix 7.4 presents a model where the production technologies for machines and the consumption good differ allowing for negative growth in p_t^x/p_t^C. In this case, low-skill wages may decline asymptotically. Fourth, even if some of the tasks (but not all) performed by high-skill workers are automatable, our results would remain similar as long as high-skill workers remain essential in production.[11]

Appendix 6.1 breaks the assumption of symmetry and assumes that new products have a higher productivity. Higher productivity comes in the form of TFP improvements (which augments the productivity of all factors, including machines) or higher (low-skill and high-skill) labor productivity. Productivity increases exponentially, so we assume that N_t grows linearly to maintain non-explosive growth, and non-automated products

[11] If all labor tasks are automatable, infinite production is possible in finite time once N is large enough. Factors such as natural resources or land are then likely to be the scarce factor.

still have a positive probability of becoming automated. We show that, if all productivity improvements are labor-augmenting, horizontal innovation becomes more low-skill biased and balanced growth is possible; a result similar to that of Acemoglu and Restrepo (2017a). However, if machines also become partly more productive in new tasks (once those are automated), balanced growth is impossible. Therefore, the invention of new non-automated tasks is not enough to ensure balanced growth, what is required is that labor but not potential machines are more productive in these new tasks.[12]

3 Endogenous innovation

We now model automation and horizontal innovation as a the result of investment. This allows us first to look at the impact of wages on technological change (the reverse of Proposition 1), second to study the transitional dynamics of the system, and third to explore the interactions between the two innovation processes. Section 3.1-3.5 characterizes the model and its solution, section 3.6 relates it to historical experience and section 3.7 provides additional results and comparative statics.

3.1 Modeling innovation

A non-automated firm can hire $h_t^A(i)$ high-skill workers to perform automation research which will result in the firm becoming automated as a Poisson process with rate $\eta G_t^{\widetilde{\kappa}} \left(N_t h_t^A(i)\right)^{\kappa}$. Once a firm is automated it remains so forever. $\eta > 0$ denotes the productivity of the automation technology, $\kappa \in (0,1)$ measures the concavity of the automation technology, $G_t^{\widetilde{\kappa}}$, $\widetilde{\kappa} \in [0, \kappa]$, represents possible knowledge spillovers from the share of automated products, and N_t represents knowledge spillovers from the total number of intermediate inputs. The spillovers in N_t ensure that both automation and horizontal innovation may take place in the long-run; they exactly compensate for the mechanical reduction in the amount of resources for automation available for each product (namely high-skill workers) when the number of product increases.[13] Automation is

[12] We have also abstracted from the accumulation of low-skill human capital. “Traditional” human capital would be equivalent to augmenting low-skill labor in (3), and, from Uzawa’s theorem, would be insufficient to guarantee balanced growth when $G_\infty > 0$ and N_t grows exponentially. Grossman et al. (2017) show how non-traditional human capital can lead to balanced growth when technological change is not purely labor-augmenting. He and Liu (2008) show that the endogenous accumulation of skill through a “schooling” technology fits the rise in the skill-ratio and the skill-premium well.

[13] These spillovers can be micro-funded as follows: let there be a fixed mass one of firms indexed by j each producing a continuum N_t of products indexed by i so that production is given by $Y_t =$

undertaken by the incumbent firm, but we could accommodate automation by entrants as long as the incumbent also automates with positive probability or captures a share of the surplus created by the automation innovation.

New intermediate inputs are developed by high-skill workers in a standard manner according to a linear technology with productivity γN_t. With H_t^D high-skill workers pursuing horizontal innovation, the mass of intermediate inputs evolves according to:

$$\dot{N}_t = \gamma N_t H_t^D.$$

We assume that firms do not exist before their product is created and therefore cannot invest in automation. As a result, new products are born non-automated, which means that "horizontal innovation" corresponds to an increase in N_t keeping $G_t N_t$ constant and (following our discussion in section 3.7) is low-skill biased under certain conditions. This is motivated by the idea that when a task is new and unfamiliar, the flexibility and outside experiences of workers allow them to solve unforeseen problems. Only as the task becomes routine and potentially codefiable a machine (or an algorithm) can perform it (Autor, 2013). Our results carry through if only a share of the new products are born non-automated as discussed in section 3.7.[14]

Therefore the rate and direction of innovation will depend on the equilibrium allocation of high-skill workers between production, automation and horizontal innovation. Defining the total mass of high-skill workers working in automation as $H_t^A \equiv \int_0^{N_t} h_t^A(i)di$, we get that high-skill labor market clearing leads to

$$H_t^A + H_t^D + H_t^P = H. \tag{18}$$

3.2 Innovation allocation

We denote by V_t^A the value of an automated firm, by r_t the economy-wide interest rate and by $\pi_t^A \equiv \pi(w_{Lt}, w_{Ht}, 1)$ the profits at time t of an automated firm. The asset pricing

$(\int_0^1 \int_0^{N_t(j)} y_t(i,j)^{\frac{\sigma-1}{\sigma}} didj)^{\frac{\sigma}{\sigma-1}}$. When a firm hires $\widetilde{H}_t^A(j)$ high-skill workers in automation each of its non-automated products gets independently automated with a Poisson rate of $\eta G_t^{\widetilde{\kappa}}[\widetilde{H}_t^A(j)/(1-G_t(j)))]^{\kappa}$. The aggregate economy would be identical to ours and have the same social planner allocation (the decentralized equilibrium would qualitatively behave qualitatively similarly but the externality in the automation technology from the number of products would be internalized).

[14]The model predicts that the ratio of high-skill to low-skill labor *in production* is higher for automated than non-automated firms, though not *overall* since non-automated firms also hire high-skill workers for the purpose of automating. In particular, new firms do not always have a higher ratio of low to high-skill workers (and at the time of its birth a new firm only relies on high-skill workers).

equation for an automated firm is given by

$$r_t V_t^A = \pi_t^A + \dot{V}_t^A. \tag{19}$$

This equation states that the required return on holding an automated firm, V_t^A, must equal the instantaneous profits plus appreciation. An automated firm only maximizes instantaneous profits and has no intertemporal investment decisions to make.

A non-automated firm invests in automation. Denoting by V_t^N the value of a non-automated firm and letting $\pi_t^N \equiv \pi(w_{Lt}, w_{Ht}, 0)$, we get the asset pricing equation:

$$r_t V_t^N = \pi_t^N + \eta G_t^{\widetilde{\kappa}} \left(N_t h_t^A\right)^{\kappa} \left(V_t^A - V_t^N\right) - w_{Ht} h_t^A + \dot{V}_t^N, \tag{20}$$

where h_t^A is the mass of high-skill workers in automation research hired by a single non-automated firm. This equation is similar to equation (19), but profits are augmented by the instantaneous expected gain from innovation $\eta G_t^{\widetilde{\kappa}} \left(N_t h_t^A\right)^{\kappa} \left(V_t^A - V_t^N\right)$ net of expenditure on automation research, $w_{Ht} h_t^A$. This gives the first order condition:

$$\kappa \eta G_t^{\widetilde{\kappa}} N_t^{\kappa} \left(h_t^A\right)^{\kappa - 1} \left(V_t^A - V_t^N\right) = w_{Ht}. \tag{21}$$

h_t^A increases with the difference in value between automated and non-automated firms, and thereby current and future low-skill wages—all else equal.

Since non-automated firms get automated at Poisson rate $\eta G_t^{\widetilde{\kappa}} \left(N_t h_t^A\right)^{\kappa}$, and since new firms are born non-automated, the share of automated firms obeys:[15]

$$\dot{G}_t = \eta G_t^{\widetilde{\kappa}} \left(N_t h_t^A\right)^{\kappa} (1 - G_t) - G_t g_t^N. \tag{22}$$

Free-entry in horizontal innovation guarantees that the value of creating a new firm cannot be greater than its opportunity cost:

$$\gamma N_t V_t^N \leq w_{Ht}, \tag{23}$$

with equality whenever there is strictly positive horizontal innovation ($\dot{N}_t > 0$).

The low-skill and high-skill representative households' problems are standard and

[15] Using that by symmetry the total amount of high-skill workers hired in automation research is $H_t^A = (1 - G_t) N_t h_t^A$, we can also write this expression as $\dot{G}_t = \eta G_t^{\widetilde{\kappa}} \left(H_t^A\right)^{\kappa} (1 - G_t)^{1-\kappa} - G_t g_t^N$.

lead to Euler equations which in combination give

$$\dot{C}_t/C_t = (r_t - \rho)/\theta, \tag{24}$$

with a transversality condition requiring that the present value of all time-t assets in the economy (the aggregate value of all firms) is asymptotically zero:

$$lim_{t\to\infty}\left(exp\left(-\int_0^t r_s ds\right) N_t\left((1-G_t)V_t^N + G_t V_t^A\right)\right) = 0.$$

3.3 Equilibrium characterization

Following Proposition 2, asymptotically high-skill wages output and consumption grow proportionately to N_t^{ψ} where $\psi \equiv ((1-\beta)(\sigma-1))^{-1}$ when $G_\infty > 0$ (hence ψ is the asymptotic elasticity Y_t with respect to N_t). Therefore to study the behavior of the system we introduce the normalized variables $\hat{v}_t \equiv w_{Ht}N_t^{-\psi}$ and $\hat{c}_t \equiv c_t N_t^{-\psi}$. As h_t^A mechanically tends to 0 as the mass of non-automated firms grows we also introduce $\hat{h}_t^A \equiv N_t h_t^A$. We further define the auxiliary variable $\chi_t \equiv \hat{c}_t^{\,\theta}/\hat{v}_t$ which allows us to simplify the system (χ_t is related to the mass of high-skill workers in production and therefore, given $\hat{h}_t^A$, to H_t^D and the growth rate of N_t). Since the economy does not feature balanced growth, we also need to keep track of the level of N_t, we do this by introducing $n_t \equiv N_t^{-\beta/[(1-\beta)(1+\beta(\sigma-1))]}$, which tends toward 0 as N_t tends toward infinity. Finally, we define $\omega_t \equiv (w_{Lt}N_t^{-\psi/(1+\beta(\sigma-1))})^{\beta(1-\sigma)}$ which asymptotes a finite positive number. The equilibrium can then be characterized by a system of differential equations with two state variables n_t, G_t, two control variables, $\hat{h}_t^A, \chi_t$ and an auxiliary equation defining ω_t (see equations (31), (32), (39), (40), (45) and (47)-(49) in Appendix 6.2). This system admits a steady-state as stipulated below (proof in Appendix 7.5.1):

Proposition 3. *Assume that*

$$\kappa^{-\kappa}\left(\gamma(1-\kappa)/\rho\right)^{\kappa-1}\rho/\eta + \rho/\gamma < \psi H, \tag{25}$$

then the system of differential equations admits a steady state $(n^*, G^*, \hat{h}^{A*}, \chi^*)$ *with* $n^* = 0$, $0 < G^* < 1$ *and positive growth* $\left(g^N\right)^* > 0$.

We will refer to the steady state $(n^*, G^*, \hat{h}^{A*}, \chi^*)$ as as an asymptotic steady state for our original system of differential equations. In addition, the assumption that $\theta \geq$

1 ensures that the transversality condition always holds.[16] For the rest of the paper we restrict attention to parameters such that there exists a unique saddle-path stable steady state $(n^*, G^*, \hat{h}^{A*}, \chi^*)$ with $n^* = 0$, $G^* > 0$. Then, for an initial pair $(N_0, G_0) \in (0, \infty) \times [0, 1]$ sufficiently close to the asymptotic steady state, the model features a unique equilibrium converging towards it.[17]

The following section describes the equilibrium and explains how the economy reaches this asymptotic steady-state even starting far from it. We focus on the evolution of the automation incentives, which, we show, are crucially linked to the level of low-skill wages.

3.4 Innovation incentives along the transitional path

A distinctive feature of this economy is that the path of technological change itself will be unbalanced through the transitional dynamics. In the following, we elaborate on this by showing that the economy goes through three phases: a Phase 1 where the incentive to automate is low and the economy behaves close to a Romer model, a Phase 2 where automation increases the share of automated products G_t and a Phase 3 where the economy approaches the steady state. This section helps guide the intuition, and formal proofs of all the results mentioned are contained in Appendices 7.5.3, 7.5.4 and 7.5.5. These are "quantitative phases" in that the incentive to automate in Phase 1 is small but not exactly 0 and in Phase 3, G_t is only approximately constant.

Following (21), the mass of high-skill workers in automation $\left(H_t^A = (1 - G_t)\, N_t h_t^A\right)$ and therefore the automation intensity rate, given by $\eta G_t^{\widetilde{\kappa}} \left(H_t^A / (1 - G_t)\right)^{\kappa}$, depends on the ratio between the gain in firm value from automation $V_t^A - V_t^N$, and its effective cost namely the high-skill wage divided by the number of products w_{Ht}/N_t, that is:

$$H_t^A = (1 - G_t) \left(\kappa \eta G_t^{\widetilde{\kappa}} \frac{V_t^A - V_t^N}{w_{Ht}/N_t} \right)^{1/(1-\kappa)}. \tag{26}$$

Crucially, as the number of products in the economy increases, the ratio $\left(V_t^A - V_t^N\right) / \left(w_{Ht}/N_t\right)$ changes value. To see this, we combine (19), (20) and (21), and integrate over the dif-

[16] To understand equation (25), let the efficiency of the automation technology η be arbitrarily large such that the model approaches a Romer model with only automated firms. Then equation (25) becomes $\rho/\gamma < \psi H$, which mirrors the condition for growth in a Romer model with linear innovation technology. With a smaller η the present value of a new product is reduced and the condition is more stringent.

[17] Multiple asymptotic steady states with $G^* > 0$ are technically possible but are not likely for reasonable parameter values (see Appendix 7.5.2). In addition, with two state variables (n_t and G_t) saddle path stability requires exactly two eigenvalues with positive real parts. In our numerical investigation, for all parameter combinations which satisfy the previous restrictions, this condition was always met.

ference in value between an automated and a non-automated firm to get:

$$V_t^A - V_t^N = \int_t^\infty \exp\left(-\int_t^\tau r_u du\right)\left(\pi_\tau^A - \pi_\tau^N - \frac{1-\kappa}{\kappa} w_{H\tau} h_\tau^A\right) d\tau, \tag{27}$$

such that the difference in value between the two types of firms is given by the discounted difference of the profit flows adjusted for the cost and probability of automation. Further, by the assumptions of Cobb-Douglas and iso-elastic demand, high-skill wages (for given H_t^P) and aggregate profits are both proportional to aggregate output. Therefore, w_{Ht}/N_t is proportional to average profits: $w_{Ht}/N_t = \left[G_t \pi_t^A + (1-G_t)\,\pi_t^N\right] / \left[\psi H_t^P\right]$. As a result, the mass of high-skill workers in automation essentially depends on the discounted flow of profits of automated versus non-automated firms divided by the average profits made by firms. Intuitively — from equation (27) — with a positive discount rate, as a first approximation $V_t^A - V_t^N$ will move like $\pi_t^A - \pi_t^N$, so that one gets

$$\frac{V_t^A - V_t^N}{w_{Ht}/N_t} \tilde{\propto} \frac{\pi_t^A - \pi_t^N}{G_t \pi_t^A + (1-G_t)\,\pi_t^N} = \frac{1 - \left(1 + \varphi w_{Lt}^{\epsilon-1}\right)^{-\mu}}{G_t + (1-G_t)\left(1+\varphi w_{Lt}^{\epsilon-1}\right)^{-\mu}}, \tag{28}$$

where we used $\pi_t^A - \pi_t^N = \left[\left(1+\varphi w_{Lt}^{\epsilon-1}\right)^\mu - 1\right]\pi_t^N$ from (7) and where $\tilde{\propto}$ denotes "approximately proportional". This highlights low-skill wages (relative to the inverse productivity of machines $\tilde{\varphi}^{-1}$) as the key determinant of automation innovations. Note, that when $w_{Lt} \approx 0$ the incentive for automation innovation is very low, whereas when $w_{Lt} \to \infty$ it approaches $1/G_t > 0$. This price effect bears similarity with Zeira (1998), where the adoption of a labor-saving technology also depends on the price of labor.[18]

First Phase. When the number of products, N_t, is sufficiently low that w_{Lt} is small relative to $\tilde{\varphi}^{-1}$, the difference in profits between automated and non-automated firms is small relative to average profits. Following (26) and (28), the allocation of high-skill labor to automation, H_t^A, is low and automation intensity is low. Consequently, growth is driven by horizontal innovation and the behavior of the economy is close to that of a Romer model with a Cobb-Douglas production function between low-skill and high-skill labor, and both wages approximately grow at a rate $g_t^N/(\sigma-1)$. This corresponds to what we label as Phase 1 of the economy (naturally, if G_t is not initially low, it must depreciate during this period following equation (22)).

[18]Beyond a focus on different empirical phenomena (an increase in inequality vs. cross-country productivity differences), there are two important differences between our model and Zeira (1998). Zeira (1998) assumes exogenous technological progress (while we model endogenous innovation) and here the innovation cost changes over time while Zeira (1998) has constant adoption cost.

Second Phase. As w_{Lt} grows relative to $\tilde{\varphi}^{-1}$, the term $(V_t^A - V_t^N)/(w_{Ht}/N_t)$ increases, in fact, following (28) it grows like $\left(1 + \varphi w_{Lt}^{\epsilon-1}\right)^{\mu} - 1$ when G_t is low. This raises the incentive to innovate in automation. Without the externality in the automation technology ($\widetilde{\kappa} = 0$), (26) directly implies that H_t^A must raise significantly above zero, and with it the Poisson rate of automation, $\eta \left(H_t^A / (1 - G_t)\right)^{\kappa}$ and thereby the share of automated products, G_t. For $\widetilde{\kappa} > 0$, the depreciation in the share of automated products during Phase 1 might gradually make the automation technology less effective which can delay or even potentially prevent the take-off of automation.[19]

In line with Proposition 1, both the increase in G_t and N_t lead to an increase in the skill premium. As we show below, the low-skill wage may temporarily decline. We will label this time period where the share of automated products in the economy increases sharply the second phase (the transition between phases is smooth and therefore the exact limits are arbitrary). Arguably, this time period is the one where our model differs the most from the rest of the literature.

Third Phase. As the share of automated products G_t is no longer near zero, the gain from automation $V_t^A - V_t^N$ and its effective cost w_{Ht}/N_t grow at the same rate (the right-hand side in (28) is now close to $1/G_t$). As a result, the normalized mass of high-skill workers in automation research ($N_t h_t^A$) stays bounded (see (26)), and so does the Poisson rate of automation, which implies that G_t converges to a constant below 1.

The economy then converges toward the asymptotic steady-state which features a constant share of automated products strictly below zero and a constant growth rate in the number of products. Following Proposition 1, part A, the skill premium grows, but the continued existence of non-automated products ensures that low-skill wages grow at a positive rate asymptotically given by $g_{\infty}^{w_L} = g_{\infty}^{w_H} / \left(1 + \beta(\sigma - 1)\right)$. In this phase, the profits made by a non-automated firm are negligible relative to the horizontal innovation cost, therefore it is the prospect of future automation which guarantees the entry of new products. This plays an important role in the interaction between the two innovation technologies. Besides, the growth rate in the number of products is always lower in Phase 3 than in Phase 1 in particular because automation reduces the share of output which goes to profits of new firms (see Appendix 7.5.5 for more details). Though in Phase 3 the share parameters of the nested CES function are constant, the model continues to differ from a generic capital deepening model in that long-run growth is endogenized

[19] Automation takes off if either G_0 and N_0 are not too low or, for any values of $N_0, G_0 > 0$ whenever $\kappa(1-\beta) + \widetilde{\kappa} < 1$—see Appendix 7.5.4. If we were to assume instead that the automation technology is given by $max\left\{\eta G_t^{\widetilde{\kappa}}, \underline{\eta}\right\}\left(N_t h_t^A\right)^{\kappa}$, then automation would always take off.

and depends on its interaction with automation.

3.5 An illustration of the transitional dynamics

We now illustrate our previous result and further analyze the behavior of our economy through the use of numerical simulations.[20] Thereafter, we relate our theoretical results to the historical experience of the US economy. Unless, otherwise specified, and in line with the results of Sections 2.3 and 3.4, the broad patterns described below do not depend on specific parameter choices and we simply choose "reasonable" parameters (Table 1). Appendix 7.7.6 gives a systematic exploration of the parameter space and section 4 calibrates a richer model to the U.S. data.

Table 1: Baseline Parameter Specification

σ	ϵ	β	H	L	θ	η	κ	$\tilde{\varphi}$	ρ	$\widetilde{\kappa}$	γ	N_0	G_0
3	4	2/3	1/3	2/3	2	0.2	0.5	0.25	0.02	0	0.3	1	0.001

Baseline Parameters. Total stock of labor is 1 with $L = 2/3$ and $\beta = 2/3$ such that absent automation and if all high-skill workers were in production the skill premium would be 1. The initial mass of products is set low at $N_0 = 1$ to ensure we begin in Phase 1. The initial share of automated products is low, $G_0 = 0.001$, but would initially decline had we chosen a higher level. We set $\sigma = 3$ to capture an initial labor share close to 2/3. We set $\tilde{\varphi} = 0.25$ and $\epsilon = 4$, so that at $t = 0$, the profits of automated firms relative to non-automated firms are only 0.004%. The innovation parameters (γ, η, κ) are chosen such that GDP growth is close to 2% both initially and asymptotically. There is no externality from the share of automated products in the automation technology, $\widetilde{\kappa} = 0$. ρ and θ are chosen such that the interest rate is around 6% initially and asymptotically.

Figure 3 plots the evolution of the economy. Based on the behavior of the *automation expenditures* (Panel C) we delimit Phase 1 as corresponding to the first 100 years and Phase 2 as the period between year 100 and year 250.

Innovation and growth. Initially, low-skill wages and hence the incentive to automate—proportional to $(V_t^A - V_t^N)/(w_{Ht}/N_t)$—are low (Panel B) and so is the share of automated firms G_t (Panel C). With growing low-skill wages, the incentive to automate picks up a bit before year 100. Then the economy enters Phase 2 as automation expenses sharply increase (up to 4% of GDP). This leads to an increase in the share of automated

[20]We employ the so-called "relaxation" algorithm for solving systems of discretized differential equations (Trimborn, Koch and Steger, 2008). See Appendix 7.6 for details.

products G_t which eventually stabilizes at a level strictly below 1. There is no simple one-to-one link between automation spending and rising inequality in our model: here, automation spending is higher in Phase 3 than in Phase 2 (Panel C in figure 3), yet the growth in the skill premium is slower.[21]

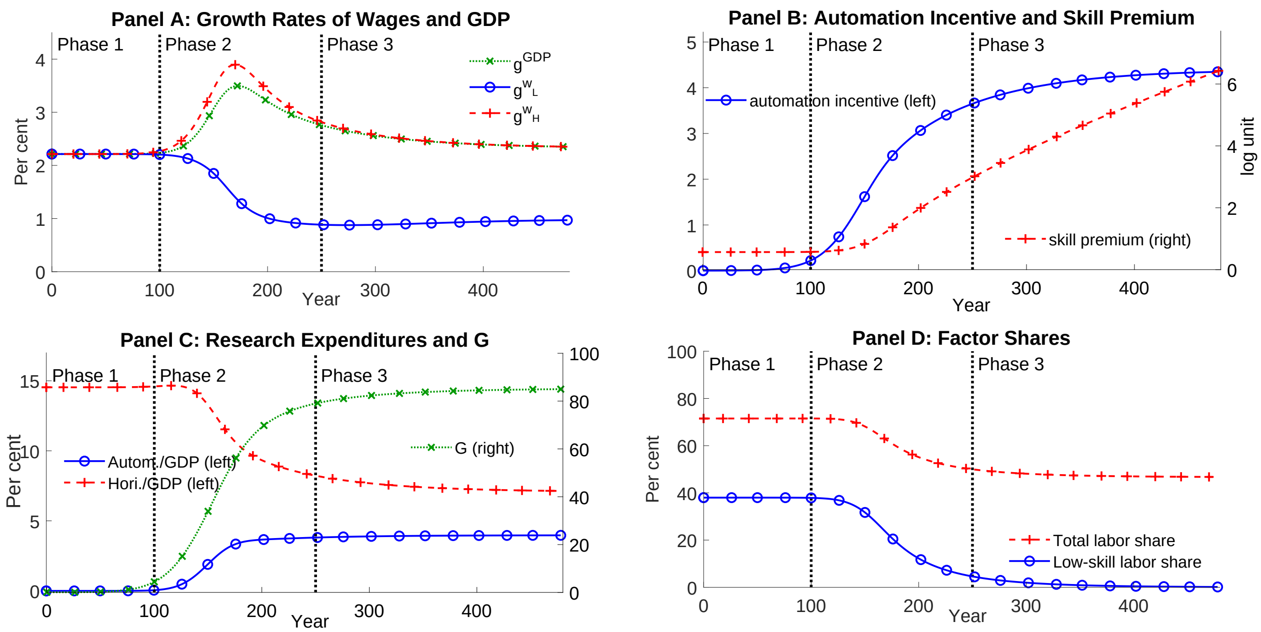

Figure 3: Transitional Dynamics for baseline parameters. Panel A shows growth rates for GDP, low-skill wages (w_L) and high-skill wages (w_H), Panel B the incentive to automate, $\left(V_t^A - V_t^N\right) / (w_{Ht}/N_t)$, and the skill premium, Panel C the total spending on horizontal innovation and automation as well as the share of automated products (G), and Panel D the wage share of GDP for total wages and low-skill wages.

As shown in Panel C, spending on horizontal innovation as a share of GDP declines during Phase 2 and for any parameter values ends up being lower in Phase 3 than Phase 1. Despite this, the growth rate of GDP is roughly the same in Phases 1 and 3 because the lower rate of horizontal innovation in Phase 3 is compensated by a higher elasticity of GDP wrt. N_t $(1/\left[(\sigma-1)(1-\beta)\right]$ instead of $1/(\sigma-1))$. As a result, the phase of intense automation—which also contributes to growth—is associated with a temporary boost of growth. This is, however, specific to parameters (Appendix 7.7.2 gives a counter-example). That intense automation need not be associated with a sharp increase in growth is important because the lack of an acceleration in GDP growth in recent decades has often been advanced in opposition to the hypothesis that a technological revolution

[21]Intuitively the elasticity of the skill-premium with respect to the skill-bias of technology is not constant in our model, contrary to a CES framework with factor-augmenting technologies.

explains the recent increase in the skill-premium (Acemoglu, 2002a).

Wages. In the first phase, and referring to figure 2, the relative demand curve stays close to the straight dotted line with slope $\frac{1-\beta}{\beta}\frac{L}{H^P}$ associated with a Cobb-Douglas production. Continuous horizontal innovation pushes the isocost curve towards the North East so that both wages grow at around 2% (Panel A).

As rising low-skill wages trigger the second phase, the relative demand curve pivots counter-clockwise and bends upwards increasing the growth rate of high-skill wages to almost 4% and suppressing the growth rate of low-skill wages to around 1%. Though our parameter values satisfy the conditions of Proposition 1 B.ii and any increase in G_t has a negative impact on w_{Lt}, the growth in N_t is sufficient to ensure that low-skill wages grow at a positive rate throughout. Section 3.7 demonstrates that this need not be the case. It is precisely this movement of the relative demand curve that allows our model to capture labor-saving innovation; a feature which a model with constant G or a capital deepening model would not include. In the third phase, the relative demand curve no longer moves, but the continuous push of the isocost curve keeps increasing the skill-premium albeit more slowly than previously.[22] Yet, at the same time, the decline in horizontal innovation reduces it. Appendix 7.7.1 presents a growth decomposition exercise where we compute the instantaneous contribution of horizontal innovation and automation to low-skill and high-skill wages.[23]

Factor shares. Panel D of figure 3 plots the labor share and the low-skill labor share. With machines as intermediate input, capital income corresponds to aggregate profits, which are a constant share of output. High-skill labor in production also earns a constant share of output. Both correspond to a rising share of GDP in Phase 2 as during this time period, the ratio Y/GDP increases since machines expenditures are excluded from GDP. Low-skill labor earns (close to) a constant share of output in Phase 1 but its share declines with automation in Phase 2 and approaches 0 in Phase 3. As a result, the labor share of GDP, constant in Phase 1, declines in Phase 2 and stabilizes at a

[22]Changes in the mass of high-skill workers in production, H_t^P, also affect the skill premium: increasing G_t requires hiring more high-skill workers in automation innovation (in the same vein as the General Purpose Technology literature; notably Beaudry et al. (2016)). Here, this effect is counteracted by a decrease in horizontal innovation and the net effect is small.

[23]At the aggregate level, our model boils down to a nested CES production function (see equation (3)), and Phase 2 corresponds to a period where the share parameter of the composite which features substitutability between machines and low-skill labor, G_t, rises. This change in the share parameter should not be confused with an increase in the elasticity of substitution between machines and low-skill labor (in fact, the Morishima's elasticity of substitution between these two factors is symmetric and declines in Phase 2 from a value close to ϵ to a value close to $1 + \beta(\sigma - 1)$).

lower level than before in Phase 3. Working contrary to this is the increase in innovation as only high-skill workers work in innovation, but the net effect is an increase in the capital share. For different parameter values, since H_t^P is not constant, the drop in the labor share can be delayed relative to the rise in the skill premium (see Appendix 7.7.3). The ratio of wealth to GDP also increases during Phase 2 and asymptotes a constant in Phase 3 (see Appendix 7.7.1).

3.6 Comparison with the historical experience

Our model suggests that technological progress is overall biased against low-skill labor, especially at later stages when low-skilled wages are sufficiently high. This bias manifests as a process of automation which causes a widening skill premium and a decline in the labor share. This is in line with the experience of the United States since the 1980s, where the college premium (considered to be a good proxy for the skill premium over that time period) has been steadily increasing (our stylized fact 1 in the introduction) and the labor share has declined by around 5 percentage points (stylized fact 3, see Karabarbounis and Neiman, 2013, or Elsby, Hobijn and Sahin, 2013). This contrasts our paper with most of the growth literature which features a balanced growth path and therefore does not have permanently increasing labor inequality.[24]

Extrapolating further in the past is more complicated, but it is worth discussing it briefly since automation did not start in the 1980s. Goldin and Katz (2008) argue that technological change has been skill-biased throughout the 20^{th} century as periods of decline in the skill premium (such as the 1970s) can be accounted for by exogenous changes in the relative supply of skills. Even before, Katz and Margo (2014) argue that the relative demand for highly skilled workers (in professional, technical and managerial occupations) has increased steadily from perhaps as early as 1820 to the present. This shift in relative demand may have been (partly) compensated for by changes in relative skill supply (see Appendix 6.4) and human capital accumulation (see Section 2.5) making growth more balanced. However, automation did not always replace the least skilled individuals: the mechanization of the 19^{th} century replaced skilled artisans, and the computerization of the last 30 years displaced more severely middle-skill workers. Therefore our model offers a framework to shed light on the historical experience

[24]For instance, in Acemoglu (1998), low-skill and high-skill workers are imperfect substitutes in production. Yet, since the low-skill augmenting technology and the high-skill augmenting technology grow at the same rate asymptotically, the relative stocks of effective units of low-skill and high-skill labor is constant, leading to a constant relative wage.

provided that the definition of 'high-skill' is narrow.

The capital share has followed a U-curve in the 20^{th} century (Piketty and Zucman, 2014 and Piketty, 2014), which our baseline model cannot account for. As the decline of the labor share in manufacturing is more persistent (9 percentage points from 1960 to 2005; Alvarez-Cuadrado, Van Long and Poschke, 2018), one way to reconcile our model with the historical pattern on factor shares would be to extend it to include multiple sectors (agriculture, manufacturing and services) with an elasticity of substitution less than 1, experiencing automation waves at different points in time. In such a model, as one sector automates, spending on the other sectors would increase (as in Acemoglu and Guerrieri, 2008) securing a higher labor share.

3.7 Additional results

We now exploit the endogenous nature of endogenous change in our model and examine the interaction between horizontal innovation and automation

Declining low-skill wages. Empirical evidence suggests that low-skill wages have been stagnating and perhaps even declining in recent periods (*Stylized fact 2*). Because our model features a labor saving innovation, it can accommodate declining low-skill wages in Phase 2, unlike a model with fixed G, a capital deepening model with perfectly elastic capital or in the previous DTC literature. Here, it happens when the relative demand curve pivots sufficiently fast counter-clockwise compared with the movement of the isocost curve in Figure 2. We can ensure this, for instance, by introducing externalities in the automation function and setting $\tilde{\kappa} = 0.49$. This delays the onset of automation, but intensifies it once it takes off, so that the relative demand curve moves quickly enough compared to the isocost curve for a decline in low-skill wages.[25] Yet, this drop must be temporary because lower low-skill wages discourage automation. See Appendix 6.3 for details and another example with $\tilde{\kappa} = 0$.

Effect of innovation parameters. In Appendix 7.5.6, we establish:

Proposition 4. *The asymptotic growth rates of GDP g_∞^{GDP} and low-skill wages $g_\infty^{w_L}$ increase in the productivity of automation η and horizontal innovation γ.*

Therefore, in the long-run, a better automation technology (a higher η) actually benefits low-skill wages: the reason is that firms automate faster which encourages horizontal innovation. During the transition, however, a higher η also means that Phase 2

[25] Horizontal innovation drops during this intense automation phase because its cost increases as there is a high demand for high-skill workers in automation innovation.

starts sooner, leading to lower low-skill wages at that point and a higher skill premium (see a numerical example in Appendix 7.7.4). These result preview those on the effect of taxes on automation in section 4.2.

Further results and extensions in the Appendix. i) Our model features elements of self-correction in the presence of exogenous shocks (as Acemoglu and Restrepo, 2017). For instance, with no automation externality ($\tilde{\kappa} = 0$), a positive exogenous shock to G_t will be followed by a period where automation is relatively less intense (as the skill premium would decline), so that eventually the asymptotic share of automated products stays the same (see Appendix 7.7.5). ii) Appendix 7.8 studies the social planner's problem. The optimal allocation is qualitatively similar to the equilibrium we described, which shows that our results are not driven by the market structure we imposed. iii) Appendix 7.9 presents a setting in which automation can only be undertaken before a firm starts production. The transition of the economy is qualitatively identical, which shows that the main results of the paper depend only on the feature that higher w_{Lt} creates incentives to automate and not on the assumption that firms are born non-automated, iv) Appendix 6.4 presents the transitional dynamics with an endogenous supply response.

4 Quantitative Exercise and Policy Experiments

In this section, we conduct a quantitative exercise to compare empirical trends for the United States with the predictions of our model. This allows us to dicipline the parameters of our model and subsequently conduct policy experiments.

4.1 Quantitative exercise

To match the data quantitatively, we need to modify the baseline model. First, since the share of high-skill workers has dramatically increased, we let H and L vary over time and here take as given their path from the data (as opposed to endogenizing labor supply as in Appendix 6.4). Second, we assume that producers rent machines from a capital stock. Capital can also be used used as structures in both automated and non-automated firms. Third, we allow for the possibility that low-skill workers are replaced by a composite of machines and high-skill workers. The production function (2) is then replaced by

$$y(i) = [l(i)^{\frac{\epsilon-1}{\epsilon}} + \alpha(i)(\tilde{\varphi} h_e(i)^{\beta_4} k_e(i)^{1-\beta_4})^{\frac{\epsilon-1}{\epsilon}}]^{\frac{\epsilon\beta_1}{\epsilon-1}} h_s(i)^{\beta_2} k_s(i)^{\beta_3}, \tag{29}$$

where $\beta_1 + \beta_2 + \beta_3 = 1$ and $\beta_4 \in [0, 1)$. The central difference between equations (2) and (29) is the introduction of $h_e(i)$ as high-skill labor which—along with machines—perform the newly automated tasks ("e" for equipment). This feature is necessary to capture a relatively low drop in the labor share. $k_s(i)$ is structures and $k_e(i)$ and $k_s(i)$ are both rented from the same capital stock K_t. K_t increases with investment in final goods and depreciates at a fixed rate Δ, so that (12) is replaced by:

$$\dot{K}_t = Y_t - C_t - \Delta K_t. \tag{30}$$

The cost advantage of automated firms now depends on the ratio between low-skill wage and the price of the high-skill labor capital aggregate namely $w_{Lt}/(w_{Ht}^{\beta_4}\tilde{r}_t^{1-\beta_4})$ where $\tilde{r}_t = r_t + \Delta$ is the gross rental rate of capital. The logic of the baseline model directly extends to this case. Proposition 1 still holds and Proposition 2.A) holds with $g_\infty^{w_H} = g_\infty^Y = g_\infty^N / \left((\beta_2 + \beta_1\beta_4)(\sigma - 1)\right)$ and $g_\infty^{w_L} = g_\infty^Y \left(1 + (\sigma - 1)\beta_1\beta_4\right) / \left(1 + \beta_1(\sigma - 1)\right)$. An equivalent to Proposition 3 holds but the system of differential equations includes three control variables and three state variables and Proposition 4 holds as well. The transitional dynamics are similar to that of the baseline model but automation innovation now depends on $w_{Lt}/(w_{Ht}^{\beta_4}\tilde{r}_t^{1-\beta_4})$. It is low in a first phase, increases in a second phase and stabilizes in the third phase as the economy approaches its asymptotic steady-state. The capital share and the capital output ratio increase in Phase 2 as equipment replaces low-skill labor in production. Details and proofs are provided in Appendix 7.11.

We match our extended model to the data (details on the data and the method in Appendix 7.12). Because of data availability and to make our exercise easily comparable to the rest of the literature, we identify low-skill workers with non-college educated workers and high-skill workers with college educated workers and focus on the years 1963-2007. We match the skill-premium and take the empirical skill-ratio as given (and normalize total population to 1). We also match the growth rate of real GDP/employment and the labor share. We further associate the use of machines with private equipment (excluding transport) and software. As pointed out by Gordon (1990) the NIPA price indices for real equipment are likely to understate quality improvements in equipment and therefore growth in the real stock of equipment. Hence, we use the adjusted price index from Cummins and Violante (2002) for equipment, and build (private) equipment and software to GDP ratios from 1963 to 2000.[26]

Our exercise has similarities with Goldin and Katz (2008) who attempt to explain the skill-premium using a constant trend in the skill bias of technical change or Krusell et al. (2000) who feed in the empirical path of equipment to explain the increase in the skill-premium. But, it is more demanding: both the technology path and the equipment stock must be endogenous.

Our model is not stochastic and cannot be directly estimated. Instead, we take a parsimonious approach and choose parameters to minimize the weighted squared log-difference between observed and predicted paths. We start the simulation 40 years before 1963 to force N_{1963} and G_{1963} to be consistent with the long-run behavior of our model.[27] Since the model requires the skill-ratio before and beyond the time period we estimate, we fit a generalized logistical function to the path of the log of the skill-ratio and use the predicted values outside 1963-2007 (over that time period, the fit is excellent).

The model features a total of 14 parameters ($\tilde{\varphi}$ can be normalized to 1 without loss of generality). Instead of exogenously restricting parameters we will fit the model with all parameters freely (other than the economically motivated boundaries imposed by the model itself) and then assess whether these parameter estimates fit with other similar estimates. Table 2 gives the resulting parameters. The elasticity of substitution across products σ is estimated at 6.7, consistent with Christiano, Eichenbaum and Evans (2005) who find that observed markups are consistent with a sigma of around 6. The elasticity of substitution between machines and workers is estimated at almost 5. $\tilde{\kappa}$ is estimated at 0.58 implying a substantial automation externality; a force which causes an accelerated Phase 2. Finally, we find β_1—the factor share of machines/low-skill workers—of 0.62 which implies sizable room for automation, though a β_4 of 0.73 means that the share of high-skill workers in the composite that replaces low-skill workers is of substantial importance. The preference parameters are within standard estimates with a ρ of 3.7% and the implied θ resulting in log-preferences. The only parameter that is estimated outside a common range is the depreciation rate Δ (but it is not precisely identified). Appendix 6.5 discusses in details how the parameters are identified.

[26]The series on the skill premium, the skill ratio, GDP/employment and the labor share directly predict a series for low-skill wages. Since this series differs substantially from the low-skill wages series of Acemoglu and Autor (2011) depicted in Figure 1 (particularly because of compositional effects), we do not attempt to match the latter.

[27]Initial values for G_t have little impact on the state of the economy several years later (see e.g. Figure 22 in Appendix 7.7.6). And the same is true for initial values of K_t. Therefore by simulating the economy 40 years prior, we ensure that the simulated moments are nearly independent of the initial values for G_t and K_t. We fix K_{1923} at its steady-state value in a model with no automation.

Table 2: Parameters from quantitative exercise

Parameter	σ	ϵ	β_1	γ	$\tilde{\kappa}$	θ	η	κ	ρ	β_2	Δ	β_4	N_{1963}	G_{1963}
Value	6.7	4.97	0.62	0.64	0.58	1	0.41	0.58	0.037	0.18	0.014	0.73	21.6	0.02

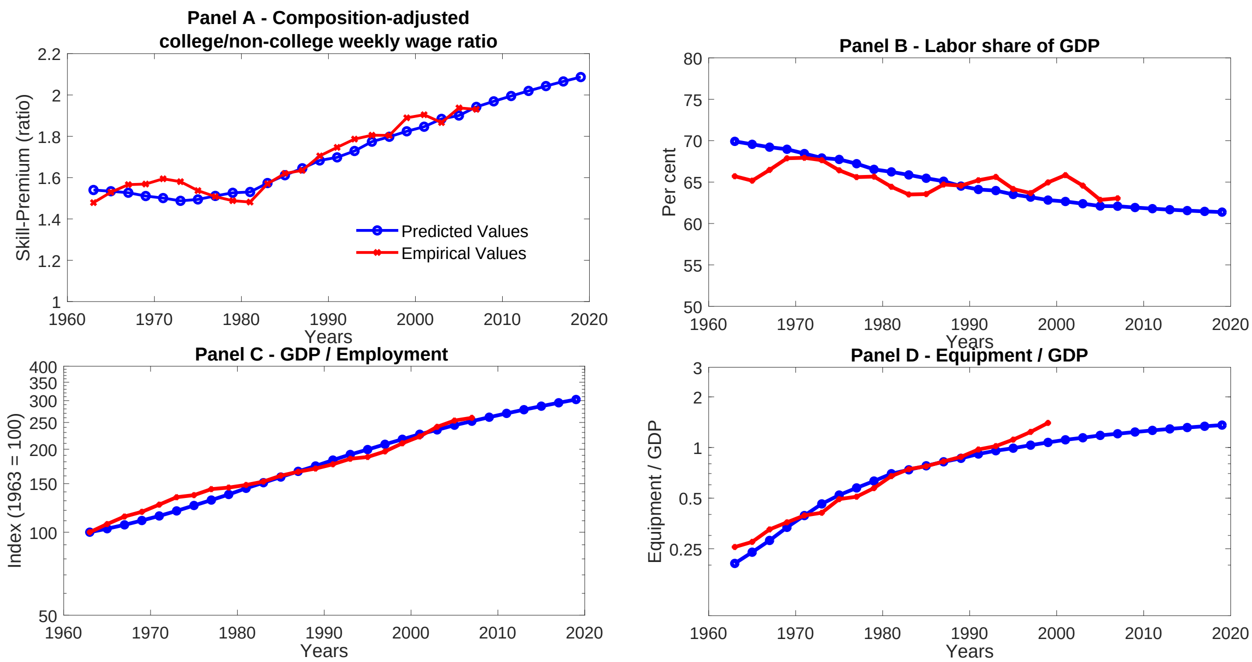

Figure 4: Predicted and empirical time paths

Figure 4 further shows the predicted path of the matched data series along with their empirical counterparts. Panel A demonstrates that the model matches the rise in the skill premium from the early 1980s and the flat skill premium in the period before reasonably well. Though less pronounced than in the data, our model also includes the more recent decline in the growth rate of the skill premium, which, computed over a 5 years moving window, peaks in 1984 at 1.3% and drops to 0.64% in the 2000s. The decline in the labor share is matched from 1970 onward. The average growth rate of the economy is matched completely as shown in Panel C. Although, the model largely captures the average growth rate of capital equipment over GDP during the period, the predicted path differs somewhat from its empirical counterpart as shown in Panel D on log-scale. Whereas the empirical path is close to exponential, the predicted path tapers off somewhat towards the end of the period.[28] In Appendix 6.5.5, we reproduce the same

[28] Interestingly, more recent data show a slow down in software investment (see Beaudry et al., 2016). Moreover, the ratio of equipment to GDP in the data is only a proxy for the ratio of machines to GDP in the model, since all equipment may not be used to replace low-skill workers.

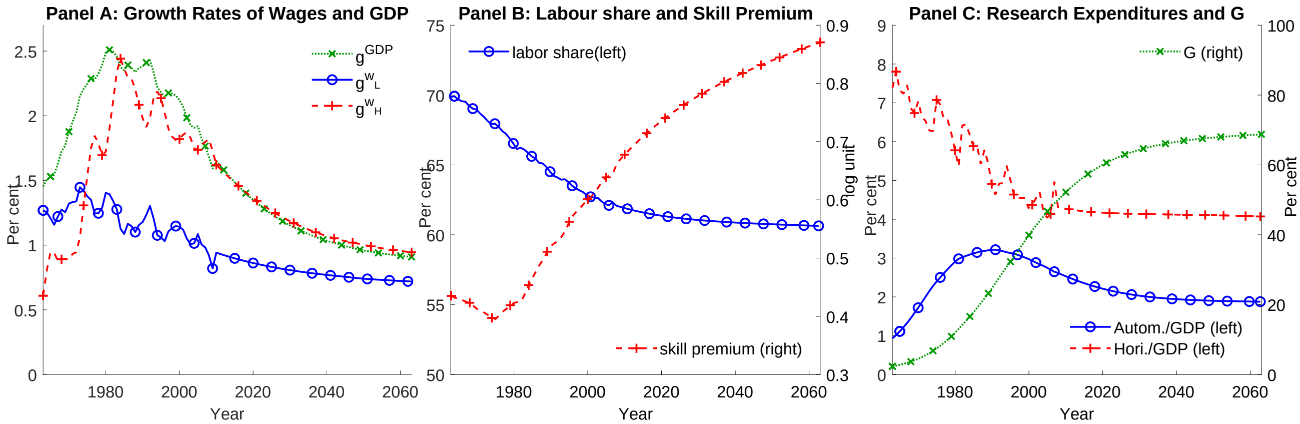

Figure 5: Transitional Dynamics with calibrated parameters (the growth rates are computed over a 5 year moving average).

exercise but only matching the first 30 years, the parameters are nearly identical and the calibrated model matches well the rest of the sample period.

Figure 5 plots the transitional dynamics from 1963 to 2063. Panel A shows that GDP growth slows down past 2007, Panel B that the skill premium keeps growing albeit at a slower rate than in the 1980s, while the labor share smoothly declines toward its steady-state value of 56%. Panel C shows that the share of automated products increases sharply through the 1963-2007 time period: In 1963 the value is 0.02 (yet this value is not precisely identified) but rises to 0.20 by the 1980s and 0.49 by 2007 to finally settle at 0.88 asymptotically—in this case, the marginal effect of automation unambiguously reduces low-skill wages on the entire path. Besides, the growth rate of the skill premium decreases in the 1990s and 2000s even though automation expenditures are roughly constant between 1980 and 2000. This constitutes a response to the critique of the literature on skill-biased technical change put forward by Card and DiNardo (2002), who argue that inequality rising the most in the early to mid 1980s and technological change continuing in the 1990s, squares poorly with the predictions of a framework based on skill-biased technological change.

4.2 Automation taxes

Among the many policy proposals to address rising income inequality, is a tax on the use of automation technology or a "robot tax". In the following we examine the consequences of two taxes: either on the use of machines—in the form of a tax on the rental rate of equipment—or the innovation of new machines—in the form of taxing high-skill workers in automation innovation. In either cases we consider the permanent unexpected introduction of a tax of 50% (See Appendix 7.11 for details).

First, consider a tax on the use of machines. To clarify the role of endogenous technology we also simulate the economy holding technology, N_t and G_t and therefore H_t^P at the baseline level. Figure 6 reports the results. The immediate effect is to discourage the use of machines and consequently low-skill wages rise by 4% on impact (Panel B) with a corresponding lower skill premium (Panel C). The endogeneity of technology amplifies the effect of the tax over time (in panel B, the gap between the endogenous and the exogenous cases widens). This results from two effects. First, the tax discourages automation innovation leading to a lower G (Panel E). Second, since high-skill workers and machines are complements, the tax reallocates high-skill workers toward horizontal innovation, increasing N (Panel D). Consequently, the positive effect on low-skill wages is eventually larger than the initial 4%. Even output will asymptotically be higher than the baseline (extending Panel A).[29]

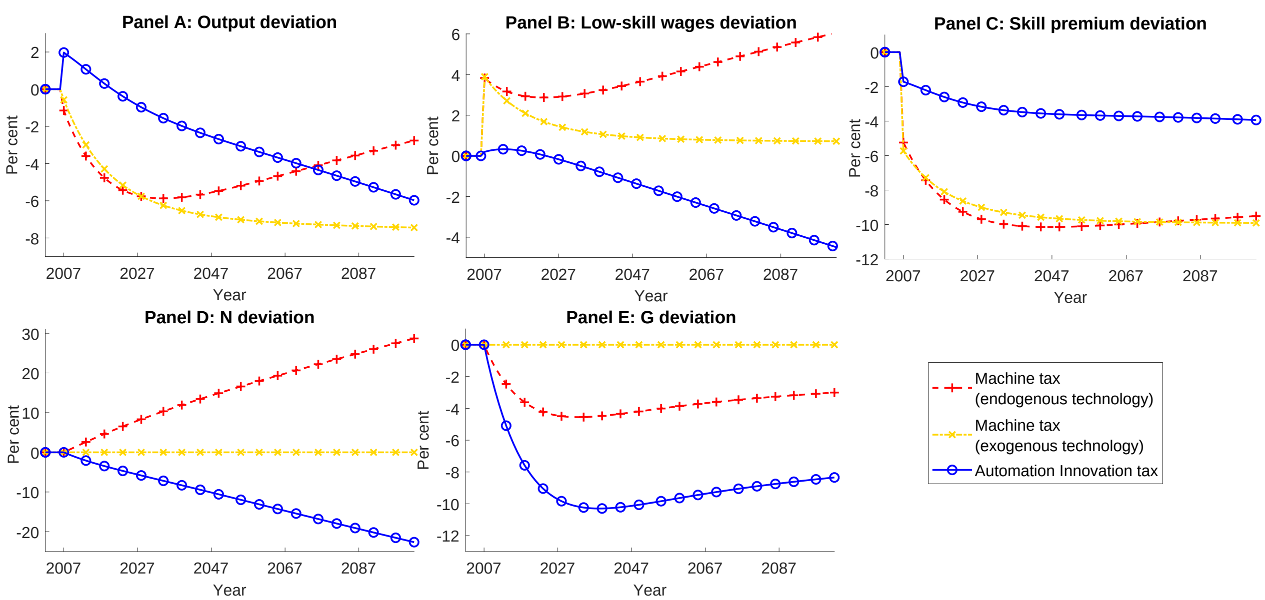

Figure 6: Effects of a machine tax and an automation innovation tax relative to baseline.

A tax on automation innovation has very different implications: First, high-skill workers reallocate from innovation in automation toward production which, on impact, boosts output and marginally low-skill wages. As the share of automated product G

[29]Asymptotically, a machine tax has no effect on G or on the growth rate of N: as using low-skill workers instead of machines becomes prohibitively expensive, the allocation of high-skill workers remains undistorted by the presence of a finite tax. See Proposition 12 in Appendix 7.11.

decreases, low-skill wages further increase though very modestly. However, discouraging automation innovation also discourages horizontal innovation which eventually reduces low-skill wages. The intuition is similar to that behind Proposition 4: a tax on automation innovation has similar effects to that of reducing the effectiveness of the automation technology. The skill-premium is reduced as the economy grows at a slower rate.

This exercise highlights the importance of endogenous technology: Though both forms of "robot" taxes increase low-skill wages on impact, the long-run effects depend crucially on whether the tax is designed to encourage or discourage overall innovation. Of course, this exercise is only a first pass and analyzing the welfare consequences of these policies or others, say minimum wage legislation, is of interest for future research.

5 Conclusion

This paper introduces automation in a horizontal innovation growth model. In such a framework, the economy undertakes a structural break. After an initial phase with stable income inequality and stable factor shares, automation picks up. During this second phase, the skill premium increases, low-skill wages stagnate and possibly decline, the labor share drops—all consistent with the US experience in the past 50 years—and growth starts relying increasingly on automation. In a third phase, the share of automated products stabilizes, but the economy still features a constant shift of low-skill employment from recently automated tasks to as of yet non-automated tasks. Low-skill wages grow in the long-run but slower than high-skill wages. A calibrated version of our model shows that a tax on machine use increases low-skill wages more than in a model with exogenous technology, while a tax on automation innovation eventually reduces low-skill wages.

A lesson from our framework is that if tasks performed by a scarce factor (say labor) can be automated but it is not presently profitable to do so, then, in a growing economy, the return to this factor will eventually increase sufficiently to make it profitable. In other words, there is a long-run tendency for technical progress to displace substitutable labor (a point made by Ray, 2014,), but this only occurs if the relevant wages are large relative to the price of machines. This in turn can only happen under three scenarios: either automation itself increases the wages of these workers (the scale effect dominates the substitution effect), or there is another source of technological progress (here, horizontal innovation), or technological progress allows a reduction in the price of machines relative to the consumption good (as in Appendix 7.4). Importantly, when machines are produced with a technology similar to the consumption good, automation can only reduce wages temporarily: a prolonged drop in wages would end the incentives to automate in the first place. Although, we focus on a general equilibrium model with low-skill labor, these insights extend to subsectors of the economy and other scarce factors.

Fundamentally, the economy in our model undertakes an endogenous structural change when low-skill wages become sufficiently high. This distinguishes our paper from most of the literature, which seeks to explain changes in the distribution of income inequality through exogenous changes: an exogenous increase in the stock of equipment as per Krusell et al. (2000), a change in the relative supply of skills, as per Acemoglu (1998), or the arrival of a general purpose technology as in the associated literature. This makes our paper closer in spirit to the work of Buera and Kaboski (2012), who argue that the increase in income inequality is linked to the increase in the demand for high-skill intensive services, which results from non-homotheticity in consumption.

The present paper is only a step towards a better understanding of the links between automation, growth and income inequality. Given that automation has targeted either low- or middle-skill workers and that artificial intelligence may now lead to the automation of some high-skill tasks, a natural extension of our framework would include more skill heterogeneity. Another natural next step would be to add firm heterogeneity and embed our framework into a quantitative firm dynamics model. Our framework could be used to study the recent phenomenon of "reshoring", where US companies that had offshored their low-skill intensive activities to China, now start repatriating their production to the US after having further automated their production process.

References

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

Acemoglu, D. (2002a). Directed Technical Change. *Review of Economic Studies*, 69(4):871–809.

Acemoglu, D. (2002b). Technical Change, Inequality, and the Labor Market. *Journal of Economic Literature*, 40(1):7–72.

Acemoglu, D. (2007). Equilibrium Bias of Technology. *Econometrica*, 75(5):1371–1409.

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

Acemoglu, D. and Autor, D. (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. *Handbook of Labor Economics*, 4(B).

Acemoglu, D. and Guerrieri, V. (2008). Capital Deepening and Nonbalanced Economic Growth. *Journal of Political Economy*, 116(3):467–498.

Acemoglu, D. and Restrepo, P. (2017a). The race between machine and man: Implications of technology for growth, factor shares and employment.

Acemoglu, D. and Restrepo, P. (2017b). Robots and jobs: Evidence from us labor markets.

Aghion, P., Deschezleprêtre, A., Hémous, D., Martin, R., and Van Reenen, J. (2016). Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, 214(1):1–51.

Aghion, P. and Howitt, P. (1996). Research and Development in the Growth Process. *Journal of Economic Growth*, 1(1):49–73.

Aghion, P., Howitt, P., and Violante, G. (2002). General Purpose Technology and Wage Inequality. *Journal of Economic Growth*, 7(4):315–345.

Alvarez-Cuadrado, F., Van Long, N., and Poschke, M. (2018). Capital-labor substitution, structural change and the labor income share. *Journal of Economic Dynamics and Control*, 87.

Autor, D. (2013). The “Task Approach” to Labor Markets: An Overview. *Journal for Labour Market Research*, 46(3):185–199.

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

Autor, D., Katz, L., and Kearney, M. (2006). The Polarization of the U.S. Labor Market. *American Economic Review*, 96(2):189–194.

Autor, D., Katz, L., and Kearney, M. (2008). Trends in U.S. Wage Inequality: Revising the Revisionists. *Review of Economics and Statistics*, 90(2):300–323.

Autor, D., Katz, L., and Krueger, A. (1998). Computing Inequality: Have Computers Changed the Labor Market? *Quarterly Journal of Economics*, 113(4):1169–1213.

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

Bartel, A., Ichniowski, C., and Shaw, K. (2007). How Does Information Technology

Really Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement and Worker Skills. *Quarterly Journal of Economics*, 122(4):1721–1758.

Beaudry, P. and Green, D. (2005). Changes in US wages, 1976-2000: ongoing skill bias or major technological change? *Journal of Labor Economics*, 23(3):609–648.

Beaudry, P., Green, D., and Sand, B. (2016). The great reversal in the demand for skill and cognitive tasks. *Journal of Labor Economics*, 34(S1):S199–S247.

Benzell, S., Kotlikoff, L., LaGarda, G., and Sachs, J. (2017). Robots are us: some economics of human replacement. Working Paper WP-785.

Buera, F. and Kaboski, J. (2012). The Rise of the Service Economy. *American Economic Review*, 102(6):2540–2569.

Card, D. and DiNardo, J. (2002). Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles. *Journal of Labor Economics*, 20(4):733–783.

Caselli, F. (1999). Technological Revolutions. *American Economic Review*, 89(1):78–102.

Christiano, E. and Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1).

Cummins, J. and Violante, G. (2002). Investment-specific technical change in the united states (1947-2000): Measurement and macroeconomic consequences. *Review of Economic Dynamics*, 5(2):243–284.

Elsby, M., Hobijn, B., and Sahin, A. (2013). The Decline of the U.S. Labor Share. *Brookings Papers on Economic Activity*, 47(2):1–63.

Feng, A. and Graetz, G. (2016). Rise of the machines: The effects of labor-saving innovations on jobs and wages. *Working Paper*.

Galor, O. and Moav, O. (2000). Ability-Biased Technological Transition, Wage Inequality, and Economic Growth. *Quarterly Journal of Economics*, 115(2):469–497.

Goldin, C. and Katz, L. (2008). *The Race between Education and Technology*. Harvard University Press.

Goos, M., Manning, A., and Salomons, A. (2009). Job Polarization in Europe. *American Economic Review*, 99(2):58–63.

Gordon, R. J. (1990). The measurement of durable goods prices. *NBER Monograph Series*.

Graetz, G. and Michaels, G. (2018). Robots at work. CEPDP 1330.

Grossman, G., Helpman, E., Oberfield, E., and Sampson, T. (2017). Balanced growth despite uzawa. *The American Economic Review*, 107(4):1293–1312.

Habakkuk, J. (1962). *American and British Technology in the Nineteenth Century.*

Cambridge University Press.

He, H. and Liu, Z. (2008). Investment-specific technological change, skill accumulation and wage inequality. *Review of Economic Dynamics*, 11.

Hémous, D. and Olsen, M. (2016). The rise of the machines: Automation, horizontal innovation and income inequality. CEPR Discussion Paper No 10244.

Hornbeck, R. and Naidu, S. (2014). When the levee breaks: Black migration and economic development in the american south. *American Economic Review*, 104(3):963–990.

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

Katz, L. and Margo, R. (2014). Technical change and the relative demand for skilled labor: The united states in historical perspective. In Boustan, L., Frydman, C., and Margo, R., editors, *Human Capital in History*, pages 15–57. University of Chicago and NBER.

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

Koh, D., Santaeulàlia-Llopis, R., and Zheng, Y. (2016). Labor share decline and intellectual property products capital. *Working Paper*.

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

Lewis, E. (2011). Immigration, skill mix and capital skill complementarity. *Quarterly Journal of Economics*, 126 (2):1029–1069.

Lloyd-Ellis, H. (1999). Endogenous Technological Change and Wage Inequality. *American Economic Review*, 89(1):47–77.

Nelson, R. and Phelps, E. (1966). Investment in Humans, Technological Diffusion, and Economic Growth. *American Economic Review*, 56(1/2):69–75.

Newell, R. G., Jaffe, A. B., and Stavins, R. N. (1999). The induced innovation hypothesis and energy-saving technological change. *The Quarterly Journal of Economics*, 114(3):941–975.

Nordhaus, W. (2007). Two Centuries of Productivity Growth in Computing. *Journal of Economic History*, 67(1):128–159.

Peretto, P. and Seater, J. (2013). Factor-Eliminating Technical Change. *Journal of Monetary Economics*, 60(4):459–473.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.

Piketty, T. and Zucman, G. (2014). Capital is Back: Wealth-Income Ratios in Rich Countries 1700-2010. *Quarterly Journal of Economics*, 129(3):1255–1310.

Ray, D. (2014). Nit-Piketty: A comment on Thomas Piketty's Capital in the Twenty First Century. Working paper, New York University.

Romer, P. (1990). Endogenous Technological Change. *Journal of Political Economy*, 98(5):71–S102.

Sachs, J. and Kotlikoff, L. (2012). Smart machines and long-term misery. NBER wp 18629.

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

Trimborn, T., Koch, K.-J., and Steger, T. (2008). Multi-Dimensional Transitional Dynamics: A Simple Numerical Procedure. *Macroeconomic Dynamics*, 12(3):301–319.

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