Technological Transitions with Skill Heterogeneity Across Generations*

Rodrigo Adão†
University of Chicago
Booth School of Business

Martin Beraja‡
Massachusetts Institute of Technology

Nitya Pandalai-Nayar§
University of Texas at Austin

August 23, 2019

Abstract

How do economies adjust to technological innovations? We develop a theory where overlapping generations of workers are heterogeneous over a continuum of technology-specific skills. Forward-looking investment decisions upon entry determine the worker's skill-type. Given a type's technology-specific wage, workers self-select into a technology. We show that this economy can be represented as a q-theory of skill investment. This allow us to sharply characterize the transitional dynamics and welfare implications of a technology-improving innovation. The adjustment is slower in economies with higher technology-skill specificity because the larger increases in relative wages induce larger, more persistent changes in the skill distribution across generations. We then empirically study the adjustment of developed economies to recent cognitive-biased technological innovations. We find strong responses of cognitive-intensive employment for young but not old generations. This suggests that cognitive-skill specificity is high and that the supply of cognitive skills is elastic at longer horizons. In such economies, ignoring the adjustment across generations by extrapolating from changes at short or long horizons alone leads to severe biases in the average and distributional welfare implications of technological innovations.

*We are grateful to Masao Fukui and Michelle Lam for excellent research assistance. For their valuable suggestions, we thank Daron Acemoglu, Fernando Alvarez, Arnaud Costinot, Ariel Burstein, V.V. Chari, Oliver Falck, Oded Galor, Erik Hurst, Greg Kaplan, Pablo Kurlat, Frank Levy, Simon Mongey, Fabrizio Perri, Chris Tonetti, Venky Venkateswaran, Gianluca Violante, and Fabrizio Zilibotti. We also thank seminar participants at Duke, MIT, NYU, Notre-Dame, UCSD, UT Austin, WashU, St. Louis Fed, UCLA, USC Marshall, IIES Stockholm, Munich, the Barcelona GSE Summer Forum, and the NBER SI Income Distribution and Macroeconomics, and Micro Data for Macro Models groups for helpful comments. Data access was provided by the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB). We thank Peter Brown and Clare Dingwell of the Harvard Economics Department, for facilitating access to IAB data. All results based on IAB microdata have been cleared for disclosure to protect confidentiality.

†Email: rodrigo.adao@chicagobooth.edu. Web: sites.google.com/site/rradao
‡Email: maberaja@mit.edu. Web: economics.mit.edu/faculty/maberaja
§Email: npnayar@utexas.edu. Web: www.nityapandalainayar.com

1 Introduction

New technologies are the key drivers of increases in living standards over long horizons. Yet, more recently, a literature has shown that they may have strong distributional consequences at shorter horizons.[1] If adjustment margins vary at different horizons, focusing on the consequences of technological innovations over short or long periods alone risks mismeasuring their overall impact on welfare and labor markets. In this paper, we develop a theory to study technological transitions where this is relevant because changes in the distribution of technology-specific skills happen slowly over generations. We then use this theory to empirically assess the implications of recent skill-biased innovations.

The theory has three distinct features. First, there are overlapping generations of workers with stochastic lifetimes, as in Yaari (1965) and Blanchard (1985). Second, within each generation, workers are heterogeneous over a continuum of skill types. A type determines the worker's productivity in the two technologies of the economy, as in Roy (1951). Given each type's technology-specific wages at a point in time, there is a threshold determining which skill types self-select into each of the two technologies. The output of the two technologies is then combined to produce a final consumption good. Third, given the expected future path for the wage distribution, workers make a costly investment upon entering the labor market that determines their skill type, similar to Chari and Hopenhayn (1991) and Caselli (1999). This gives rise to differences in technology-specific skill heterogeneity across generations.

The equilibrium of this economy is a joint path for the skill distribution, the assignment of skill types to technologies, and relative technology-specific wages and output. It entails a high-dimensional fixed-point problem: forward-looking entrants make skill investment decisions based on the expected future path for the wage distribution, which determine how the skill distribution evolves over time and, ultimately, the actual equilibrium path of the wage distribution and all other outcomes.

Our first result establishes that the approximate equilibrium of this economy can be represented as a q-theory of skill investment.[2] The path for the skill distribution is only a function of two variables at each point in time: the present-discounted value of log-relative wages in efficiency units (q) and the threshold determining the assignment of skills to technologies (which plays the role of the pre-determined variable). The equilibrium dynamics of these two variables is described by a simple system of linear differential equations. Our characterization of the equilibrium dynamics thus consists of "tracking" variables determining the evolution of the skill distribution rather than the high-dimensional skill distribution itself. Such an approach is reminescent of the one in Lucas and Moll (2014) and Perla and Tonetti (2014) that also reduce the analysis of the equilibrium dynamics of a high-dimensional endoge-

[1] See Durlauf and Aghion (2005) for a review of the literature on the impact of new technologies and innovation on long-run living standards. See Acemoglu and Autor (2011) and Autor and Salomons (2017) for reviews of the literature documenting the impact of new technologies on employment and wages of workers associated with different skills, occupations, industries, and firms.

[2] See Tobin (1969) and Hayashi (1982) for the original q-theory of capital investment.

nous object–the distribution of firm productivity– to the characterization of the evolution of a threshold over time.

Our second result derives in closed-form the transitional dynamics following a one-time, permanent increase in the productivity of all skill types employed in one of the technologies. We refer to this as a skill-biased technological innovation. The logic of the economy's adjustment follows immediately from the q-theory representation of the equilibrium. The relative productivity increase leads to an increase in the relative labor demand and wages in the improved technology. On impact, marginal skill-types now sort into that technology. The extent to which they do so crucially depends on how different skill types are in their relative technology-specific productivity. That is, the degree of technology-skill specificity determines the short-run skill supply elasticity (η). The increase in current and future relative wages leads younger entering generations to invest in skills that are complementary to the improved technology. The extent to which they do so crucially depends on a parameter (ψ) governing the cost of investing in technology-specific skills. That is, ψ determines the skill supply elasticity at longer horizons. Along the transition, q falls and relative output increases as younger generations replace older generations, expanding the supply of skills that became more valuable after the technological innovation.

The above result indicates that the impact of new technologies on the economy may significantly change over time as a result of differences between the skill supply elasticity at shorter versus longer horizons. Our third result then investigates how the interaction of the two distinct mechanisms shapes the economy's adjustment. It establishes that an economy where technology-skill specificity is higher has a slower adjustment path to the new long-run equilibrium. Specifically, it has more persistent dynamics of q and relative output as measured by their cumulative impulse responses. The q-theory analogy again delivers the intuition for this novel result. When technology-skill specificity is higher (and η is lower), the reallocation of worker skill-types across technologies is smaller following the skill-biased innovation and, therefore, the increase in relative wages is larger—as is the case in static assignment models. In our model though, the larger increase in relative wages implies that younger entering generations have stronger incentives to invest in the skills that became more valuable. As a result, there are larger differences in skill heterogeneity across generations. Thus, the economy's adjustment is slower because larger changes in the skill distribution take place as younger generations replace older generations. Importantly, a higher long-run skill supply elasticity (ψ) amplifies the impact of technology-skill specificity on the persistence of the economy's adjustment.

We conclude the theoretical part of the paper by analyzing how the persistence of the adjustment affects the average and distributional welfare consequences of a skill-biased technological innovation. In our theory, persistence arises from the dynamics of the skill distribution. So, in order to evaluate its consequences, we consider a static version of our model in which we shut down any skill investment of young workers. However, we allow this static

model to match labor market responses over a fixed horizon. When targeting short-run responses, this approach misses the inequality decline and the output increase that happens at longer horizons, thus understating the average welfare benefits and overstating the lifetime welfare inequality increases caused by the innovation. In contrast, targeting responses over long horizons leads to the opposite biases. Crucially, the magnitude of these biases are larger in economies that exhibit more persistent transition dynamics due to larger differences in the skill distribution across generations.

In the second part of the paper, we empirically assess the role that our two main theoretical mechanisms play in the adjustment of economies to recent skill-biased technological innovations. We start by connecting the parameters governing technology-skill specificity and the cost of skill investment to observable dynamic responses of worker allocations within and between generations. We do so by exploiting the closed-form expressions for the economy's transitional dynamics. Intuitively, for older generations of workers with a given skill distribution, the innovation-induced employment reallocation is larger if skills are less specific to each technology (i.e., η is higher). Relative to older generations, younger workers adjust their skills in response to the technological innovation. This generates between-generation differences in relative employment that are larger whenever the cost of investing in skills is lower (i.e. ψ is higher).

We then explore this insight to provide three pieces of evidence indicating that these two separate mechanisms are relevant to understand the economy's adjustment to the recent arrival of cognitive-biased technologies.[3] First, we analyze the employment trends in nine broad occupation groups in eighteen developed countries. We document that, in all countries, employment growth in the three most cognitive-intensive occupations was stronger for younger workers than for older workers. Second, we use microdata to provide a more detailed investigation of these responses in Germany. Controlling for a number of confounding factors, we show that employment and payroll grew more in occupations that require more time spent performing cognitive-intensive tasks. We find that the effect of cognitive intensity on these variables is strong for younger generations, but weak for older generations. Finally, following Falck, Gold, and Heblich (2014), we use pre-determined conditions of the German telephone network to obtain quasi-experimental variation across regions in the adoption timing of broadband internet in the early 2000s. By comparing late to early adopting regions, we estimate causal impulse response functions that show an increase in the relative employment and payroll of more cognitive-intensive occupations starting after 2005. The estimates are again different for older and young generations. The impact on relative employment is small and nonsignificant for older generations at all horizons, but it is positive and statistically significant for younger generations.

[3]Our approach follows an extensive literature documenting that the recent arrival of new technologies in the workplace, like the computer and the internet, augmented the productivity of jobs intensive in cognitive and analytical tasks while substituted jobs intensive in routine tasks —e.g. Autor, Levy, and Murnane (2003), Spitz-Oener (2006), Autor and Dorn (2013), Akerman, Gaarder, and Mogstad (2015), Acemoglu and Restrepo (2017), and, for a review, Acemoglu and Autor (2011).

Our empirical results suggest that skill supply elasticity is low at short horizons but it is high at longer horizons. Through the lens of our theory, such patterns arise whenever skills are very specific to cognitive-intensive activities (i.e., η is low) and the cost of accumulating skills is smaller for younger entering generations than for older incumbent generations (i.e., ψ is positive). As discussed above, this is precisely the environment that is likely to lead to substantial welfare biases from ignoring the slow adjustment across generations.

To quantify this, we calibrate our model to match our estimated dynamic responses for Germany. We consider a skill-biased innovation that increases the employment share of a technology from 25 percent to 50 percent across long-run equilibria. This is about the range of employment shares across developed economies. We find that the consumption equivalent average welfare increase across all generations is 52 percent and the lifetime welfare inequality increase is 38 percent. We then compute the biases of ignoring the adjustment across generations implied by our theory. If we match the same short-run increase in the employment share and only consider the changes on impact, we instead find a lower average welfare increase of 25 percent and a larger inequality increase of 49 percent. If we match the same long-run increase in relative employment and only consider changes across steady states, we instead find a higher average welfare increase of 70 percent and a lower inequality increase of 29 percent. These biases are smaller in economies with lower skill-specificity because the economy's adjustment is faster.

Related literature. Our paper is related to several strands of the literature. A long literature has analyzed the labor market consequences of technological innovations. We depart from the canonical framework in Katz and Murphy (1992) by modeling the supply of skills across technologies at different time horizons. Specifically, given the skill distribution at any point in time, the short-run skill supply to each technology arises from the static sorting decision of workers. This static assignment structure has been used in a recent literature analyzing how labor markets respond to a variety of shocks – e.g, Costinot and Vogel (2010), Acemoglu and Autor (2011), Hsieh, Hurst, Jones, and Klenow (2013), Burstein, Morales, and Vogel (2016), and Adão (2016). In addition, our theory entails slow-moving changes in skill supply that arise from the entry of young generations with different skills than those of previous generations, as in Chari and Hopenhayn (1991), Caselli (1999) and Galor and Moav (2002).[4] We show that the combination of these features yields tractable expressions for the equilibrium dynamics that resemble a q-theory of skill investment. We exploit the parsimony of our theory to establish that higher levels of technology-skill specificity and long-run skill supply elasticity generate slower adjustments following skill-biased innovations. We then link the two margins of skill supply in our theory to observable dynamic responses of labor

[4]Recent papers have documented that demand-driven shocks in relative wages affect educational attainment decisions – e.g., Atkin (2016) and Charles, Hurst, and Notowidigdo (Forthcoming). Our theory builds on the insight in Ben-Porath (1967) that workers make the bulk of their skill investment early in the life cycle, implying that young workers have a higher elasticity to changes in relative wages than old workers (e.g., Lee and Wolpin (2006)).

market outcomes within and across generations. Our empirical application indicates that separately allowing for these two forces is important in the context of the recent experiences of developed countries, in general, and Germany, in particular.

The main source of dynamics in our theory is the endogenous change in the supply of technology-specific skills over time. Several papers have proposed alternative sources of dynamics to study the transition following technological innovations, including sluggish labor mobility across sectors (Matsuyama, 1992), technology diffusion across firms (Atkeson and Kehoe, 2007), firm-level investment in R&D (Atkeson, Burstein, and Chatzikonstantinou, 2018), endogenous creation of new tasks for labor in production (Acemoglu and Restrepo, 2018), and permanent changes in the returns to wealth accumulation following increases in automation (Moll, Rachel, and Restrepo, 2019). Our paper complements this literature by analyzing empirically and theoretically how the endogenous dynamics of skill heterogeneity across generations affects the economy's adjustment to skill-biased technological innovations.

An extensive literature has estimated the distributional consequences of new technologies – for a review, see Acemoglu and Autor (2011). Our empirical analysis follows the literature showing the impact of new technologies on occupations with different task intensity – e.g., Autor, Levy, and Murnane (2003), Autor and Dorn (2013) and Acemoglu and Restrepo (2017). As in Akerman, Gaarder, and Mogstad (2015), we exploit regional characteristics to estimate the labor market consequences of broadband internet adoption. While they focus on the impact of this technology on the educational composition of employment in Norwegian firms, we estimate its effect on the occupation composition of employment in German regional labor markets. Similar to Card and Lemieux (2001) and Autor and Dorn (2009), we find that the impact of new technologies varies across worker generations. We complement this literature by showing that such short-run empirical analysis alone may miss part of the impact of new technologies on average welfare and lifetime inequality in economies with high technology-skill specificity, but it is an essential input in the measurement of the main mechanisms in our theory that control the transitional dynamics triggered by the arrival of new technologies.

Our paper is also related to the literature analyzing structural transformation in the form of long-run worker reallocation across sectors – e.g., Ngai and Pissarides (2007), Buera and Kaboski (2012), Herrendorf, Herrington, and Valentinyi (2015) and, for a review, Herrendorf, Rogerson, and Valentinyi (2014). Recently, Young (2014) and Lagakos and Waugh (2013) show that endogenous skill-sector sorting affects the process of structural transformation. Moreover, a number of papers have also emphasized the adjustment across generations. Kim and Topel (1995) document that the expansion of the manufacturing sector in Korea was driven almost entirely by new, young entrants to the labor force. Porzio and Santangelo (2019) documents substantial variation across countries in the extent to which the reallocation out of agriculture happens within- or between-cohorts. Hobijn, Schoellman, and Vindas

(2019) also document that many countries exhibit large between-generation differences in worker reallocation between agriculture, manufacturing and services. Relative to this literature, we make three contributions. First, we provide a tractable theory to analyze the role of skill heterogeneity within and across generations in the transitional dynamics following technological innovations. Second, we estimate impulse response functions to a technological innovation in Germany and show how they discipline the key parameters of our theory. Third, we point out which features of the economy (e.g., technology-skill specificity) lead to slow adjustment dynamics and, as result, large biases from welfare calculations that ignore them.

Outline. Our paper is organized as follows. Section 2 presents our model and establishes the q-theory representation of its equilibrium. In Section 3, we analyze the dynamic adjustment to skill-biased technological innovations. The welfare consequences implied by our theory are evaluated in Section 4. Section 5 links our theory to responses in observable outcomes for different generations of workers. It then presents our empirical analysis. Section 6 shows our quantitative analysis. Section 7 concludes.

2 A Model of Skilled-biased Technological Transitions

We propose a model with overlapping generations of workers that make forward-looking decisions to invest on a continuum of skills upon entry. Conditional on their skills, workers self-select to work with different technologies throughout their lives. Our model yields a q-theory of skill investment that determines the equilibrium path of the wage and output of workers employed with different technologies. As a function of these outcomes, we characterize the evolution of the economy's skill distribution and technology-skill assignment.

2.1 Environment

We consider a closed economy in continuous time indexed by t. There is a single final output whose production uses the input of two intermediate goods. The production technology of each intermediate good requires workers to perform a technology-specific task bundle. We denote the two technologies as high-tech ($k = H$) and low-tech ($k = L$). There is a continuum of worker skill types, $i \in [0, 1]$. The skill type determines the worker's productivity with each production technology.

Final product. Production of the final product is a CES aggregator of the two intermediate inputs:

$$Y_t = \left[(A_t Y_{Ht})^{\frac{\theta-1}{\theta}} + (Y_{Lt})^{\frac{\theta-1}{\theta}} \right]^{\frac{\theta}{\theta-1}} \tag{1}$$

where $\theta > 0$ is the demand elasticity of substitution between the low-tech and the high-tech intermediate products, and A_t is a shifter of the relative productivity of the high-tech input.

Conditional on the price of intermediate inputs, the cost minimization problem of firms producing the final product implies that the relative spending on the high-tech input is

$$y_t = \left(\frac{\omega_t}{A_t}\right)^{1-\theta}, \tag{2}$$

where $\omega_t \equiv \omega_{Ht}/\omega_{Lt}$ is the relative price of the high-tech input. We normalize the price of the low-tech input to one, $\omega_{Lt} \equiv 1$.

We consider a competitive environment, so that profit maximization implies that the equilibrium price of the final output is

$$P_t = (1 + y_t)^{\frac{1}{1-\theta}}. \tag{3}$$

In the model, y_t measures the relative importance of high-tech production in total output, and ω_t is the relative value of high-tech production. In the next section, we analyze how these outcomes adjust following unanticipated permanent changes in A_t. As in Katz and Murphy (1992), A_t is a shifter of the relative productivity of the high-tech input.

Allocation of skills to technology. We assume that a worker's skill type determines her productivity with the two technologies in the economy. For a worker of type i, $\alpha(i)$ is the overall productivity and $\sigma(i)$ is the differential productivity in high-tech production. Specifically, we assume that the production function of the low-tech good is

$$X_{Lt} = \int_0^1 \alpha(i) s_{Lt}(i) di, \tag{4}$$

and that of the high-tech good is

$$X_{Ht} = \int_0^1 \alpha(i)\sigma(i) s_{Ht}(i) di, \tag{5}$$

where $s_{kt}(i)$ is the density function of workers employed with technology k at time t.[5]

We assume a competitive labor market with zero profit in low-tech and high-tech production. In equilibrium, the wage rates of skill type i with the H and L technologies are respectively given by

$$w_{Ht}(i) = \omega_t \sigma(i)\alpha(i) \quad \text{and} \quad w_{Lt}(i) = \alpha(i). \tag{6}$$

As in Roy (1951), workers self-select across technologies to maximize labor income. Thus,

[5] Appendix C.1 provides a microfoundation for the production functions in (4)–(5). In this environment, production of each good combines the output of workers where each worker's output is a Cobb-Douglas function of cognitive and non-cognitive tasks performed on the job. Production of the H and L goods require different shares of cognitive and non-cognitive tasks. The function $\sigma(i)$ is determined by each skill-type's ability to perform task bundles.

the wage of a worker with skill type i is

$$w_t(i) = \max\{\omega_t \sigma(i), 1\}\alpha(i). \tag{7}$$

Recent papers have considered similar structure of endogenous sorting of workers to different technologies – e.g., Acemoglu and Autor (2011), Costinot and Vogel (2010), Adão (2016). As each point in time t, the sorting decision in (7) generates a mapping of skill types to technology. Such a mapping depends on the relative value of one unit of effective labor employed in high-tech production, ω_t. In the model, ω_t is a natural measure of inequality as it captures the relative wage rate of skill types employed in different technologies conditional on their productivity. The sorting decision also depends on the exogenous relative productivity of different skill types in the H technology. That is, it depends on the function $\sigma(i)$ that measures type i's comparative advantage in high-tech production. Without loss of generality, it is always possible to order skill types such that high-i types have a higher relative productivity in high-tech production. Thus, the slope of the function $\sigma(i)$ captures the degree of technology-skill specificity: how much the relativity productivity in technology H varies across skill types. To simplify the analysis, we assume that $\sigma(i)$ is strictly increasing.

Skill investment. We consider an overlapping generations setting in which the birth and death of workers follows a Poisson process with rate δ. At each point in time, workers use their labor earnings to purchase the final good. Utility is the present value of a logarithmic flow utility discounted by a rate ρ. For a worker of type i born at time t, lifetime utility is

$$V_t(i) = \int_t^\infty e^{-(\rho+\delta)(s-t)} log\left(\frac{w_s(i)}{P_s}\right) ds. \tag{8}$$

Crucially, as in Chari and Hopenhayn (1991) and Caselli (1999), we allow workers to acquire different skills at birth taking into account the value of future earnings streams. Given the wage function $\{w_s(i)\}_{s>t}$, workers born at time t can pay a utility cost to select a lottery $\tilde{s}_t(i)$ over skill types. If they do not pay the cost, their type is drawn from an exogneous distribution of innate ability, $\bar{s}_t(i)$.[6] A worker's type is then fixed during their lifetime. Intuitively, workers can pay an ex-ante cost to invest on human capital that allows them to target specific skill types. The skill investment may occur through schooling or on-the-job while workers are young. Yet, we assume that there is uncertainty regarding the return of the investment in terms of the specific skill realization.[7]

Formally, we assume that the cost of the lottery is proportional to the Kullback-Leibler

[6] Appendix C.2 considers the case of "learning from others" by allowing $\bar{s}_t(i)$ to depend on the current skill distribution in the economy $s_t(i)$. Specifically, we assume that $\bar{s}_t(i) = (s_t(i))^\gamma (\bar{s}(i))^{1-\gamma}$ where γ governs the strength of the "learning-from-others" effect: when γ is higher, it is easier for workers to target skills that are already abundant in the economy. As in Chari and Hopenhayn (1991), this introduces complementarity in the skill investment decision of different generations.

[7] Our approach is consistent with the evidence in Carneiro, Heckman, and Vytlacil (2011) that, conditional on observable educational investment, individuals with different unobserved skills have heterogeneous educational returns. Our model treats this heterogeneity in unobserved skills through the uncertainty of the type realization. This approach is isomorphic to an environment where individuals with heterogeneous costs of skill acquisition select their skill type.

divergence between the lottery $\tilde{s}_t(i)$ and the baseline distribution $\bar{s}_t(i)$.[8] Specifically, the cost of lottery $\tilde{s}_t(i)$ is

$$\frac{1}{\psi}\int_0^1 log\left(\tilde{s}_t(i)/\bar{s}_t(i)\right)\tilde{s}_t(i)di, \qquad \psi > 0.$$

Thus, the skill investment problem is

$$\max_{\tilde{s}_t(.):\int_0^1 \tilde{s}_t(i)di=1} \int_0^1 \left(V_t(i) - \frac{1}{\psi} log\left(\frac{\tilde{s}_t(i)}{\bar{s}_t(i)}\right)\right)\tilde{s}_t(i)di. \tag{9}$$

The parameter ψ governs the cost of targeting particular skill types. In the limit when $\psi \to 0$, the cost of targeting a particular skill is infinite and the economy's skill distribution does not respond to changes in the lifetime earnings of different skill types. Whenever $\psi > 0$, the optimal lottery $\tilde{s}_t(i)$ endogenously responds to the relative present discounted value of different skill types, $V_t(i)$.

Equilibrium. The assumption that only new cohorts have the ability to choose a skill lottery implies that the evolution of the skill distribution $s_t(i)$ follows the Kolmogorov-Forward equation,

$$\frac{\partial s_t(i)}{\partial t} = -\delta s_t(i) + \delta\tilde{s}_t(i). \tag{10}$$

Finally, the economy's equilibrium must satisfy market clearing for all t. By Walras law, it is sufficient to only consider the relative demand and supply of the high-tech intermediate input:

$$y_t = \omega_t x_t \tag{11}$$

where y_t is given by (2) and x_t is the ratio of the production functions in (4)–(5).

Definition 1 (Competitive Equilibrium) *Given an initial skill distribution $s_0(i)$, a competitive equilibrium is a path of the technology-skill assignment $\{G_t(i) : i \in [0,1] \to \{H,L\}\}_t$, the skill distribution $\{s_t(i)\}_t$, the skill lottery $\{\tilde{s}_t(i)\}_t$, the relative price and ideal price index $\{\omega_t, P_t\}_t$, such that*

1. *Given ω_t, the technology-skill assignment is given by the worker self-selection decision in (7).*
2. *The skill distribution $s_t(i)$ satisfies the Kolmogorov-Forward equation (10) with the skill lottery $\tilde{s}_t(i)$ given by the skill investment decision of new cohorts implied by (9).*
3. *The markets for high-tech and low-tech goods clear by satisfying (11). The final good market clears by satisfying (3).*

[8] This cost function has a long tradition in macroeconomics that uses it as a tractable form of comparing distance between distributions – e.g., frameworks with rational inattention (Sims (2003)) and model uncertainty (Hansen and Sargent (2008)). As discussed below, in our setting, this function yields the continuous type analog of the skill acquisition solution in an environment in which worker's ability to acquire a discrete number of skills follows a Type 1 extreme-value distribution.

2.2 Static and Dynamic Equilibrium Conditions

We now proceed to characterize the equilibrium in two steps. First, we consider the static conditions that, given the skill distribution $s_t(i)$ at time t, determines the relative wage and output of the high-tech input, $\{\omega_t, y_t\}_t$, that clears the intermediate good market. Second, we consider dynamic conditions that, given the path of the relative wage $\{\omega_t\}_t$, determine the optimal skill lottery chosen by young cohorts $\{\tilde{s}_t(i)\}_t$ and the evolution of the skill distribution $\{\tilde{s}_t(i)\}_t$.

Static equilibrium conditions. The endogenous sorting decision in (7) determines the assignment of skill types to technologies. We graphically represent this decision in Figure 1. The solid and dashed lines represent the potential log-wage of workers of type $i \in [0,1]$ if employed in high- and low-tech production respectively. Equation (7) implies that types self-select to work with the technology that yields the highest labor earnings. Thus, high-i (low-i) types receive higher relative earnings in high-tech (low-tech) production and choose to be employed with that technology. In equilibrium, the assignment is described by a threshold l_t characterizing the type that is indifferent between working with any of the two technologies.[9] That is,

$$\omega_t \sigma(l_t) = 1. \tag{12}$$

We summarize this discussion in the following lemma.

Lemma 1 (Equilibrium Assignment) *Worker types $i \leq l_t$ are employed in low-tech production with labor income of $w_t(i) = \alpha(i)$. Worker types $i > l_t$ are employed in high-tech production with labor income of $w_t(i) = \omega_t \sigma(i)\alpha(i)$. The threshold is determined by the indifference condition in (12).*

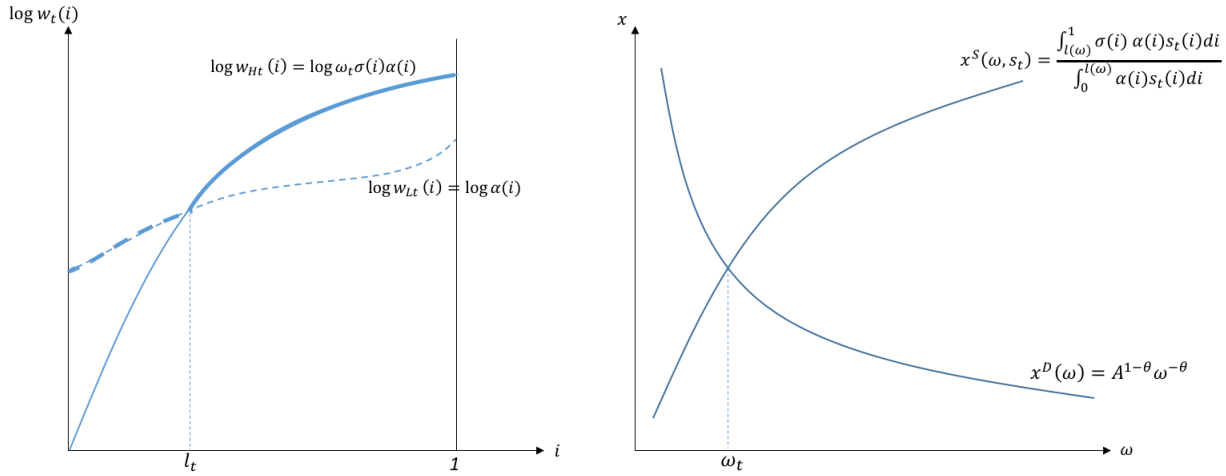
Lemma 1 links the relative wage ω_t to the allocation of skill types across technologies. Condition (12) is central to understand the impact of technological shocks on the allocation of workers across technologies. It implies that the inverse elasticity of $\sigma(i)$ determines how much workers reallocate across technologies in response to changes in the relative wage. We refer to the elasticity of $\sigma(i)$ as the technology-skill specificity. The inverse elasticity of $\sigma(i)$ plays the role of short-run skill supply across technologies, which we formally define as

$$\eta \equiv \left| \frac{\partial \log l_t(\omega_t)}{\partial \log \omega_t} \right| = \left(\frac{\partial \log \sigma(l_t)}{\partial \log i} \right)^{-1}.$$

where $l_t(\omega_t)$ is the implicit function defined by (12).

[9] Notice that, since $\sigma(i)$ is strictly increasing, the two curves in Figure 1 must cross at most once. Since the demand function in (2) goes to infinity as ω_t goes to zero, there is positive employment with both technologies in equilibrium and, therefore, the two curves in Figure 1 must cross at least once. Figure 1 illustrates the case in which high-i types are more productive in absolute terms, receiving higher earnings if employed with any of the two technologies. Adão (2016) shows that, depending on the shape of $\alpha(i)$, there are different possible configurations for Figure 1 which determine the correlation between income and technology allocation across skill types.

Figure 1: Static equilibrium



Panel A: Equilibrium Assignment

Panel B: Market Clearing

The technology-skill assignment in Lemma 1 determines the relative supply of high-tech production as a function of the relative wage ω_t. Conditional on the skill distribution $s_t(i)$, equations (4)–(5) imply that the relative supply is

$$x_t(\omega_t, s_t) = \frac{\int_{l_t(\omega_t)}^{1} \sigma(i)\alpha(i)s_t(i)di}{\int_0^{l_t(\omega_t)} \alpha(i)s_t(i)di}. \tag{13}$$

The relative price ω_t is then determined by the market clearing condition in (11). We illustrate this condition in Panel B of Figure 1. Whenever ω_t is low, relative demand for input H is high, but its relative supply is low due to the small share of types employed in high-tech production. Whenever ω_t is high, the opposite is true. In equilibrium, relative demand and supply are equalized. The following lemma formalizes the existence of a unique equilibrium value of ω_t for any given distribution $s_t(i)$.

Lemma 2 (Equilibrium Threshold) *Given the skill-type distribution $s_t(i)$ at time t, there is a unique equilibrium relative wage such that*

$$A_t^{\theta-1}\omega_t^{-\theta}\int_0^{l_t(\omega_t)} \alpha(i)s_t(i)di = \int_{l_t(\omega_t)}^{1} \alpha(i)\sigma(i)s_t(i)di, \tag{14}$$

and $l_t(\omega_t)$ is the implicit function defined by (12).

Proof. See Appendix A.1. ■

Dynamic equilibrium conditions. We now turn to the entrant's forward-looking problem of choosing their skill lottery $\tilde{s}_t(i)$ conditional on the future path of the relative technology-specific wage $\{\omega_s\}_{s>t}$. The solution of the maximization problem in (9) immediately yields the following optimal lottery.

Lemma 3 (Equilibrium Lottery) *Define* $log\left(Q_t(i)\right) \equiv \int_t^\infty e^{-(\rho+\delta)(s-t)} \max\{log\left(\omega_s \sigma(i)\right), 0\} ds$. *The equilibrium lottery is*

$$\tilde{s}_t(i) = \frac{\bar{s}_t(i)\alpha(i)^{\frac{\psi}{\rho+\delta}} Q_t(i)^{\psi}}{\int_0^1 \bar{s}_t(j)\alpha(j)^{\frac{\psi}{\rho+\delta}} Q_t(j)^{\psi} dj} \tag{15}$$

Proof. See Appendix A.2. ■

The optimal lottery in (15) is an extension of the multinomial logit function for a continuum of types. It shows that the investment on high-i types is a function of the present value of the relative wage in high-tech production as captured by $Q_t(i)$. The parameter ψ governs the sensitivity of the optimal lottery to changes in relative lifetime earnings. To see this more clearly, consider the stationary equilibrium with $\omega_t = \omega$ such that

$$s(i) = \tilde{s}(i) = \frac{\bar{s}(i) W(i)^{\psi}}{\int_0^1 \bar{s}(j) W(j)^{\psi} dj} \tag{16}$$

where $W(i) = \left(\alpha(i) \max\left\{\omega\sigma(i), 1\right\}\right)^{\frac{1}{\delta+\rho}}$ is the present value of the utility flow of skill type i.

In this case, the skill distribution is a constant-elasticity function of relative wages across types, where the elasticity is ψ.[10] Thus, a higher ψ implies that the long-run supply of high-i types is more sensitive to changes in the relative wage in high-tech production. Accordingly, we refer to ψ as the long-run skill supply across technologies, which we formally define as

$$\psi \equiv \frac{\partial \log s(i)/s(i')}{\partial \log W(i)/W(i')}.$$

2.3 Equilibrium Characterization: a q-theory of skill investment

We now combine the static equilibrium condition in (14) and dynamic equilibrium condition in (15) to characterize the equilibrium path of the relative technology-specific wage rates, $\{\omega_t\}_t$, and the skill distribution, $\{s_t(i)\}_t$. We begin by noting that the only endogenous state variable in our model is the skill distribution $s_t(i)$. In addition, the only control variable is the optimal lottery, $\tilde{s}_t(i)$, that is a function of the path of relative wage $\{\omega_s\}_{t>s}$. This implies that the dynamic equilibrium of the economy entails a fixed-point problem, since the market

[10] Notice that the log-run equilibrium of our model is a generalization with a continuum of types of the extension of the assignment model in Acemoglu and Autor (2011) with endogenous skill supply – see Section 4.6 in Acemoglu and Autor (2011). In our framework however, along the transitional equilibrium, the skill distribution differs from the stationary skill distribution.

clearing condition in (14) implies that ω_t is a function of $s_t(i)$ through its effect on the relative high-tech output supply.

Our first result approximates the equilibrium solution of this fixed-point problem by considering a log-linear expansion around the stationary equilibrium. We denote with " $\hat{}$ " variables in log-deviations from the stationary equilibrium. As a first step in the equilibrium characterization, we use the following lemma to reduce the dimansionality of the fixed-point problem determining the economy's equilibrium path.

Lemma 4 (Worker allocations) *Given $\hat{s}_0(i)$ and the corresponding $\hat{y}_t, \hat{q}_t$, the evolution of the skill distribution $\hat{s}_t(i)$, the optimal lotteries $\hat{s}_t(i)$, and the assignment threshold $\hat{l}_t$ are approximated by:*

$$\hat{s}_t(i) = \hat{s}_0(i)e^{-\delta t} + \int_0^t e^{\delta(\tau-t)}\hat{\hat{s}}_\tau(i)d\tau \tag{17}$$

$$\hat{\hat{s}}_t(i) = \left(\mathbb{I}_{i>l} - \int_l^1 s(i)di\right)\psi\hat{q}_t + o_t(i) \tag{18}$$

$$\hat{l}_t = \frac{\eta}{\theta-1}\hat{y}_t \tag{19}$$

where $o_t(i)$ is such that $\int s(i)o_t(i)di = 0$.

Proof. See Appendix A.3 ■

Lemma 4 links the equilibrium path of the skill distribution and the assignment threshold to the joint dynamics of $\{\hat{q}_t, \hat{y}_t\}$, where $log(q_t) \equiv \int_t^\infty e^{-(\rho+\delta)(s-t)}log(\omega_s)ds$ is the present discounted value of the log-relative wage and y_t is the relative high-tech output. Notice that the change in the optimal skill lottery in (17) along the transition depends centrally on the evolution of the relative return of skills employed in the high-tech production. In particular, changes in the present discounted value of relative prices $\hat{q}_t$ affect the relative lifetime earnings of types employed in the H-technology in the new stationary equilibrium ($i > l$). The parameter ψ controls the sensitivity of the optimal skill investment to such changes. The overall skill distribution in (18) is then simply a population-weighted average of the skill distributions of each generation. Since generations are born and die at rate δ, the population share at time t of the initial generation is $e^{-\delta t}$ whereas entering generation τ has a weight $\delta e^{\delta(\tau-t)}$. Finally, the assignment threshold changes are driven by changes in relative prices $\hat{\omega}_t$. The assignment threshold's sensitivity to such wage changes depends on the degree of technology-skill specificity controlled by the short-run skill supply elasticity η. We obtain the expression in (19) because the demand equation in (2) implies that $\hat{\omega}_t = \frac{1}{\theta-1}\hat{y}_t$.

We are now ready to derive our first main result. The following proposition characterizes a system of differential equations that determines the equilibrium path of $\{\hat{q}_t, \hat{y}_t\}$, given an initial $\hat{y}_0$.

Proposition 1 (q-theory of Skill Investment)

1. *Consider the initial condition* $\hat{y}_0$ *and the terminal condition* $\lim_{t\to\infty} \hat{y}_t = 0$. *The dynamic system describing the joint evolution of* $\{\hat{q}_t, \hat{y}_t\}$ *is*

$$\frac{\partial \hat{y}_t}{\partial t} = -\delta \hat{y}_t + \frac{\theta - 1}{\theta + \kappa\eta}\delta\psi\hat{q}_t \tag{20}$$

$$\frac{\partial \hat{q}_t}{\partial t} = (\rho + \delta)\hat{q}_t + \frac{1}{\theta - 1}\hat{y}_t, \tag{21}$$

where κ *is a positive constant.*

2. *The equilibrium for* $\{\hat{q}_t, \hat{y}_t\}$ *is saddle-path stable and given by:*

$$\hat{y}_t = \hat{y}_0 e^{-\lambda t} \quad \textit{and} \quad \hat{q}_t = \zeta \hat{y}_0 e^{-\lambda t} \tag{22}$$

where

$$\lambda = -\frac{\rho}{2} + \sqrt{\left(\frac{\rho}{2}\right)^2 + \delta\left((\rho + \delta) + \frac{\psi}{\theta + \kappa\eta}\right)} \quad \textit{and} \quad \zeta = -\frac{1}{\theta - 1}\frac{1}{\rho + \delta + \lambda}$$

Proof. See Appendix A.4. ■

The system in Proposition 1 is a rather standard one in macroeconomics. Relative output, $\hat{y}_t$, acts as a state variable whose law of motion needs to be solved backward. The optimal lottery is the control variable that needs to be solved forward. The present value of future relative technology-specific wage, q_t, is the only endogenous variable affecting the optimal lottery choice, so it is the only variable that needs to be solved forward.[11] Importantly, this result indicates that the equilibrium dynamics of $\{\hat{q}_t, \hat{y}_t\}$ can be solved without keeping track of the evolution of the skill distribution $s_t(i)$. This is a direct consequence of the fact that the equilibrium condition in (14) determining the evolution of $\hat{l}_t$ and, therefore, of $\hat{y}_t$ only contains integrals of $s_t(i)$ whose equilibrium dynamics only depend on $\hat{q}_t$ as shown in Lemma 4.

The second part of Proposition 1 shows that the equilibrium always exists and is unique, which is a consequence of saddle-path stability. Conditional on the overall adjustment in relative output $\hat{y}_0$, relative output y_t and the present discounted value of the relative wage q_t have a constant convergence rate of λ to the stationary equilibrium. The system immediately yields the path of the relative wage along the transition since $\hat{\omega}_t = \frac{1}{1-\theta}\hat{y}_t$. Whenever $\theta > 1$, the changes in relative output and relative wage in the H-technology are negatively correlated along the transition. Intuitively, the endogenous evolution of the skill distribution shifts the relative labor supply in high-tech production, which generates an increase in the relative high-tech supply and, consequently, a decrease in the relative wage $\hat{\omega}_t$ and its presented discounted value $\hat{q}_t$.

[11] Such log-linear expansions are common when characterizing transitional dynamics in macroeconomic models. Specifically, the system's structure is similar to that of the neoclassical growth model in which the equilibrium is characterized by a dynamic system with one state and one control – capital and consumption – and two equations – a forward looking Euler equation and backward looking capital accumulation equation.

The system describing the evolution of $\{\hat{q}_t, \hat{y}_t\}$ is isomorphic to the q-theory of capital investment (Tobin, 1969, Hayashi, 1982) where, in our model, $\hat{q}_t$ is the present discounted value of relative price of high-tech output. In other words, it is the shadow price of the human capital "asset" associated with having one additional unit of high-tech output. Whenever this marginal value is higher, the incentives to invest in high-i skills are stronger, which has a positive impact on the relative value of high-tech output $\hat{y}_t$. As in the q-theory, the elasticity of relative output value $\hat{y}_t$ with respect to $\hat{q}_t$ depends on parameters governing the costs of adjustment (i.e, the term $\delta\psi$). However, as opposed to the traditional q-theory where all units of the capital stock are homogenous and perfectly substitutable, our model features imperfect substitution of human capital across technologies. Thus, the elasticity of y_t to q_t also depends on the degree of technology-skill specificity measured by η.

3 The Economy's Adjustment to a Skill-biased Technological Innovation

We now analyze the dynamic adjustment of our economy to one-time permanent increase in the relative productivity A. Because this innovation increases the relative productivity of workers with higher skill-types i that are sorted into the H sector, we refer to it as a skill-biased technological innovation. We first use the results above to characterize in closed-form the joint dynamics of q_t and y_t, as well as the evolution of the skill distribution and the technology-skill assignment. We then analyze how the parameters governing the short-run and long-run skill supply affect different features of the economy's dynamic adjustment to the skill-biased technological innovation.

3.1 Dynamic Responses of Equilibrium Outcomes

We analyze the economy's adjustment following an unanticipated increase in the productivity of high-tech production at time $t = 0$, $\Delta \log(A)$. We assume that immediately prior to the shock at time $t = 0^-$, the economy of Section 2 is in a stationary equilibrium with skill distribution given by $s_0(i)$. The shock only augments the productivity of skill types employed in high-tech production. Given the positive sorting of types into the H technology, the innovation is biased towards high-i types.

We start by characterizing the dynamic responses of q_t and y_t. Let $\Delta \log(q_t) \equiv \log(q_t/q_{0^-})$ and $\Delta \log(y_t) \equiv \log(y_t/y_{0^-})$ denote change in outcomes from their initial levels.

Proposition 2 (Dynamic responses) *Given a skill-biased technological innovation* $\Delta \log(A)$*, the*

dynamic responses $\Delta \log(y_t)$ *and* $\Delta \log(q_t)$ *are approximated by:*

$$\Delta \log(y_t) = \left(\frac{1+\kappa\eta}{\theta+\kappa\eta} + \frac{\psi}{\chi} \frac{\theta-1}{\theta+\kappa\eta} (1 - e^{-\lambda t}) \right) (\theta - 1) \Delta \log(A)$$
$$\Delta \log(q_t) = \frac{1}{\chi} \left(1 + \frac{\lambda - \delta}{\delta} e^{-\lambda t} \right) (\theta - 1) \Delta \log(A)$$

where $\chi \equiv \left(\theta + \kappa\eta + \frac{\psi}{\rho+\delta} \right) (\rho + \delta)$.

Proof. See Appendix A.5. ■

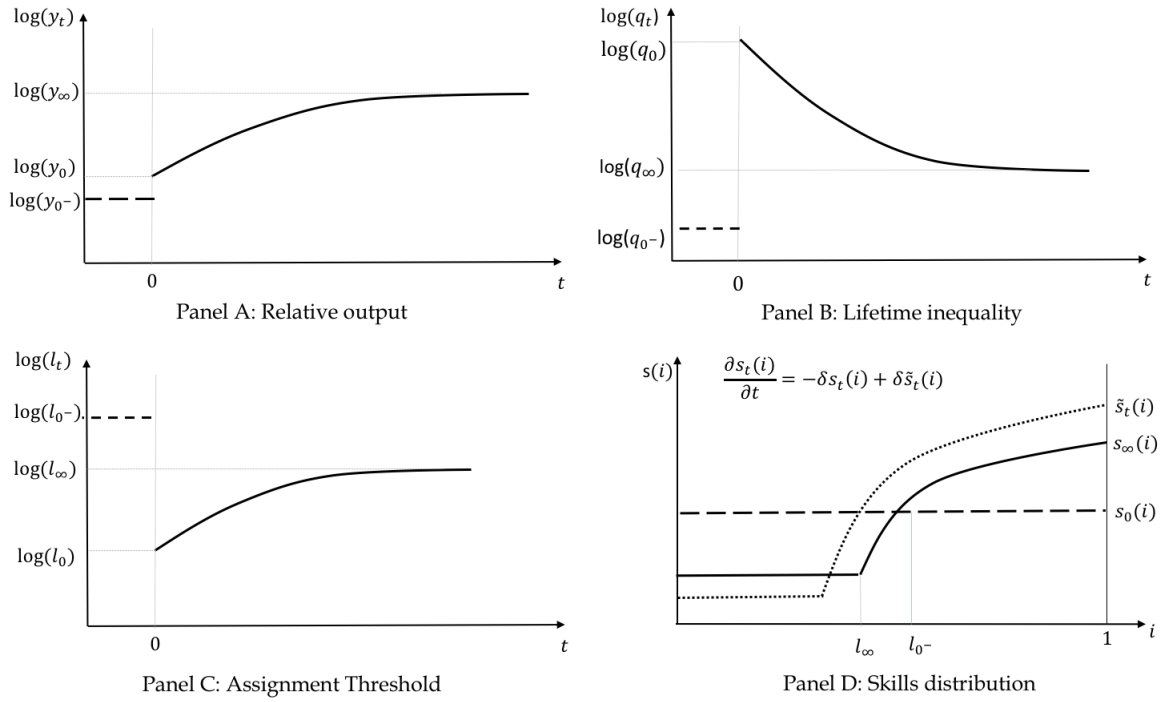
The proposition shows that, when $\theta > 1$, both y_t and q_t increase on impact and in the long-run. However, along the transition, q_t falls while y_t increases at rate λ. To derive these expressions, first note that the results in Proposition 1 immediately yield the transitional dynamics of y_t and q_t given an initial condition for $\hat{y}_0 = \log(y_0/y_\infty)$. Then, to characterize the full dynamic responses, we derive expressions for the short-run and long-run changes in y_t implied by the technological shock, $\Delta \log(A)$, given the economy's initial skill distribution $s_0(i)$. The expressions highlight how both the short-run and long-run skill supply parameters, η and ψ, crucially shape the dynamic responses. We return to this point in Section 3.2 when we analyze how these parameters affect the economy's impulse response functions to the technological innovation.

We are now ready to provide a unified account of the economy's adjustment to a skill-biased technological shock by combining this proposition with Lemma 4's characterization of the evolution of the skill distribution and the assignment threshold. Figure 2 provides a graphical illustration.

The productivity increase causes an increase in the demand for workers in the high-tech production. In equilibrium, the relative price increases on impact ($\Delta \log(\omega_0) > 0$). This causes workers of the old generation that had relatively low skill-types to enter the high-tech sector ($\Delta \log(l_0) < 0$). Along the transition, younger entering workers decide to invest in high-i skills that are complementary to high-tech production in anticipation of higher relative wages (as captured by $\Delta \log(q_t) > 0$). Then, the overall skill distribution changes as older generations are replaced with younger generations at rate δ. This increase in the mass of high-i workers expands the supply of high-tech output over time ($\Delta \log(y_t) > 0$), triggering a decline in the current and present discounted value of relative wages, ω_t and q_t, and an increase in the assignment threshold, l_t. This implies that intermediate-i types that initially entered the high-tech sector are displaced over time. Finally, compared to the initial equilibrium of the economy, the new long-run equilibrium entails a higher relative wage and output, and a larger mass of workers in the high-tech sector (driven both by a stationary skill distribution with higher mass in high-i types and the lower assignment threshold).

These results indicate the importance of considering the dynamics generated by changes in the skill distribution across generations. If we focus only on responses observed at any fixed horizon, then we miss part of the economy's adjustment. For instance, an extrapola-

Figure 2: The economy's adjustment to a skill-biased technological shock



tion from observed responses at short horizons misses the relative output increase and the relative wage decline that happen for future generations. In contrast, extrapolating from observed responses at longer horizons misses the fact that older generations experienced higher relative wages and lower relative output along the transition.

3.2 Comparative Statics: Short-Run and Long-Run Skill Supply Elasticities

We now turn to a detailed investigation of how the parameters that govern the short-run and long-run skill supply elasticities affect the economy's adjustment to the skill-biased technological innovation at different horizons.

Short-run skill supply elasticity. Consider first how the short-run skill supply elasticity affects the impulse response function of labor market outcomes. In our theory, this elasticity is inversely related to the extent of technology-skill specificity – i.e., how different skill types are in terms of relativity technology-specific productivity. Thus, this comparative static exercise can be thought as a comparison of the transitional dynamics following shocks to technologies with different degrees of technology-skill specificity. In other words, this exercise speaks to events in which skills used to operate one technology are more easily transferable to be used in the new improved technology.

The following proposition formally establishes how η affects the impulse response functions to skill-biased technological innovations.

Proposition 3 (Comparative statics with respect to η)

1. *Short-run adjustment*

$$\frac{\partial \Delta log(y_0)}{\partial \eta} > 0 > \frac{\partial \Delta log(q_0)}{\partial \eta}$$

2. *Long-run adjustment*

$$\frac{\partial \Delta log(y_\infty)}{\partial \eta} > 0 > \frac{\partial \Delta log(q_\infty)}{\partial \eta}$$

3. *Persistence*

$$\frac{\partial \left(\int_0^\infty |\hat{y}_t| \, dt \right)}{\partial \eta} < 0, \quad \frac{\partial \left(\int_0^\infty \hat{q}_t dt \right)}{\partial \eta} < 0$$

Proof. See Appendix A.6. ■

To fix ideas, Figure 3 illustrates the results in Proposition 3 with the impulse response functions of two economies. The black lines show the responses of an economy with a high value of η (i.e., high short-run skill supply elasticity or low technology-skill specificity). The blue lines show the responses of an economy with a low value of η (i.e., low short-run skill supply elasticity or high technology-skill specificity).

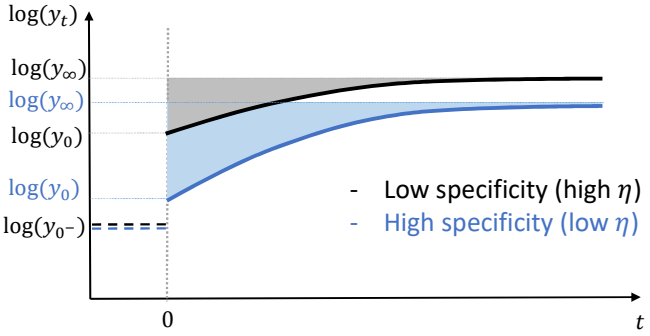
In the short run, the economy with a higher η exhibits a larger change in the relative output of high-tech production and a lower change in the relative lifetime wage in high-tech production. This follows from the fact that, in the economy with a higher η, it is easier to reallocate workers across technologies in response to the shock, implying that

$$\frac{\partial |\Delta log(l_0)|}{\partial \eta} > 0 \quad \text{and} \quad \frac{\partial \Delta \log(\omega_0)}{\partial \eta} < 0.$$

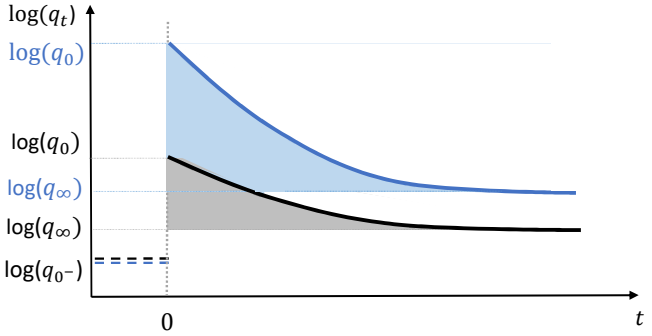
As a consequence of the smaller change in the relative lifetime wage in high-tech production when η is higher, younger workers have weaker incentives to adjust their skill investment compared to older workers. Thus, as we can see from Lemma 4, the overall change in the skill distribution implied by the shock is smaller as well. This then implies that the transitional dynamics from the short- to the long-run are less persistent, as measured by the cumulative impulse response (e.g., $\int_0^\infty \hat{q}_t dt$). Finally, while the smaller change in the skill distribution could have implied a smaller (larger) overall long-run adjustment in relative output (lifetime inequality), it turns that that larger (smaller) short-run response dominates. Thus the long-run adjustment in relative output (lifetime inequality) is larger (smaller).

To summarize, these results show that, compared to the low-η economy, adjustments in the high-η economy are stronger for relative output but weaker for relative lifetime earnings and relative skills. Importantly, these adjustments are front-loaded in the high-η economy.

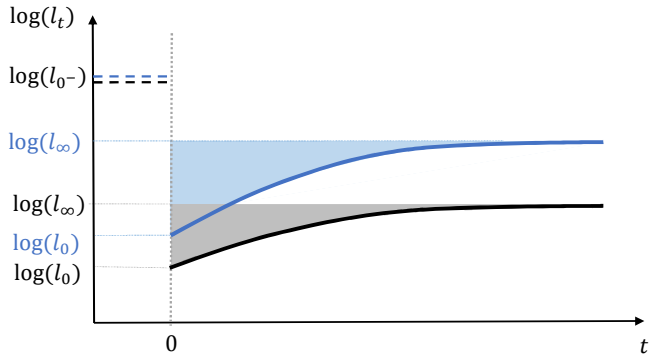
Figure 3: Comparative statics with respect to η



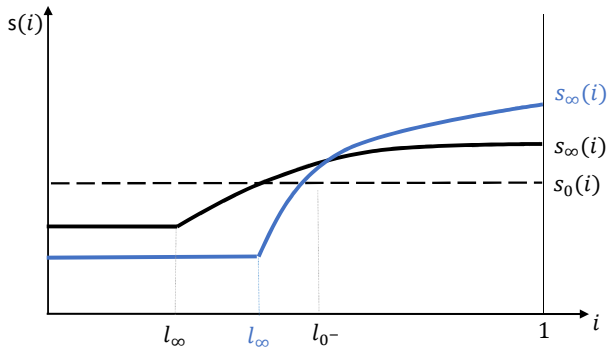
Panel A: Relative output



Panel B: Lifetime inequality



Panel C: Assignment Threshold



Panel D: Skills distribution

Long-run skill supply elasticity. We now consider how the parameter ψ affects the economy's adjustment to the technological innovation in high-tech production. This comparative statics exercise illustrates how economies with different costs of skill investment—and thus different degrees of long-run skill supply elasticity—respond to skill-biased technological shocks over time. This comparative static exercise can be thought as a comparison of the transition to technological shocks in two economies that differ with respect to the ability of young workers to adjust skill investment. It speaks to situations in which it is easier to tilt investment towards skills in high demand in the labor market due to, for example, the availability of vocational training or a more flexible educational system.

Proposition 4 (Comparative statics with respect to ψ)

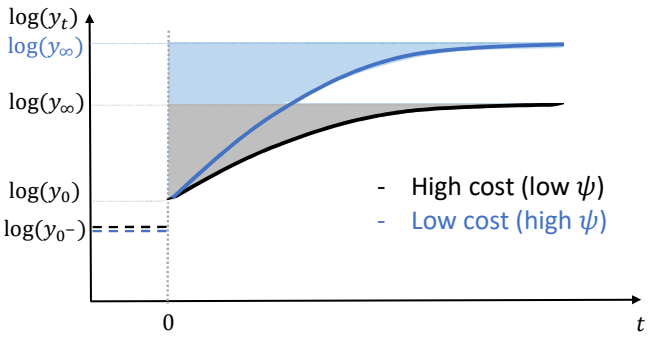
1. *Short-run adjustment*

$$\frac{\partial \Delta log(y_0)}{\partial \psi} = 0 > \frac{\partial \Delta log(q_0)}{\partial \psi}$$

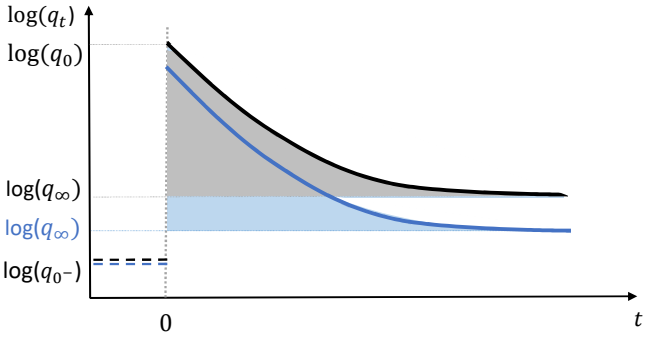
2. *Long-run adjustment*

$$\frac{\partial \Delta log(y_\infty)}{\partial \psi} > 0 > \frac{\partial \Delta log(q_\infty)}{\partial \psi}$$

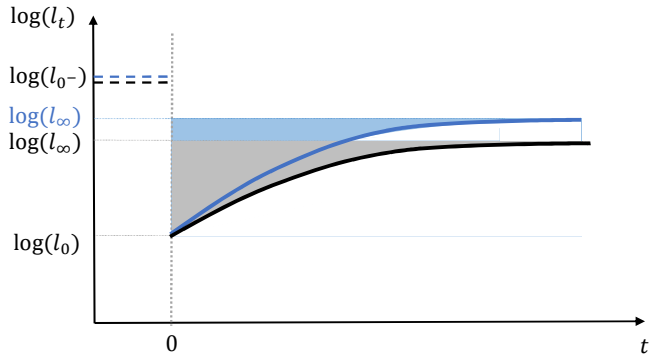
Figure 4: Comparative statics with respect to ψ



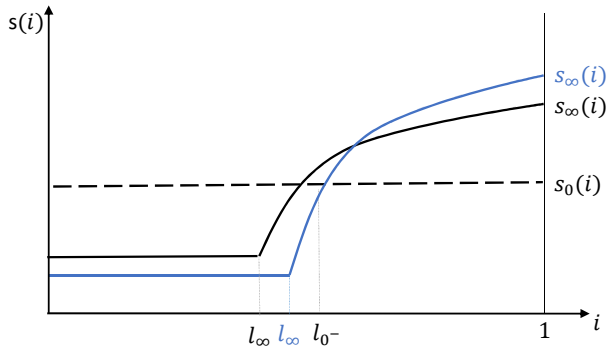
Panel A: Relative output



Panel B: Lifetime inequality



Panel C: Assignment Threshold



Panel D: Skills distribution

3. Persistence

$$\left.\frac{\partial\left(\int_0^\infty |\hat{y}_t|\,dt\right)}{\partial\psi}\right|_{\psi=0} > 0, \quad \left.\frac{\partial\left(\int_0^\infty \hat{q}_t dt\right)}{\partial\psi}\right|_{\psi=0} > 0$$

Proof. See Appendix A.6. ■

Again, we illustrate the results in Proposition 4 with a graphical representation of the impulse response functions of two economies in Figure 4. The blue lines depict the adjustment of an economy with a high value of ψ, and the black lines represent the responses of an economy with a low value of ψ. Accordingly, the "black" economy is closer to a static model with an exogenous skill distribution. In contrast, the "blue" economy entails stronger dynamic forces arising from the higher ability of younger generations to adjust their skill invest decisions.

The first part of Proposition 4 indicates that the short-run response of relative output is identical in both economies. This is because the parameter ψ does not affect the relative employment of old generation born before the shock. However, a higher ψ attenuates the short-run increase of lifetime inequality because future relative wages fall by more as consequence of the larger increase in the future supply of high-i skills (as can be seen from

Lemma 4). The later also implies that relative output (lifetime inequality) increases more (less) in the long-run. Finally, due to the larger change in the skill distribution along the transition, the persistence of both lifetime inequality and relative output increases when ψ is higher. Overall, in comparison to a low-ψ economy, a high-ψ economy back-loads the adjustment in relative output and wages.

At this point, it is important to highlight the differences that short- and long-run skill supply elasticity create in how new technologies affect the economy. In the long-run, a higher η or a higher ψ have similar qualitative implications for the adjustment of relative output and lifetime inequality. However, the two parameters have different implications for the transitional dynamics of the economy due to their different implications for the equilibrium path of the skill distribution. On the one hand, a higher ψ reduces the cost of adjusting skills, generating larger changes in the skill distribution along the transition. This implies that adjustments are back-loaded as ψ increases. On the other hand, a higher η reduces the change in relative wages at impact, which attenuates changes in the skill distribution. This implies that adjustments are front-loaded as η increases. The next section shows that, because of their different dynamic implications, η and ψ have different quantitative implications for the welfare consequences of skill-biased technological innovations.

4 Welfare Consequences of a Skill-biased Technological Innovation

We now evaluate the welfare consequences of skill-biased technological innovations. We analyze the same one-time permanent increase in the relative productivity of the H-technology. Our analysis has two parts. First, we characterize the impact of the shock on average welfare and lifetime inequality for different worker generations. Second, we show that ignoring the dynamics of the skill distribution across generations biases calculations of the welfare consequences of new technologies.

4.1 Welfare Analysis

Our welfare measure is the ex-ante expected utility of individuals born at each point in time. This is equivalent to the average across all individuals in a cohort of their discounted lifetime utility. In our setting, the ex-ante expected utility of the worker cohort born at time τ is given by the solution of the utility maximization problem in (9). Combining this solution with the expression for wages in (7), we can write the ex-ante expected utility of cohort τ as

$$U_\tau = \int_0^1 \tilde{s}_\tau(i) \left[\log\left(\alpha(i)^{\rho+\delta} Q_\tau(i)\right) - \int_\tau^\infty e^{-(\rho+\delta)t} log\left(P_t\right) dt - \frac{1}{\psi} \log\left(\frac{\tilde{s}_\tau(i)}{\bar{s}(i)}\right) \right],$$

where $\tilde{s}_\tau(i)$ is the skill distribution of cohort τ, $Q_\tau(i)$ is the present-discounted value of $\max\{\log(\omega_t \sigma(i)), 0\}$ defined in Lemma 3, and P_t is the ideal price index defined in (3).

In order to obtain a welfare measure for the economy, it is necessary to aggregate the welfare of the different worker cohorts. We take an utilitarian approach by considering a weighted sum of the ex-ante utility of different cohorts. Specifically, we define the economy's discounted average welfare at t as

$$\bar{U}_t = r \int_t^\infty e^{-r\tau} U_\tau d\tau.$$

Our social welfare function multiplies cohort τ's ex-ante utility by a cohort-specific weight of $re^{-r\tau}$. The parameter r captures the idea that the welfare of future generations may be discounted at each point in time. To see this, it is useful to consider two extreme cases. When $r \to \infty$, the social welfare function completely ignores the welfare of all future cohorts. In the other extreme, when $r \to 0$, the social welfare function only gives positive weight to generations born in the new stationary equilibrium.

We also consider the impact of the new technology on welfare inequality. Notice that the relative technology-specific wage is the only component of relative earnings across skill types that responds to the technological shock. This implies that it will be useful to define the lifetime welfare inequality for cohort τ as the present discounted value of the relative technology-specific wage, $\log q_\tau$. We again aggregate different worker cohorts by defining the economy's discounted lifetime welfare inequality at t as

$$\bar{\Omega}_t = r \int_t^\infty e^{-r\tau} \log(q_\tau) d\tau.$$

We now use these measures to characterize the welfare consequences of the one-time permanent change in A. As in Section 3, we assume that the economy is in a stationary equilibrium right before the shock at $t = 0^-$, so that $\bar{U}_{0^-} = U_{0^-}$ and $\bar{\Omega}_{0^-} = \log(q_{0^-})$. The following proposition characterizes the shock-induced changes in average welfare $\Delta\bar{U} \equiv \bar{U}_0 - U_{0^-}$ and lifetime inequality $\Delta\bar{\Omega} \equiv \bar{\Omega}_0 - log(q_{0^-})$ as a function of the exogenous technological innovation, $\Delta \log(A)$, and the endogenous change in the discounted value of relative wages, $\log(q_t)$.[12]

Proposition 5 (Average welfare and lifetime welfare inequality) *The changes in average welfare $\Delta\bar{U}$ and lifetime inequality $\Delta\bar{\Omega}$ are approximately:*

$$\Delta\bar{U} = \frac{1}{\rho+\delta}\frac{y_\infty}{1+y_\infty}\Delta\log(A) - \left(\frac{y_\infty}{1+y_\infty} - \frac{e_\infty}{1+e_\infty}\right)\Delta\bar{\Omega}$$
$$\Delta\bar{\Omega} = \Delta\log(q_\infty) + \frac{\lambda r}{r+\lambda}\int_0^\infty \hat{q}_\tau d\tau$$

where $e_\infty \equiv \left(\int_{l_\infty}^1 s(i)di\right) / \left(\int_0^{l_\infty} s(i)di\right)$ is the relative H-specific employment in the new stationary equilibrium.

[12]Notice that, because of the assumption of log-utility, $(\rho+\delta)\Delta\bar{U}$ has the interpretation of a consumption-equivalent variation for a worker cohort.

Proof. See Appendix A.7. ■

The change in average welfare, $\Delta\bar{U}$, combines the impact of the shock on both the cost of the final consumption good and the average lifetime earnings of workers. Consider first the change in the price of the final good. Since the price of the low-tech good is the numeraire, only changes in the cost of the high-tech good affect the price index. Such an effect is proportional to the share of the H-good in the consumption bundle, $\frac{y_\infty}{1+y_\infty}$. The cost of the high-tech good itself changes for two reasons: (i) the exogenous technology shock captured by the term $\frac{1}{\rho+\delta}\frac{y_\infty}{1+y_\infty}\Delta\log(A)$, and (ii) the endogenous relative wage change captured by the term $\frac{y_\infty}{1+y_\infty}\Delta\bar{\Omega}$. Finally, the change in average welfare also responds to the change in the average lifetime earnings of workers. Again, since only the H-specific wage changes, average lifetime earnings only responds to the change in lifetime inequality $\Delta\bar{\Omega}$. This average wage response is proportional to the share of workers employed with the H-technology, $\frac{e_\infty}{1+e_\infty}$.

Proposition 5 indicates that inequality has a negative impact on welfare whenever the H-technology accounts for a higher share of output than employment. Intuitively, in this case, changes in the relative wages induce changes in the price index for everyone that are larger than the increase in the average wage of those employed in the H-technology. In our theory, this happens whenever the labor earnings of workers in the H-technology are above the economy's average labor earnings.[13]

Turning to the change in welfare inequality in the second part of the proposition, we can see that $\Delta\bar{\Omega}$ combines the long-run change and the persistence of changes in $\log(q_t)$. The relative importance of persistence is increasing in λ since it governs how fast lifetime inequality decays along the transition, and as such, the differences *between* generations in lifetime inequality. When λ is higher, generations far in the future have similar lifetime inequality to the initial generation. Then, because such future generations are discounted at rate r, the average lifetime inequality across generations increases.

Taken together, the results from Propositions 5 highlight the importance of distinguishing between economies that differ in terms of adjustment persistence. In our theory, such a persistence arises from the endogenous change in the skill distribution across generations whose magnitude depends on the short- and long-run skill supply elasticities (as shown in Propositions 3 and 4). Thus, ignoring changes in skills across generations introduces biases when evaluating the welfare consequences of new technologies.

4.2 Welfare Consequences of Adjustment Across Generations

We now investigate how calculations of the welfare consequences of technological shocks are affected by the persistence of changes in labor market outcomes along the transition to the new equilibrium. In our theory, persistence arises from the dynamics of the skill distribution. So, in order to evaluate its consequences, we consider a static version of our model in which

[13] Such a case arises if absolute advantage is positively correlated with comparative advantage to operate the H-technology – i.e., $\alpha(i)$ is increasing in i.

we shut down any skill investment of young workers. However, we allow this static model to match labor market responses over one particular horizon. This exercise thus speaks directly to the risks of ignoring the adjustment across generations by focusing on estimates of the impact of new technologies on labor market outcomes over fixed time horizons.

To be more precise, we engage in the following thought experiment. Consider an economy subject to a one-time permanent shock $\Delta \log A$. Suppose that this economy behaves according to the theoretical predictions described in Section 3 with short- and long-run skill supply elasticity given by η and ψ, respectively. We consider a researcher that relies on a static assignment model to analyze how this economy responds to the technological shock. Through the lens of our theory, this researcher considers a misspecified parametrization of the economy in which the long-run elasticity equals zero. This parametrization shuts down any dynamics in the economy because the skill distribution is the same for all generations.

We assume that this researcher observes responses in labor market outcomes over a fixed horizon $t = T$. We focus on changes in lifetime inequality since this is the main endogenous outcome entering the welfare computations in Proposition 5. We consider two ways in which the researcher may decide to use the static model to match the observed inequality response, $\Delta \log q_T$. In the first approach, the researcher observes the true shock ($\Delta \log A^1 = \Delta \log A$), and selects η^1 to match $\Delta \log q_T$ with $\psi^1 = 0$. In the second, the researcher observes the true parameter ($\eta^2 = \eta$), and selects the size of the shock $\Delta \log A^2$ to match $\Delta \log q_T$ with $\psi^2 = 0$.[14]

The following proposition shows that, despite matching inequality responses at time T, this researcher misses the economy's transitional dynamics triggered by the evolution of the skill distribution across generations. This introduces biases in the evaluation of the welfare consequences of the technological innovation.

Proposition 6 *Consider an economy in which η and ψ are positive. Assume that $\Delta \log A$ generates a change in lifetime inequality between $t = 0$ and $t = T$ of $\Delta \log q_T$. Consider predictions under two alternative static parametrizations of the model ($\psi^1 = \psi^2 = 0$).*

1. *Suppose $\Delta \log A^1 = \log A$ is known such that $\frac{\Delta \log(A)}{\Delta \log(q_T)} > \frac{\theta(\rho+\delta)}{\theta-1}$. There exists η^1 that matches $\Delta \log q_T$ with an associated T^1 such that $\Delta \bar{\Omega}^1 > \Delta \bar{\Omega}$ and $\Delta \bar{U}^1 < \Delta \bar{U}$ if, and only if, $T < T^1$.*
2. *Suppose $\eta^2 = \eta$ is known. There exists $\Delta \log A^2$ that matches $\Delta \log q_T$ with an associated T^2 such that $\Delta \bar{\Omega}^2 > \Delta \bar{\Omega}$ and $\Delta \bar{U}^2 < \Delta \bar{U}$ if $T < T^2$.*

Proof. See Appendix A.8. ■

This proposition shows that there are multiple ways in which researchers can use a static version of our model to match observed inequality responses over a fixed horizon. All versions ignore the transitional dynamics of labor market outcomes generated by changes in the

[14] There are many variants of this exercise. We could consider for example the two approaches above with the researcher targeting changes in relative output instead of changes in lifetime inequality. This alternative exercise exploits the relationship between changes in q_t and y_t in Proposition 2. Another possibility is to consider an exercise in which the researcher selects $\Delta \log A$ and η to jointly match changes in both relative output and relative wages.

skill distribution across generations. This introduces biases in the evaluation of the welfare consequences of new technologies. If the researcher only matches inequality responses in short horizons (i.e. T is low), then she will think that inequality will remain high in the future. This makes her overpredict the present value of lifetime inequality, and underpredict the average welfare gain. Alternatively, a researcher using the first approach would reach the opposite conclusions if she matches inequality responses in long horizons (i.e. T is high).

Such biases will be larger when there is higher persistence in labor market outcomes due to larger changes in the skill distribution along the transition. As shown in Section 3.2, this is the case whenever the skill supply elasticity is low in the short-run (i.e, η is low) but large in the long-run (i.e., ψ is large).

This discussion highlights the potential for biases of ignoring the skill adjustment across generations. We now turn to an empirical investigation of different observable implications of our theory that speak to the magnitude of the shock-induced changes in the skill distribution. In other words, we analyze responses in labor market outcomes that are informative about the size of the skill supply elasticity in the short- and long-run. In Section 6, we use this empirical analysis to quantify the biases of ignoring the skill adjustment across generations.

5 Within- and Between-Generation Consequences of New Technologies

Our theoretical results establish the distinct roles of the short- and long-run skill supply elasticity in shaping how skill-biased technological shocks affect labor market outcomes and welfare. In this section, we empirically assess the importance of these mechanisms in transitions triggered by recent skill-biased technological innovations. We start by connecting the parameters of short- and long-run skill supply elasticity respectively to the employment adjustment of old and young generations. We then explore this insight to provide three pieces of evidence indicating that these two separate mechanisms are relevant to understand the economy's adjustment to the recent arrival of new cognitive-biased technologies.

First, in 18 developed countries, employment growth in the most cognitive-intensive occupations was stronger for young workers than for old workers. Second, turning to a detailed investigation of these responses in Germany, we show that, in the cross-section of occupations, growth of employment and payroll was increasing in the time spent performing cognitive-intensive tasks. We find that these responses are stronger for younger than for older generations. Finally, we explore cross-regional variation in adoption timing to obtain empirical impulse response functions to one cognitive-biased technological innovation: the arrival of broadband internet in the early 2000s. We find that the impact on relative employment is small for older generations at all horizons, but increasing over time for younger generations. These estimates suggest that the long-run skill elasticity is positive, but the short-run skill elasticity is low.

5.1 Observable Responses in Labor Market Outcomes Between and Within Generations

This section shows how technological shocks affect relative employment for different generations of workers. Although Section 3 links the parameters governing short- and long-run skill supply to the response of several outcomes, many of them are not readily observable to researchers. For instance, as in the q-theory of capital investment, q_t is a forward-looking variable whose measurement requires knowledge of the entire path of relative wages. In addition, without taking an explicit stance on what observable attributes determine worker skills in different activities, it is not possible to directly measure the skill distribution and the technology-skill assignment. However, we now show that the observable generation-specific employment responses combine the dynamic adjustment in these variables and, therefore, depend on our two main theoretical mechanisms.

As before, we consider a one-time permanent change in A at $t = 0$, and derive the dynamic adjustment of relative high-tech employment for old and young generations. We consider older generations as those born before period $t = -x$ and younger generations as those born at period $t = -x$. In period $t \geq 0$, the relative high-tech employment of these worker generations are given by

$$e_t^{old} = \frac{\int_{l_t}^{1} s_0(i)di}{\int_0^{l_t} s_0(i)di} \quad \text{and} \quad e_t^{young} = \frac{\tilde{x}_0 e^{-\delta t}\int_{l_t}^{1} s_0(i)di + \delta\int_0^t e^{\delta(\tau-t)}\int_{l_t}^{1}\tilde{s}_\tau(i)did\tau}{\tilde{x}_0 e^{-\delta t}\int_0^{l_t} s_0(i)di + \delta\int_0^t e^{\delta(\tau-t)}\int_0^{l_t}\tilde{s}_\tau(i)did\tau},$$

where $\tilde{x}_0 \equiv 1 - e^{-\delta x}$ is the population share of the young generation at $t = 0$.

For both worker groups, the technology-skill assignment is identical and determined by the threshold l_t. Notice that all workers of the old generations have the pre-shock skill distribution, $s_0(i)$. However, the skill distribution of young generations combines the pre-shock distribution, $s_0(i)$, and the post-shock lotteries, $\tilde{s}_\tau(i)$. The overlapping generation structure of the model implies that the relative share of workers in the young generation with the pre-shock skill distribution decays at the constant rate δ.[15]

Relative employment of old generation: Short-run skill supply elasticity. In Appendix A.9, we show that the change in the relative employment of old generations is

$$\Delta \log e_t^{old} \approx \frac{\eta}{\theta + \kappa\eta}\frac{1}{e_H}\left(1 - \frac{\psi}{\chi}(1 - e^{-\lambda t})\right)(\theta - 1)\Delta \log A, \tag{23}$$

where e_H is the high-tech employment share at $t = 0^-$.

[15] We allow the young group to include workers born before the shock (since $x \geq 0$). This circumvents the challenge of identifying the cohorts that start adjusting their skills, which arises because, in practice, technologies may not be adopted instantaneously and young workers may still invest on skills after entering the labor force (in the form of vocational training or on-the-job learning). It is also possible to allow part of the workers born before the shock to adjust their skills at $t = 0$. In this case, rather than $s_0(i)$, the initial skill distribution would be a mix of $s_0(i)$ and $\tilde{s}_0(i)$. This extension does not alter our main qualitative insights, but reduces the magnitude of the short-to-long adjustment in the economy.

Among old generations, the increase in the relative productivity of high-tech production induces the reallocation of older workers towards high-tech production whenever $\theta > 1$. The expression indicates that this positive effect on the relative high-tech employment becomes weaker over time. As discussed in the previous section, this follows from the expansion of high-i skills among younger generations, which displaces old workers with marginal skills from high-tech production – i.e., those with skills $i \in (l_0, l_\infty)$.

Importantly, expression (23) shows that the magnitude of the increase in high-tech relative employment is increasing in the short-run skill supply elasticity, η. To see this more clearly, consider the relative employment response of old workers at impact:

$$\frac{\Delta \log e_0^{old}}{\Delta \log A} \approx \frac{\eta}{\theta + \kappa\eta} \frac{\theta - 1}{e_H}.$$

This expression indicates that, conditional on the shock size, a higher η induces larger changes in relative employment in the short-run. Whenever η is small, old workers do not change their relative employment following the shock. Intuitively, this parameter controls the technology-skill specificity and, therefore, how much changes in the relative high-tech wage affect the assignment threshold and relative employment.

Relative employment of young generation: Long-run skill supply elasticity. Turning to the employment response among young generations, Appendix A.9 also establishes that

$$\Delta \log e_t^{young} \approx \Delta \log e_t^{old} + \frac{\psi}{\chi} \frac{1 - e^{-\lambda t}}{1 - (1 - \tilde{x}_0) e^{-\delta t}} (\theta - 1) \Delta \log A. \tag{24}$$

This expression indicates that the evolution of the allocation of young workers has two components. The first term captures the change in technology-skill assignment and, since it is the only determinant of the relative employment of old generations, it can be approximated by $\Delta \log e_t^{old}$. The second term captures the change in the skill investment decision of incoming cohorts. At each point in time, this term is positive as young workers distort skill investment towards high-i skills that became more valuable in high-tech production. We can also show that the between-generation difference grows shortly after the shock.

Expression (23) crucially shows that the between-generation difference in relative employment response is increasing in the parameter of long-run skill supply elasticity, ψ. To see this more clearly, consider the between-generation relative employment response in the long-run:

$$\frac{\Delta \log e_\infty^{young} - \Delta \log e_\infty^{old}}{\Delta \log A} \approx \frac{\psi}{\psi + (\theta + \kappa\eta)(\rho + \delta)} (\theta - 1).$$

Conditional on the shock size, a higher ψ yields stronger employment differences across generations in the long-run. Intuitively, this parameter controls the sensitivity of the skill supply of incoming cohorts to changes in relative lifetime earnings. A higher elasticity implies that young workers adjust more intensively their skills, amplifying differences between

their relative employment and that of older generations.

The rest of this section relies on the insights obtained from (23)–(24) to empirically document the importance of short- and long-run skill supply elasticities. Specifically, we evaluate the impact of new technologies on the labor market outcomes of older and younger generations of workers.[16]

5.2 Cognitive-Intensive Employment Growth in Developed Economies

We define cognitive-intensive occupations as being the set of production activities that were disproportionately augmented by recent technological innovations – that is, they correspond to the H technology in the theory. Our approach follows an extensive literature documenting that the recent arrival of new technologies in the workplace, like the computer and the internet, had different effects on jobs with different task content —e.g. Autor, Levy, and Murnane (2003), Spitz-Oener (2006), Autor and Dorn (2013), Akerman, Gaarder, and Mogstad (2015), and, for a review, Acemoglu and Autor (2011). Specifically, this literature has documented that these new technologies augmented the productivity of cognitive-intensive jobs whose daily activities require problem-solving, creativity, or complex interpersonal interactions. On the other hand, these recent technological innovations substituted for routine-intensive jobs whose tasks follow well-understood procedures that can be codified in computer software, performed by machines or, alternatively, offshored over computer networks to foreign work sites.[17]

We analyze the evolution of the occupation employment composition of 18 developed countries. We use data on the number of males employed by occupation for two age groups: "Young" workers aged 15-39 yrs and "Old" workers aged 40-64 yrs. We consider employment in 9 aggregate occupation groups.[18] Using the German BERUFNET dataset, we rank occupations according to their share of time spent on tasks that intensively require analytical non-routine and interactive skills. We classify as cognitive-intensive the top 3 occupations in this ranking: Managers, Professionals, Technicians and Associate Professionals.[19]

[16] Notice the dynamic responses of relative employment also depend on the economy's speed of adjustment, λ. As shown in Proposition 2, this parameter controls the dynamic adjustment of relative output, y_t, which, in our theory, is equivalent to the dynamic adjustment of relative payroll across technologies. Thus, we investigate how technological innovations affect relative employment and payroll over time. In Section 6, we combine the theoretical and empirical impulse response functions of employment and payroll to calibrate the model's parameters.

[17] In Appendix B.1, we use the German Qualification and Working Conditions Survey to show that internet and computer usage is strongly correlated with time spent on cognitive tasks across occupations. We also document that there are no systematic differences in internet and computer usage across different cohorts of workers employed in the same occupation.

[18] Our sample of countries includes Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, United Kingdom, United States. As data sources, we use Eurostat for European countries and IPUMS International for Non-European countries. For all countries, these data sources report the number of persons employed in the following 2-digit ISCO occupations: Managers, Professionals, Technicians and Associate Professionals, Clerical Workers, Service and Sales Workers, Skilled Agricultural Workers, Craft Trades workers, Plant and Machine Operators, and Elementary Occupations.

[19] The German Federal Employment Agency produces the BERUFNET dataset using expert knowledge about the skills required to perform the daily tasks in each occupation. We define an occupation's cognitive intensity as the simple average of the time spent on analytical non-routine and interactive tasks in the years of 2011-2013.

Figure 5 displays the recent trends of employment in cognitive-intensive occupations for several developed countries. The dashed bars indicate that employment in cognitive-intensive occupations has been expanding in 16 out of the 18 countries in our sample. This trend is a reflection of the occupation polarization process documented by Goos, Manning, and Salomons (2009) for European countries, Autor and Dorn (2013) for the United States, and Green and Sand (2015) for Canada.

Figure 5 also shows how employment growth in cognitive-intensive differed for younger and older generations of workers. While older workers increased their employment in cognitive-intensive occupations in most countries, this increase was substantially stronger for younger generations. Across all countries, the average log-change in cognitive employment of younger workers was 75% higher than that of older workers. The young-old gap is higher whenever overall reallocation is higher: across countries, there is a correlation of 0.47 between the young-old gap in cognitive-intensive employment growth and that of all workers. These new stylized facts complement the finding in Autor and Dorn (2009) that the average age of workers employed in contracting middle-wage occupations increased in the United States between 1980 and 2005.

As discussed in Section 5.1, the different employment responses for young and old workers suggest that the skill supply elasticity in the short-run may be substantially different from that in the long-run. However, the aggregate trends in Figure 5 are subject to concerns about potential confounding shocks driving the expansion of cognitive-intensive employment. Moreover, by not relying on a specific innovation, they are not informative about the dynamic adjustment of economies to new technologies. For these reasons, we now consider a detailed investigation of the impact of new technologies on the German labor market.

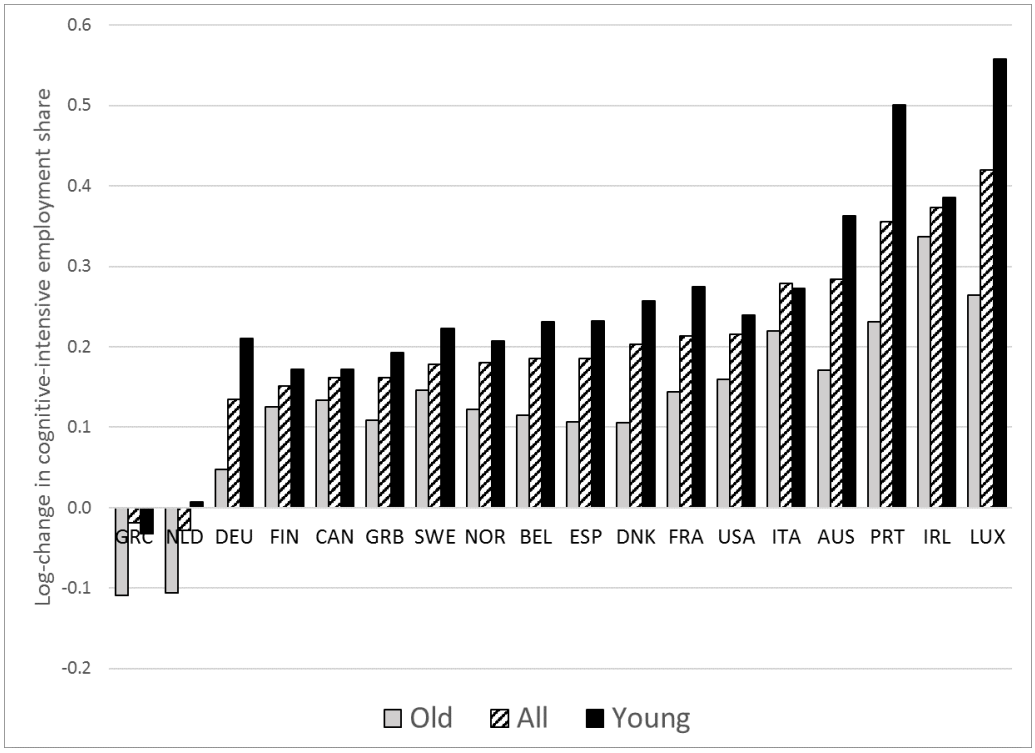
5.3 Within- and Between-Generation Consequences of New Technologies: Evidence from Germany

We next study how the German economy adjusted to recent skill-biased technological shocks. We first describe the data used in our analysis. We then investigate how employment and payroll growth varied with the time spent on cognitive tasks across occupations in Germany. Finally, we exploit quasi-experimental cross-regional variation in adoption timing of broadband internet to estimate the differential impact of this new technology on occupations with a higher cognitive intensity over time.

5.3.1 Data

Our main source of information on German labor market outcomes is the LIAB Longitudinal Model between 1995 and 2014. We follow Card, Heining, and Kline (2013) to construct a sample of full-time employed males aged 20-60 years old residing in West Germany. We first use individual-level information to construct yearly series of employment and payroll

Figure 5: Recent trends in cognitive-intensive employment growth in developed countries



Note. The figure reports the log-change in the share of males employed in cognitive-intensive occupations in 1997-2017 for European countries, in 2000-2010 for the United States, and in 2001-2011 for Canada. Sample of males in two age groups: "Young" workers aged 15-39yrs and "Old" workers aged 40-64yrs. Cognitive-intensive occupations defined as the 3 occupation groups spending more time performing cognitive tasks on the job among the 9 occupation groups in the 2-digit ISCO classification: Managers, Professionals, Technicians and Associate Professionals.

for 120 occupations. While our theory features only two technologies, this is an abstraction, and we obtain more variation empirically by using more detailed occupation information. We therefore now move from the sharp predictions of the two-technology theory to look at employment trends across occupations more generally.

We then construct a second dataset with annual data on employment and payroll for each occupation in 323 regional labor markets. Following Dauth, Findeisen, and Suedekum (2014) and Huber (2018), we use administrative districts to define regional labor markets in West Germany.[20] We use the BERUFNET dataset discussed previously to define each occupation's cognitive intensity as the share of time spent performing analytical non-routine and interactive skills.

We consider labor market outcomes for two generations of workers. We define the "Young" generation as all individuals born after 1960. The young generation was at most 35 years old in the beginning of our period of analysis in 1995, representing 57.5% of the German labor force in that year. Over time, the young generation increased its overall em-

[20] We construct our data using the district of the establishment of the main job of each individual in any given year. Since this information is only available after 1999, we use the establishment's district in 1999 to construct the worker's district affiliation in 1995-1998.

ployment share, reaching around 89% by the end of the analysis period in 2014 (when the young generation was at most 54 years old). Appendix B shows that all qualitative results in this section are robust to defining worker generations using different cohort or age groups.

5.3.2 Cognitive Intensity and Labor Market Outcomes Across Occupations

We now study the relationship between employment growth and cognitive intensity across occupations in Germany. Motivated by the model's predictions in (23)–(24), we estimate the following linear regression for each worker generation g and year t:

$$\log Y^g_{o,t} - \log Y^g_{o,1995} = \beta^g_t \bar{C}_o + \epsilon^g_{o,t} \tag{25}$$

where $Y^g_{o,t}$ is a labor market outcome in occupation o at year t of workers of generation g, and $\bar{C}_o$ is the cognitive intensity of occupation o.[21]

Table 1 reports the estimation of equation (25) in the periods of 1995-2000 (Panel A), 1995-2005 (Panel B), 1995-2010 (Panel C), and 1995-2014 (Panel D). We report the estimated impact of the occupation's cognitive intensity on its log-employment growth in columns (1)–(3) and log-payroll growth in columns (4)-(6).

Over all horizons, columns (1) indicates that occupations with a higher cognitive intensity experienced stronger growth in employment. Compared to the least cognitive-intensive occupation, the employment growth in the most cognitive-intensive occupation was around 143 percent higher by 2014. These results show that the German trends in Figure 5 also hold when we consider variation across occupations with different levels of cognitive intensity. Comparing the responses by generation in columns (2)–(3), we find that the entry into cognitive-intensive occupations was weaker for older generations than for younger generations. In fact, the coefficient estimates for the old generations are about 1/3-1/2 the size of that for the young generations at all horizons. Column (2) shows that young generations display very strong employment growth in occupations with a higher cognitive intensity in all sample periods.

Columns (4)-(6) show the relative payroll responses are slightly stronger than the relative employment responses between 1995 and 2014. This suggests that there were only small changes in the relative wage of more cognitive-intensive occupations over the sample period. In our theory, these relative payroll responses include the rise in the marginal productivity of labor in more cognitive-intensive occupations, as well as the change in overall productivity of workers employed in cognitive-intensive occupation (i.e., the selection effect created by the change in worker allocations).

We now qualitatively relate the evidence in Table 1 to the two main features of our model. The difference in employment responses of old and young generations indicates that the skill

[21] We do not include any controls in our baseline specification. Appendix Table B3 shows that results are similar when we include controls that capture potential confounding effects from the occupation's exposure to immigration and trade shocks in the period of analysis.

Table 1: Cognitive intensity and labor market outcomes across occupations in Germany

Dependent variable:	Employment Growth			Real Payroll Growth		
	All	Young	Old	All	Young	Old
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Change in 1995-2000						
Cognitive intensity	0.388***	0.650***	0.113***	0.340***	0.616***	0.157***
	(0.076)	(0.098)	(0.043)	(0.048)	(0.070)	(0.037)
Panel B: Change in 1995-2005						
Cognitive intensity	0.778***	1.150***	0.290***	0.741***	1.158***	0.404***
	(0.111)	(0.130)	(0.079)	(0.086)	(0.114)	(0.063)
Panel C: Change in 1995-2010						
Cognitive intensity	1.110***	1.523***	0.454***	1.036***	1.525***	0.539***
	(0.137)	(0.149)	(0.125)	(0.111)	(0.133)	(0.091)
Panel D: Change in 1995-2014						
Cognitive intensity	1.488***	1.894***	0.871***	1.535***	2.029***	1.044***
	(0.225)	(0.234)	(0.229)	(0.227)	(0.238)	(0.223)

Note. Sample of 120 occupations. Each panel reports the estimate for the dependent variable over the indicated time period. Young cohort defined as all workers born after 1960 and Old cohort as all workers born before 1960. Robust standard errors in parentheses. *** $p < 0.01$

supply elasticity is much higher in the long-run than in the short-run. In our theory, such a difference arises from the ability of younger generations to adjust their skill investment. In addition, the much weaker relative employment response for old generations suggests that the skill supply elasticity may be low in the short-run. In our theory, this is the case whenever skills are very specific to occupations with the same cognitive content.

5.3.3 Dynamic Adjustment to the Broadband Internet Adoption

The evidence above establishes that employment and payroll responses to the arrival of new technologies differ across worker generations. Although this evidence qualitatively speaks to the main mechanisms in our model, it cannot be quantitatively mapped to the observable predictions of Section 5.1 since it does not allow the estimation of impulse response functions to one-time permanent shocks. Specifically, the occupation-level responses above may be driven by different innovations introduced throughout the period of analysis – e.g. computers, industrial robots, or the internet.

Thus, in this section, we analyze the dynamic response to one cognitive-biased technological innovation: the introduction of broadband internet in the early 2000s.[22] There are two main reasons to focus on this particular innovation in Germany. First, it resembles the one-time permanent shock studied in Section 3 since its adoption was fast: the share of

[22] As shown Akerman, Gaarder, and Mogstad (2015), broadband internet expanded the relative demand for skilled workers in non-routine jobs inside firms. In Appendix B.1, we show that this new technology is disproportionately used by individuals employed in more cognitive-intensive occupations.

households with broadband access increased from 0% in 2000 to over 90% in 2009. Second, it is possible to explore cross-regional variation in adoption timing to estimate the impulse response functions of labor market outcomes for different worker generations. Our strategy relies on the fact that the timing of broadband adoption was spatially heterogeneous: across German districts in 2005, the mean share of household with broadband internet access was 76% and the standard deviation was 16%. In addition, following Falck, Gold, and Heblich (2014), we isolate exogenous spatial variation in adoption timing implied by the suitability of pre-existing local telephone networks for broadband internet transmission.

Empirical Strategy

Our goal is to estimate the dynamic impact of broadband internet adoption on labor market outcomes across districts in Germany. For each year between 1996 and 2014, we estimate the following linear specification

$$Y_{io,t} - Y_{io,1999} = (\alpha_t + \beta_t \bar{C}_o)\, DSL_i + \delta_{o,t} + X_{io,t}\gamma_t + \epsilon_{io,t}, \tag{26}$$

where o denotes an occupation and i district in Germany. In this specification, $Y_{io,t}$ is a labor market outcome for occupation o of district i at year t (either employment or payroll), and DSL_i is the broadband internet penetration in district i in 2005 (normalized to have standard deviation of one). As before, $\bar{C}_o$ is the time-invariant measure of the cognitive intensity of occupation o. The term $\delta_{o,t}$ is an occupation-year fixed-effect that absorbs any confounding shock that has the same impact on occupations in all regions. Similarly, $X_{io,t}$ is a control vector which includes the dependent variable pretrend growth in 1995-1999 and initial district demographic characteristics. These controls account for differential performance of cognitive-intensive occupations in regions with characteristics that may affect the profitability of broadband internet adoption.[23]

To compare responses across generations, we also estimate a similar specifications for different worker generations g:

$$Y^g_{io,t} - Y^g_{io,1999} = \sum_{c \in \{\text{young, old}\}} (\alpha^c_t + \beta^c_t \bar{C}_o)\, 1_{[g=c]} DSL_i + \delta_{o,t} + \zeta_{g,t} + X^g_{io,t}\gamma^g_t + \epsilon^g_{io,t}, \tag{27}$$

where $Y^g_{io,t}$ is a labor market outcome for individuals of cohort g employed in occupation o of district i at year t. As above, we consider two generations: the old generation born before 1960 and the young generation born after 1960. Notice that this specification also includes generation-year fixed effects that capture nationwide trends in employment of different worker cohorts. In this specification, we include the group's pretrend control in $X^g_{io,t}$.

[23] We follow Dix-Carneiro and Kovak (2017) and Freyaldenhoven, Hansen, and Shapiro (2018) by explicitly controlling for pretrends. As argued by the latter paper, pretrends due to unobserved confounding effects might exist even when they are not actually observed in the data, due to estimation error, implying they should be controlled for during estimation. The demographic controls are the college graduate population share, the manufacturing employment share, the immigrant employment share, and the age composition of the labor force. Appendix Table B4 shows results for different control sets.

We are mainly interested on the impact of broadband internet adoption on the relative outcome of cognitive-intensive occupations: β_t in (26) for the all workers, and β_t^g in (27) for generation g. To understand the interpretation of this coefficient, consider region A whose broadband internet penetration in 2005 was one standard deviation higher than that of region B. In each year t, β_t is the difference between regions A and B in the relative outcome of a more cognitive intensive occupation. Similarly, β_t^c is the equivalent difference among workers of cohort c.

The consistent estimation of equations (26)–(27) requires an exogenous source of variation on the adoption of broadband internet across German districts in 2005. However, the cross-regional variation in internet penetration is unlikely to be random since adoption should be faster in regions with workers more suitable to use that technology. For instance, this would be the case if broadband internet expands first in regions with a growing number of young individuals specialized in cognitive-intensive occupations. To circumvent this issue, we follow Falck, Gold, and Heblich (2014) to obtain exogenous variation in broadband internet adoption across German districts stemming from pre-existing conditions of the regional telephone networks. In West Germany, the telephone network constructed in the 1960s used copper wires to connect households to the municipality's main distribution frame (MDF). The initial roll-out of DSL internet access in Germany used these pre-existing copper wire lines to provide high-speed internet to households. As argued by Falck, Gold, and Heblich (2014), the copper wire transmission technology did not support high-speed internet provision over long distances. In fact, provision was impossible in areas located more than 4200m away from an existing main distribution frame (MDF). It was necessary to set up an entirely new system to provide DSL access to areas connected to a MDF located more than 4200m away. Thus, areas initially located close to MDFs obtained broadband internet access before areas located far away from them.

This discussion suggests that the initial location of MDFs is an exogenous shifter of DSL access in 2005. This requires that, conditional on controls, the determinants of MDF construction in the 1960s were orthogonal to the determinants of changes in labor market outcomes in the 2000s, except through their effect on broadband internet penetration in 2005.[24] Building on this idea, we construct two instrumental variables at the district-level that measure the region's population share located in areas where the existing telephone network could not be used to supply high-speed internet. These variables are aggregates of the municipality-level instrumental variables used in Falck, Gold, and Heblich (2014). The first variable is a simple count of the number of municipalities in the district that did not have a MDF within the municipality, and whose population center (measuring as a population-weighted centroid)

[24] While some of these MDFs were built in population centers, others were built in locations where large empty building sites were available. Falck, Gold, and Heblich (2014) provide a detailed discussion of why the main orthogonality assumption is plausible in this setting. Our strategy is similar to the geographic barriers exploited in Akerman, Gaarder, and Mogstad (2015) to estimate the impact of broadband internet on within-firm skill upgrading in Norway. In contrast, our empirical strategy uncovers reduced-form responses in regional outcomes, which combine adjustment margins within and between firms at the regional-level.

was further than the cut-off threshold of 4200m to the MDF used by the municipality. We refer to this variable as the "MDF density measure." The second variable counts the number of municipalities that satisfied the conditions in the first variable, but were further hampered by the lack of any MDFs in neighboring municipalities that were closer than 4200m. The municipalities in the second group required the installation of completely new networks since it was not possible to install copper wire lines connecting them to *any* existing MDF. We refer to this variable as "Alternative MDF availability."

Let Z_i denote the district-level instrument vector with the district's "MDF density measure" and "Alternative MDF availability." Since the observations in equation (26) vary at the occupation-district level, we estimate this equation with an instrument vector that includes Z_i interacted with a constant and the cognitive intensity $\bar{C}_o$ for each occupation o. Similarly, to equation (27), we also interact the instrumental variable vector with dummies for each generation g.

Results

We start by examining the first-stage regression that relates the initial telephone network to DSL access. Although equations (26)–(27) vary by district-occupation or district-occupation-generation, the exogenous variation in the instrument vector is only across districts. Therefore, to provide a clear picture of the exogenous variation underlying the model's first-stage, we first examine the impact of the instrument vector Z_i on the district's share of population with broadband internet access, DSL_i.[25] That is, we begin by estimating the following linear regression for year t:

$$DSI_{i,t} = Z_i\rho_t + X_i\gamma_t + \epsilon_{i,t} \tag{28}$$

where X_i is the vector of district-level controls used in the estimation of (26)–(27).

Table 2 shows that districts with adverse initial conditions for internet adoption had a lower share of households with high-speed internet in 2005. This difference is similar in 2007, and still significant. However, regional differences in broadband penetration are converging throughout the period of analysis; as discussed above, the share of households with broadband access exceeds 90% by 2009. Columns (1) and (3) report the first-stage estimates controlling for the baseline set of district-level controls. We can see that the F statistic of excluded variables remains high in the presence of these controls.

We now turn to the estimation of β_t in (26). Figure 6 reports the estimates of broadband internet expansion on the relative employment (Panel A) and the relative payroll (Panel B) of

[25] As discussed above, equations (26)–(27) have multiple endogenous variables since they include DSL access interacted with occupation cognitive intensity and worker generation dummies. To test for weak instruments in this setting, we provide the Sanderson-Windmeijer F-statistics (Sanderson and Windmeijer, 2016) for the first stage of each specification in Appendix Table B6. This test statistic checks for whether any of our endogenous variables are weakly instrumented, as well as whether there are sufficiently many strong instruments to instrument the multiple endogenous variables. As shown in the table, we obtain uniformly high first-stage SW F-statistics in all specifications.

Table 2: First-stage regressions – Share of households with DSL access in 2005 and 2007

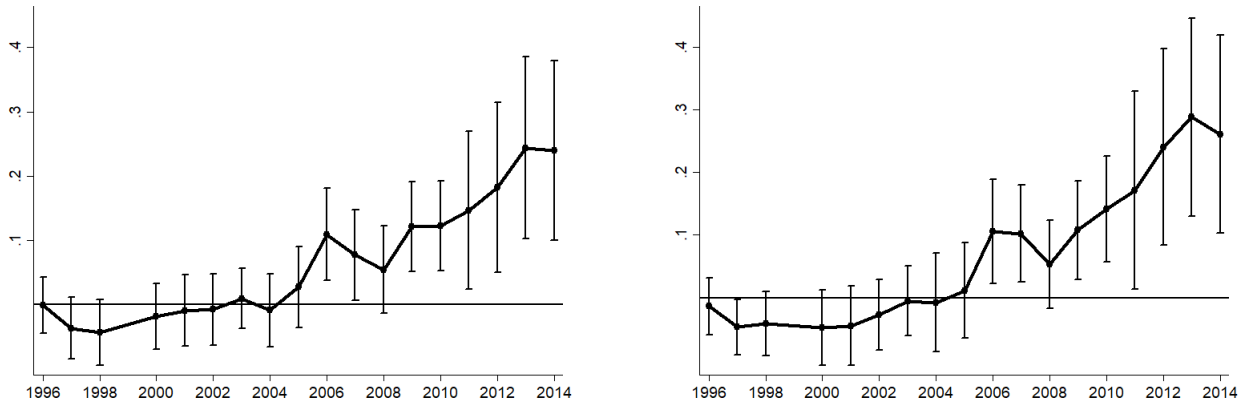
	2005		**2007**	
	(1)	(2)	(3)	(4)
MDF density measure	-0.020***	-0.018***	-0.026***	-0.024***
	(0.005)	(0.005)	(0.005)	(0.005)
Alternative MDF availability	0.002	-0.001	0.002*	-0.00003
	(0.001)	(0.002)	(0.001)	(0.001)
Baseline controls	Yes	No	Yes	No
F statistic	26.49	43.06	23.51	38.62

Note. Sample of 323 districts in West Germany. All regressions are weighted by the district population size in 1999. Baseline controls include the following district variables in 1999: college graduate population share, manufacturing employment share, immigrant employment share and workforce age composition. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

more cognitive-intensive occupations. For both outcomes, we find no evidence of responses in the pre-shock period of 1996-2005. Starting in 2005, our estimates indicate a slow and steady increase in the relative employment of more cognitive-intensive occupations. In 2014, the point estimate suggests that a region with a one-standard deviation higher broadband internet penetration in 2005 had 0.3 log-points higher employment in the most cognitive-intensive occupation than in the least cognitive-intensive occupation.

These results are consistent with the predictions of our model. We interpret the introduction of broadband internet as a positive shock to the relative productivity of the occupations that use this technology more intensively: cognitive-intensive occupations. In line with the results in Section 3, we find that employment and payroll increase more in these occupations. Such a positive impact becomes larger along the transition to the new stationary equilibrium.

Figure 6: Impact of early DSL adoption on more cognitive-intensive occupations: All generations



(a) Relative aggregate employment response

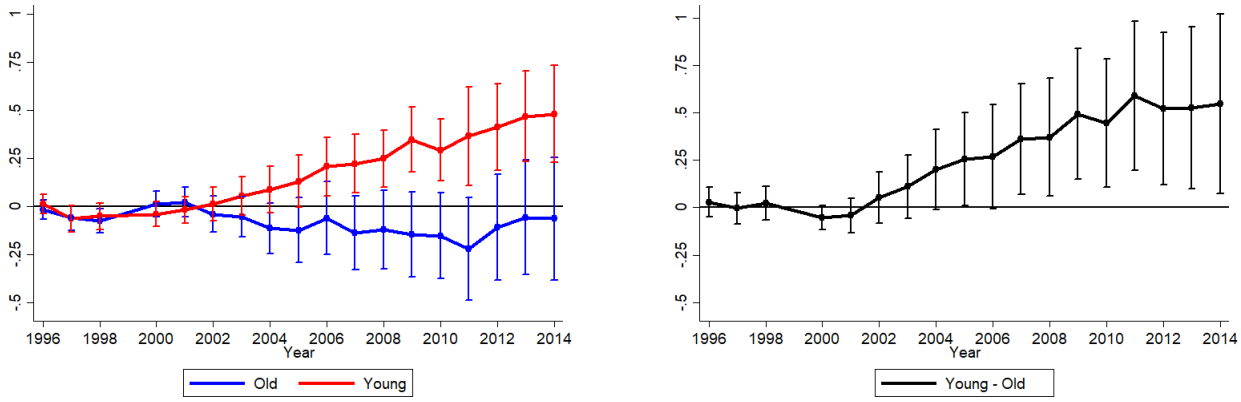
(b) Relative aggregate wagebill response

Note. Estimation of equation (26) in the sample of 120 occupations and 323 districts. Dependent variable: log employment (left) and log payroll (right). All regressions are weighted by the district population size in 1999 and include occupation-time fixed-effects. Baseline controls include the following district variables in 1999: college graduate population share, manufacturing employment share, immigrant employment share, district age composition, and the dependent variable pre-shock growth in 1995-1999. For each year, the dot is the point estimate of β^t, and the bar is the associated 90% confidence interval implied by the standard error clustered at the district level.

We now turn to the estimation of employment responses for each generation: β_t^{old} and β_t^{young} obtained from (27). Panel A of Figure 7 reports the estimates for each year between 1996 and 2014. Prior to 2003, regions with early DSL expansion did not experience differential growth in the relative outcomes of cognitive-intensive occupations for old and young workers. After 2005, we find a significant impact on the relative employment of young cohorts in cognitive-intensive occupations. In contrast, we do not find such an effect for old cohorts – if anything, the effect is negative. Panel B of Figure 7 shows that the between-generation difference in relative employment growth is statistically significant in every year after 2006. In line with our model's prediction, the between-generation component grows shortly after the shock and then starts to stabilize.

We can use our model to interpret the results in Figure 7. The small relative employment response of old generations suggests that the short-run skill supply elasticity is close to zero. In this case, old generations do not switch occupations as their skills would have a lower value in the more cognitive-intensive occupations augmented by the technological innovation. Alternatively, the positive between-generation difference in the relative employment response indicates that incoming cohorts adapt their skill investment decision towards skills appropriate for performing cognitive-intensive jobs. This suggests that the elasticity of skill supply is positive in the long-run.

Figure 7: Impact of early DSL adoption on more cognitive-intensive occupations: Within- and between generations



(a) Relative employment response by generation

(b) Between employment response and aggregate

Note. Estimation of equation (27) in the sample of 2 cohorts, 120 occupations and 323 districts. Dependent variable: log employment. All regressions are weighted by the district population size in 1999 and include occupation-time and cohort-time fixed-effects. Baseline controls include the following district variables in 1999: college graduate population share, manufacturing employment share, immigrant employment share, district age composition, and the dependent variable pretrend growth in 1995-1999. For each year, the dot is the point estimate of β^t, and the bar is the associated 90% confidence interval implied by the standard error clustered at the district level.

5.3.4 Robustness

Appendix B.3 presents additional results that attest the robustness of the findings presented in this section. First, Tables B2 and B3 present robustness exercises of the cross-occupations

estimates presented in Table 1. Specifically, we show that the positive relative employment growth in cognitive occupations is driven by the top third of occupations by cognitive intensity (for all workers and separately for each worker generation). We also report similar estimates when restricting the sample to native-born Germans, changing the definition of the young generation, or including controls for each occupation's exposure to trade and immigration shocks.

In Appendix Tables B4 and B5, we evaluate the robustness of the results in Figures 6-7 to variations in both our sample definition and control set. As shown in Appendix Table B4, results are qualitatively similar when we drop the pretrend control, but estimated coefficients are less precise and slightly smaller in magnitude. We also attest that results are similar when controlling for district-generation-year fixed-effects. This is reassuring as this restrictive set of controls absorb all potential confounding shocks that affect each district-generation pair in a year. In this case, identification comes purely from the differential effect of DSL access expansion on occupations with a higher cognitive intensity. That is, this control set captures any variation that might have resulted in a district receiving broadband access early, including differential immigration into a district that received DSL or differential aging or birth patterns in the district over time.

Table B5 presents results when we vary the definition of the young generation, as well as we restrict the sample to only native-born males. We consider several definitions of the young generation: those born after 1955, 1965 or 1970, and those aged less than 35, 40 or 45 in each year. Once again, the results are qualitatively similar across specifications. However, in line with our theory, the estimated coefficients for the young generation are stronger when we restrict the young generation to include only those born in more recent years. In addition, we consider a sample only with workers employed in small establishments. The rationale is that larger establishments might have received DSL access prior to the roll-out across Germany. In this case, we would expect adjustment in these establishments to have occurred earlier, biasing our results to zero. In line with this intuition, we find that our results are quantitatively stronger in the sample of small establishments.

6 Quantitative Analysis

We conclude our analysis by using our theory to quantify the magnitude of the impact that cognitive-biased innovations have on the economy over different horizons. We first describe the calibration of the main parameters in our theory using the empirical impulse response functions reported in Section 5. We then analyze the economy's adjustment to a large shock that increases the employment share in cognitive-intensive occupations from 25% to 50%. For this exercise, we quantify the biases generated by ignoring the dynamics of the skill distribution across generations.

6.1 Calibration

As in Section 5, we map the H technology in our theory to the set of production activities performed by cognitive-intensive occupations. We calibrate our theory in two steps. In the first step, we exogenously specify a subset of parameters and functions in the theory. We set the discount rate to match an annual interest rate of 2%, $\rho = 0.02$. We calibrate the elasticity of substitution across cognitive and non-cognitive intensive occupations to $\theta = 3$. Finally, for all welfare calculations, we specify welfare-weights re^{-rt} with $r = \rho + \delta$ so that the social discounting of future generations is identical to the discounting of worker's future utility.

We also specify functional forms for the productivity of skill types in the two technologies. We abstract from differences in non-cognitive productivity across skills by normalizing $\alpha(i) \equiv 1$. This implies that, for any given worker generation, employment and payroll responses are both driven by the degree of technology-skill specificity in the economy.[26] In addition, we assume that $\sigma(i)$ takes the form of a logistic function:

$$\sigma(i) = \frac{e^{\sigma(i-l)}}{1 + e^{\sigma(i-l)}}$$

where l is the assignment threshold in the initial stationary equilibrium. This specification is a tractable manner of capturing technology-skill specificity in the economy. It implies that the equilibrium exists for any $\sigma > 0$ since the relative productivity is bounded. Also, by setting the midpoint of the function to l, the parameter σ controls the elasticity of $\sigma(i)$ for the marginal skill-types in the initial equilibrium (i.e., i close to l). Thus, σ specifies the magnitude of the the short-run skill supply elasticity, η.

In the second step, we use the estimated responses of Section 5 to calibrate (δ, σ, ψ). We select $\delta = 0.057$ and $\tilde{x}_0 = 40\%$ to match the decline in the share of the old generation in total employment from 40% in 1997 to 15% in 2014. In line with the discussion in Section 5.1, we select ψ to match the estimated impulse response function of the between-generation difference in relative cognitive-intensive employment. The positive estimated coefficients in Panel B of Figure 7 yields $\psi = 0.35$. In addition, we select η to match the estimated impulse response function of the relative cognitive-intensive employment of the old generation. Due to the nonsignificant estimates reported in Panel B of Figure 7, we set the short-run skill supply to the low value of $\eta = 0.02$. We formally present the calibration procedure in Appendix B.4, along with an analysis of the model fit.[27]

[26] The function form of $\alpha(i)$ controls how labor earnings respond to changes in the employment composition across technologies – for a discussion, see Adão (2016). Alternative specifications of $\alpha(i)$ can thus be used to match responses in relative earnings for different worker generations.

[27] In this calibration, we select the distribution of innate ability distribution, $\bar{s}(i)$, to generate a uniform distribution of skills in the initial equilibrium: $s_0(i) \equiv 1$. In our theory, this normalization is innocuous since it does not affect changes in the skill distribution for a given change in q (Lemma 4) and η matches the short-run employment change.

6.2 Dynamic Adjustment to Cognitive-biased Technological Innovations

We now present the quantitative predictions of our theory regarding the economy's dynamic adjustment to the arrival of cognitive-biased technologies. We calibrate the magnitude of the shock, $\Delta \log A$, to match a long-run increase in the employment share in cognitive-intensive occupations from 25% to 50%. These values approximately correspond to the cognitive-intensive employment share in 1997 of the countries with the lowest and the highest cognitive-intensive employment share among those listed in Figure 5 (Portugal and Netherlands, respectively). Thus, our quantitative results can be seen as analyzing the transitional dynamics of a skill-biased shock that generates international convergence in cognitive-intensive employment shares. Given our calibration, this interpretation requires all countries to have the same parameters of short- and long-run skill supply elasticity of Germany.

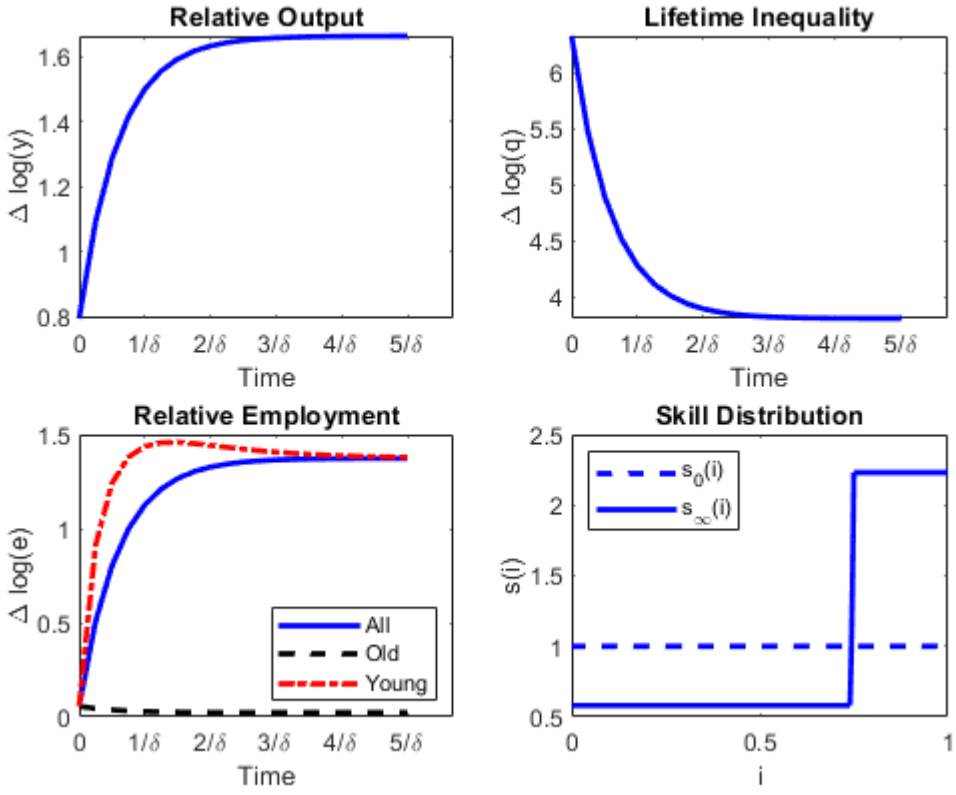
Figure 8 presents the predicted impulse response functions of labor market outcomes. Consider first response at impact. Given that our theory abstracts from several additional sources of dynamics, it would be wrong to interpret the impact adjustment as happening instantaneously in reality. We view this short-run response as capturing changes over the time window encompassing dynamic forces triggered by other variables that are likely to move faster than the distribution of skills (e.g., physical capital). In other words, we prefer to interpret the "length" of the impact adjustment as related to the time that it takes for such faster moving variables to converge to the new long-run equilibrium.

Results show that there is a substantial increase in the relative cognitive-intensive output in the short-run. This large response is a consequence of the large magnitude of the shock. This becomes clear when we take into account that relative employment almost does not change at impact because of the low value of the short-run skill supply elasticity. The combination of the large increase in relative output and the small increase in relative employment translates into large changes in lifetime inequality.

Our results also indicate that the responses in all outcomes change substantially over time (measured in terms of worker generations, $1/\delta \approx 18 yrs$). Over the course of the two generations following the shock, the responses in relative output doubles in magnitude due to the reallocation of workers across technologies. Such a reallocation is entirely driven by incoming generations of young workers. This pattern is a consequence of the change in the skill distribution across generations. The bottom right panel shows that the initial spike in lifetime inequality induces young workers to invest in high-i skills allocated to cognitive-intensive occupations. The economy's slow process of skill accumulation triggers a decline in lifetime inequality, which recedes by more than 30% over the course of two generations.

Table 3 presents our quantification of the innovation's impact on the economy's average welfare and lifetime inequality. Using the consumption equivalent measures defined in Section 4.1, the gain in average welfare across all generations, $\Delta \bar{U} * (\rho + \delta)$, is 52% and the increase in lifetime welfare inequality is 38%. These large welfare effects follow from the

Figure 8: Transitional dynamics to a cognitive-biased innovation at $t = 0$



Note. The figure reports the theoretical impulse response function with a shock calibrated to increase the employment share in cognitive-intensive occupations from 25% to 50% between stationary equilibria. Baseline calibration described in Section 6.1.

substantial shock size necessary to induce the reallocation of one-fourth of the economy's labor force.

Next, we compare these figures to those obtained with two simple calculations that ignore the economy's transition dynamics across generations. The 'Short-run' calculation assumes that changes observed at impact are permanent, while the 'Long-run' calculation assumes that the changes observed in the long-run were permanent and happened at impact. We can see that these two calculations lead to substantial biases in welfare analysis. The 'Short-run' calculation severely understates the average welfare gains and overstates the inequality increases. The opposite is true for the 'Long-run' calculation. These biases arise because both approaches do not capture the adjustment in relative output and wages along the transition. Precisely, the 'Short-run' approach misses the future accumulation of skills that increases relative output– thus reducing the ideal price index and increasing average real wages– and reduces relative wage of cognitive-intensive occupations. In contrast, the 'Long-run' approach misses the fact that it takes generations for the economy to accumulate the cognitive-biased skills necessary to achieve the levels of relative output and wages observed in the long-run.

Our last exercise instead quantifies the potential biases resulting from confounding short-

Table 3: Welfare consequences of a cognitive-Biased innovation

Change in:	Average welfare	Lifetime welfare inequality
	$\Delta\bar{U} * (\rho + \delta)$	$\Delta\bar{\Omega} * (\rho + \delta)$
Baseline	52%	38%
Short-run	25%	49%
Long-run	70%	29%

Note. The table reports the the changes in average welfare $\Delta\bar{U}$ and lifetime welfare inequality $\Delta\bar{\Omega}$ implied by a shock calibrated to increase the employment share in cognitive-intensive occupations from 25% to 50% between stationary equilibria. Baseline calibration described in Section 6.1. 'Baseline' corresponds to the measures defined in Section 4.1 that account for the economy's transitional dynamics. 'Short-run' assumes that changes at impact are permanent. 'Long-run' assumes that long-run changes happened at impact.

and long-run skill supply elasticities. As in Section 4.2, we consider the conclusions obtained by a researcher that uses a static assignment model to study technological transitions resulting from our baseline economy (shown in Figure 8). This corresponds to studying a special case of our theory with $\psi = 0$. We consider a situation in which the researcher externally calibrates the degree of technology-skill specificity to match estimates in the literature of the equivalent of our theory's short-run skill supply elasticity. In particular, we target typical estimates of 0.5 obtained from the degree of wage dispersion in a cross-section of workers (Hsieh et al., 2013).[28] We then let the researcher calibrate the shock magnitude, $\Delta \log(A)$, to match the observed response of relative output at a fixed horizon of one-quarter of a generation after impact.[29]

Table 4 shows the biases in welfare calculations from the above exercise. The results are similar to the ones from the exercise in Table 3, except that the average welfare bias is somewhat smaller. The reduction in the bias follows from the fact that this calibration already captures part of the increase in relative output that happens along the transition. Nevertheless, this calibrated static model still severely understates the true average welfare increase and overstates the lifetime welfare inequality.

Table 4: Welfare biases: Baseline dynamic model $\times$ Static assignment model

Baseline ($\eta \approx 0, \psi = 0.35$)		Static ($\eta = 0.5, \psi = 0$)	
$\Delta\bar{U} * (\rho + \delta)$	$\Delta\bar{\Omega} * (\rho + \delta)$	$\Delta\bar{U} * (\rho + \delta)$	$\Delta\bar{\Omega} * (\rho + \delta)$
52%	38%	38%	46%

[28] Estimates in the literature are typically not reported for the short-run skill supply elasticity η. Instead, they report the elasticity of relative employment with respect to a change in relative wages: $\frac{\Delta \log(e)}{\Delta \log(\omega)}$. The typical estimate for this elasticity is about 1. Given that the new long-run equilibrium employment share is 0.5 in our exercise, a reported relative employment share elasticity of 1 implies an $\eta = 0.5$.

[29] Given our calibration for δ, this implies an "observation window" of about 5 years after impact. As discussed above, this window also includes the adjustment period of variables with a faster convergence rate than changes in the skill distribution over generations.

7 Conclusion

In this paper, we developed a tractable theory to study how skill-biased technological innovations affect the economy over different horizons. Our theory has three central ingredients: (i) overlapping generations of workers with heterogeneous technology-specific skills, (ii) endogenous sorting of worker skill-types to technologies, and (iii) forward-looking investment in different skills by young workers. Following a technology-improving innovation, we characterize in closed-form the dynamic responses of labor market outcomes for different worker generations as a function of parameters governing the technology-skill specificity and the cost of adjusting skill investment. We then show how these dynamic responses affect the measurement of the welfare consequences of new technologies.

A novel insight from the theory is that economies will adjust more slowly to skill-biased technological innovations when technology-skill specificity is higher. This is because the larger increases in relative wages (as a result of a lower short-run skill supply elasticity) induce larger, more persistent changes in the skill distribution across generations. As such, responses of labor market outcomes at shorter and longer horizons significantly differ. We show that the average and distributional welfare implications of technological innovations are biased when ignoring the adjustment across generations by extrapolating from changes at short or long horizons alone. These biases are severe precisely in economies with large observed changes in inequality and differences in outcomes across generations because they suggest a high degree of skill specificity at short horizons and an elastic supply of skills at longer horizons.

Finally, we empirically investigate these issues by analyzing the adjustment of developed economies to recent cognitive-biased technological innovations, with a particular focus on Germany. Several distinct pieces of evidence show strong responses in employment and payroll in cognitive-intensive occupations to such innovations. Yet, such responses are strong for young workers but weaker for older workers, suggesting that the cognitive-skill supply elasticity is low at short horizons (a high degree of skill-specificity) but higher at longer horizons. Calibrating our model to match this evidence, we conclude that ignoring the slow adjustment across generations by only considering short-run changes severely understates the average welfare benefits and overstates the lifetime inequality increases following cognitive-biased technological innovations. Considering long-run changes alone generates biases in the opposite direction.

Taken together, we derive three broad lessons from this piece. First, recent technological transitions may have been slower and more unequal than those of the past due to differences in the specificity of skills, themselves a consequence of differences in the sectors and occupations involved as well as the educational and training institutions. Second, that caution should be exercised when interpreting technological transitions based only on empirical observations that span much less than a generation. This may lead to overly pessimistic views

of the consequences of new technologies for inequality and average welfare. Specially when the innovation affects activities with a high degree of technology-skill specificity due to the slow moving nature of such transitions. Third, that empirical observations at short horizons are indeed valuable if complemented with observations for different generations. Looking at the decisions of younger workers allows us to "see the future" and thus appropriately derive the full implications for average welfare and inequality of technological innovations.

References

Acemoglu, Daron and David Autor. 2011. "Skills, tasks and technologies: Implications for employment and earnings." In *Handbook of labor economics*, vol. 4. Elsevier, 1043–1171.

Acemoglu, Daron and Pascual Restrepo. 2017. "Robots and Jobs: Evidence from US Labor Markets." Working Paper 23285, National Bureau of Economic Research.

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

Adão, Rodrigo. 2016. "Worker heterogeneity, wage inequality, and international trade: Theory and evidence from brazil." *Unpublished paper, MIT* .

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

Atkeson, Andrew, Ariel Burstein, and Manolis Chatzikonstantinou. 2018. "Transitional Dynamics in Aggregate Models of Innovative Investment." Tech. rep., National Bureau of Economic Research.

Atkeson, Andrew and Patrick J. Kehoe. 2007. "Modeling the Transition to a New Economy: Lessons from Two Technological Revolutions." *American Economic Review* 97 (1):64–88.

Atkin, David. 2016. "Endogenous Skill Acquisition and Export Manufacturing in Mexico." *American Economic Review* 106 (8):2046–85.

Autor, David and David Dorn. 2009. "The Skill Content of Jobs and the Evolution of the Wage Structure-This Job is" Getting Old": Measuring Changes in Job Opportunities using Occupational Age Structure." *American Economic Review* 99 (2):45.

Autor, David and Anna Salomons. 2017. "Robocalypse now: does productivity growth threaten employment?" *NBER Chapters* .

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

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

Ben-Porath, Yoram. 1967. "The production of human capital and the life cycle of earnings." *Journal of political economy* 75 (4, Part 1):352–365.

Blanchard, Olivier J. 1985. "Debt, deficits, and finite horizons." *Journal of political economy* 93 (2):223–247.

Buera, Francisco J and Joseph P Kaboski. 2012. "The rise of the service economy." *American Economic Review* 102 (6):2540–69.

Burstein, Ariel, Eduardo Morales, and Jonathan Vogel. 2016. "Changes in between-group inequality: computers, occupations, and international trade." Tech. rep., mimeo.

Card, David, Joerg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *The Quarterly Journal of Economics* 128 (3):967–1015.

Card, David and Thomas Lemieux. 2001. "Can falling supply explain the rising return to college for younger men? A cohort-based analysis." *The Quarterly Journal of Economics* 116 (2):705–746.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil. 2011. "Estimating marginal returns to education." *American Economic Review* 101 (6):2754–81.

Caselli, Francesco. 1999. "Technological revolutions." *American economic review* 89 (1):78–102.

Chari, Varadarajan V and Hugo Hopenhayn. 1991. "Vintage human capital, growth, and the diffusion of new technology." *Journal of political Economy* 99 (6):1142–1165.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. Forthcoming. "Housing Booms and Busts, Labor Market Opportunities, and College Attendance." *American Economic Review* .

Costinot, Arnaud and Jonathan Vogel. 2010. "Matching and inequality in the world economy." *Journal of Political Economy* 118 (4):747–786.

Dauth, Wolfgang, Sebastian Findeisen, and Jens Suedekum. 2014. "The rise of the East and the Far East: German labor markets and trade integration." *Journal of the European Economic Association* 12 (6):1643–1675.

Dix-Carneiro, Rafael and Brian K. Kovak. 2017. "Trade Liberalization and Regional Dynamics." *American Economic Review* 107 (10):2908–46.

Durlauf, Steven N and Philippe Aghion. 2005. "Handbook of economic growth." .

Falck, Oliver, Robert Gold, and Stephan Heblich. 2014. "E-lections: Voting Behavior and the Internet." *American Economic Review* 104 (7):2238–65.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M Shapiro. 2018. "Pre-event Trends in the Panel Event-study Design." Working Paper 24565, National Bureau of Economic Research.

Galor, Oded and Omer Moav. 2002. "Natural Selection and the Origin of Economic Growth." *The Quarterly Journal of Economics* 117 (4):1133–1191.

Goos, Maarten, Alan Manning, and Anna Salomons. 2009. "Job polarization in Europe." *American economic review* 99 (2):58–63.

Green, David A and Benjamin M Sand. 2015. "Has the Canadian labour market polarized?" *Canadian Journal of Economics/Revue canadienne d'économique* 48 (2):612–646.

Hansen, Lars Peter and Thomas J Sargent. 2008. *Robustness*. Princeton university press.

Hayashi, Fumio. 1982. "Tobin's marginal q and average q: A neoclassical interpretation." *Econometrica: Journal of the Econometric Society* :213–224.

Herrendorf, Berthold, Christopher Herrington, and Akos Valentinyi. 2015. "Sectoral technology and structural transformation." *American Economic Journal: Macroeconomics* 7 (4):104–33.

Herrendorf, Berthold, Richard Rogerson, and Ákos Valentinyi. 2014. "Growth and structural transformation." In *Handbook of economic growth*, vol. 2. Elsevier, 855–941.

Hobijn, Bart, Todd Schoellman, and Alberto Vindas. 2019. "Wages during Structural Transformation: The Importance of Cohort Labor Supply Decisions." .

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow. 2013. "The allocation of talent and us economic growth." Tech. rep., National Bureau of Economic Research.

Huber, Kilian. 2018. "Disentangling the Effects of a Banking Crisis: Evidence from German Firms and Counties." *American Economic Review* 108 (3):868–98. URL http://www.aeaweb.org/articles?id=10.1257/aer.20161534.

Katz, Lawrence F and Kevin M Murphy. 1992. "Changes in relative wages, 1963–1987: supply and demand factors." *The quarterly journal of economics* 107 (1):35–78.

Kim, Dae-Il and Robert H Topel. 1995. "Labor markets and economic growth: lessons from Korea's industrialization, 1970-1990." In *Differences and changes in wage structures*. University of Chicago Press, 227–264.

Lagakos, David and Michael E. Waugh. 2013. "Selection, Agriculture, and Cross-Country Productivity Differences." *American Economic Review* 103 (2):948–80.

Lee, Donghoon and Kenneth I Wolpin. 2006. "Intersectoral labor mobility and the growth of the service sector." *Econometrica* 74 (1):1–46.

Lucas, Robert E and Benjamin Moll. 2014. "Knowledge growth and the allocation of time." *Journal of Political Economy* 122 (1):1–51.

Matsuyama, Kiminori. 1992. "A simple model of sectoral adjustment." *The Review of Economic Studies* 59 (2):375–387.

Moll, Benjamin, Lukasz Rachel, and Pascual Restrepo. 2019. "Uneven Growth: Automation's Impact on Income and Wealth Inequality." .

Ngai, L Rachel and Christopher A Pissarides. 2007. "Structural change in a multisector model of growth." *American economic review* 97 (1):429–443.

Perla, Jesse and Christopher Tonetti. 2014. "Equilibrium imitation and growth." *Journal of Political Economy* 122 (1):52–76.

Porzio, Tommaso and Gabriella Santangelo. 2019. "Does Schooling Cause Structural Transformation?" .

Roy, Andrew Donald. 1951. "Some thoughts on the distribution of earnings." *Oxford economic papers* 3 (2):135–146.

Sanderson, Eleanor and Frank Windmeijer. 2016. "A weak instrument F-test in linear IV models with multiple endogenous variables." *Journal of Econometrics* 190 (2):212–221.

Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of monetary Economics* 50 (3):665–690.

Spitz-Oener, Alexandra. 2006. "Technical change, job tasks, and rising educational demands: Looking outside the wage structure." *Journal of labor economics* 24 (2):235–270.

Tobin, James. 1969. "A general equilibrium approach to monetary theory." *Journal of money, credit and banking* 1 (1):15–29.

Yaari, Menahem E. 1965. "Uncertain lifetime, life insurance, and the theory of the consumer." *The Review of Economic Studies* 32 (2):137–150.

Young, Alwyn. 2014. "Structural transformation, the mismeasurement of productivity growth, and the cost disease of services." *American Economic Review* 104 (11):3635–67.

Appendix A Proofs

A.1 Proof of Lemma 2

We obtain (14) by applying this expression into the relative supply expression in (13) and the relative demand expression in (2). Existence and uniqueness follow from applying Bolzano's theorem to (14). The left-hand side is strictly decreasing in ω_t, converges to zero as $\omega_t \to \infty$, and converges to infinity as $\omega_t \to 0$. Notice that $l(\omega_t)$ is decreasing in ω_t and $l(\omega_t) \in [0,1]$. Thus, the right-hand-side is strictly increasing in $l(\omega_t)$, it converges to infinity as $l(\omega_t) \to 0$ and it converges to zero if $l(\omega_t) \to 1$.

A.2 Proof Lemma 3

The FOC of workers' skill-accumulation problem are:

$$V_t(i) - \frac{1}{\psi}\left(1 + log\left(\frac{\tilde{s}_t(i)}{\bar{s}_t(i)}\right)\right) - \lambda_t = 0$$
$$\lambda_t\left(\int_0^1 \tilde{s}_t(x)dx - 1\right) = 0$$

Integrating over $i \in [0,1]$, we obtain an equation characterizing λ_t:

$$log\left(\int_0^1 \bar{s}_t(i)e^{\psi V_t(i)}di\right) = \psi\lambda_t + 1$$

Therefore,

$$\tilde{s}_t(i) = \frac{\bar{s}_t(i)e^{\psi V_t(i)}}{\int_0^1 \bar{s}_t(j)e^{\psi V_t(j)}dj}.$$

Using the wage expressions and assignment function in Lemma 1, we can write the value function of a worker i at time t as

$$\begin{aligned} V_t(i) &= \int_t^\infty e^{-(\rho+\delta)(s-t)} log(w_s(i))ds \\ &= \int_t^\infty e^{-(\rho+\delta)(s-t)} \left(log(\omega_s\sigma(i)\alpha(i))\mathbb{I}_{i\geq l_s} + log(\alpha(i))\left(1 - \mathbb{I}_{i<l_s}\right)\right)ds \\ &= \frac{log(\alpha(i))}{\rho+\delta} + \int_t^\infty e^{-(\rho+\delta)(s-t)} log\left(\omega_s\sigma(i)\right)\mathbb{I}_{i\geq l_s}ds \end{aligned}$$

By defining $Q_t(i) \equiv e^{\int_t^\infty e^{-(\rho+\delta)(s-t)} log(\omega_s\sigma(i))\mathbb{I}_{i\geq l_s}ds}$, we obtain

$$\tilde{s}_t(i) = \frac{\bar{s}_t(i)\alpha(i)^{\frac{\psi}{\rho+\delta}}Q_t(i)^\psi}{\int_0^1 \bar{s}_t(j)\alpha(j)^{\frac{\psi}{\rho+\delta}}Q_t(j)^\psi dj}.$$

A.3 Proof of Lemma 4

Notice that $\int s(i)\hat{\tilde{s}}_t(i)di = \int(\tilde{s}_t(i) - s(i))di = 0$. Using (A.7) , we have that

$$\begin{aligned}
0 =& \int_0^1 s(i)\hat{\tilde{s}}_t(i)di \\
=& \hat{\tilde{s}}_t(l) - \psi \int_0^l \left(\hat{q}_t \mathbb{I}_{i<l_t} + \hat{q}_{t+\tau(i)}\mathbb{I}_{i\in(l_t,l)}\right) s(i)di \\
=& \hat{\tilde{s}}_t(l) - \left(\int_0^l s(i)di\right)\psi\hat{q}_t + \psi\int_{l_t}^l \left(\hat{q}_t - \hat{q}_{t+\tau(i)}\right) s(i)di
\end{aligned}$$

We can use use the same arguments as in A.4 to show that the last term is of second order. Thus,

$$\hat{\tilde{s}}_t(l) = \left(\int_0^l s(i)di\right)\psi\hat{q}_t$$

and, therefore,

$$\hat{\tilde{s}}_t(i) = \left(\int_0^l s(i)di\right)\psi\hat{q}_t - \psi\hat{q}_t\mathbb{I}_{i<l} + \psi(\hat{q}_t - \hat{q}_{t+\tau(i)})\mathbb{I}_{i\in(l_t,l)}.$$

To prove the result, we use that $\hat{q}_{t+\tau(i)} = \hat{q}_t e^{-\lambda\tau(i)}$. So,

$$\begin{aligned}
\hat{\tilde{s}}_t(i) &= \left(\int_0^l s(i)di\right)\psi\hat{q}_t - \psi\hat{q}_t\mathbb{I}_{i<l} + \psi(\hat{q}_t - \hat{q}_{t+\tau(i)})\mathbb{I}_{i\in(l_t,l)} \\
&= \mathbb{I}_{i>l}\psi\hat{q}_t - \left(1 - \int_0^l s(i)di\right)\psi\hat{q}_t + \psi\hat{q}_t(1 - e^{-\lambda\tau(i)})\mathbb{I}_{i\in(l_t,l)} \\
&= \left(\mathbb{I}_{i>l} - \int_l^1 s(i)di\right)\psi\hat{q}_t + o_t(i)
\end{aligned}$$

where $o_t(i) \equiv \psi\hat{q}_t(1 - e^{-\lambda\tau(i)})\mathbb{I}_{i\in(l_t,l)}$ and has $\int s(i)o_t(i)di = 0$.

Finally, the dynamics of the skill distribution and the threshold were already derived in equations A.1 and A.2.

A.4 Proof of Proposition 1

First, we do a first order approximation around the stationary equilibrium of equations (10), (12) and (14). We obtain:

$$\frac{\partial\hat{s}_t(i)}{\partial t} = -\delta\hat{s}_t(i) + \delta\hat{\tilde{s}}_t(i) \tag{A.1}$$

$$\hat{l}_t = \frac{\eta}{\theta - 1}\hat{y}_t \tag{A.2}$$

$$\hat{l}_t = \frac{\eta}{\kappa\eta + \theta}\left(\int_l^1 \hat{s}_t(i)\frac{\alpha(i)\sigma(i)s(i)}{\int_l^1 \alpha(i)\sigma(i)s(i)di}di - \int_0^l \hat{s}_t(i)\frac{\alpha(i)s(i)}{\int_0^l \alpha(i)s(i)di}di\right) \tag{A.3}$$

where

$$\kappa \equiv \frac{\alpha(l)s(l)l}{\int_0^l \alpha(i)s(i)di} + \frac{\alpha(l)\sigma(l)s(l)l}{\int_l^1 \alpha(i)\sigma(i)s(i)di}.$$

Differentiating (A.3) annd (A.2) with respect to time, we get that

$$\frac{\partial \hat{y}_t}{\partial t} = \frac{\theta - 1}{\kappa\eta + \theta}\left(\int_l^1 \frac{\partial \hat{s}_t(i)}{\partial t}\frac{\alpha(i)\sigma(i)s(i)}{\int_l^1 \alpha(i)\sigma(i)s(i)di}di - \int_0^l \frac{\partial \hat{s}_t(i)}{\partial t}\frac{\alpha(i)s(i)}{\int_0^l \alpha(i)s(i)di}di\right)$$

Applying (A.1) to this expression, we obtain

$$\frac{\partial \hat{y}_t}{\partial t} = -\delta\hat{y}_t + \frac{\theta - 1}{\kappa\eta + \theta}\delta\left(\int_l^1 \hat{\tilde{s}}_t(i)\frac{\alpha(i)\sigma(i)s(i)}{\int_l^1 \alpha(i)\sigma(i)s(i)di}di - \int_0^l \hat{\tilde{s}}_t(i)\frac{\alpha(i)s(i)}{\int_0^l \alpha(i)s(i)di}di\right). \tag{A.4}$$

Furthermore, we will guess an verify that l_t converges monotonically along the equilibrium path. We show the proof starting from $\hat{l}_0 < 0$. The proof for $\hat{l}_0 > 0$ is analogous and omitted.

Whenever $\hat{l}_0 < 0$ and increases monotonically along the equilibrium path, we have that for all $s > t$, types $i < l_t$ are employed in technology L and types $i > l$ are employed in technology H. Also, for workers with $i \in (l_t, l)$, there exist a $\tau(i)$ such that they work in H for all $t < s < t + \tau(i)$ and in L for all $s > t + \tau(i)$.

Then, from equation (15), we have

$$Q_t(i) = \begin{cases} 1 & i \leq l_t \\ e^{\int_t^{t+\tau(i)} e^{-(\rho+\delta)(s-t)} log(\omega_s \sigma(i))ds} & i \in (l_t, l) \\ \sigma(i)^{\frac{1}{\rho+\delta}} q_t & i > l \end{cases} \tag{A.5}$$

So, we can write the optimal lottery as

$$\tilde{s}_t(i) = \begin{cases} \frac{\tilde{s}(i)}{\tilde{s}(l)}\tilde{s}_t(l)e^{-\psi\int_t^\infty e^{-(\rho+\delta)(s-t)}log\left(\frac{\omega_s}{\omega}\right)ds} & i \leq l_t \\ \frac{\tilde{s}(i)}{\tilde{s}(l)}\left(\frac{\sigma(i)}{\sigma(l)}\right)^{\frac{\psi}{\rho+\delta}(1-e^{-(\rho+\delta)\tau(i)})}\tilde{s}_t(l)e^{-\psi\int_{t+\tau(i)}^\infty e^{-(\rho+\delta)(s-t)}log\left(\frac{\omega_s}{\omega}\right)ds} & i \in (l_t, l) \\ \frac{\tilde{s}(i)}{\tilde{s}(l)}\tilde{s}_t(l) & i \geq l \end{cases} \tag{A.6}$$

Log-linearizing (A.6) we obtain that

$$\hat{\tilde{s}}_t(i) = \hat{\tilde{s}}_t(l) - \psi\hat{q}_t\mathbb{I}_{i<l_t} - \psi\hat{q}_{t+\tau(i)}\mathbb{I}_{i\in(l_t,l)} \tag{A.7}$$

Replacing in the expression inside the parenthesis in (A.4), we obtain

$$\left(\int_l^1 \hat{\tilde{s}}_t(i)\frac{\alpha(i)\sigma(i)s(i)}{\int_l^1 \alpha(i)\sigma(i)s(i)di}di - \int_0^l \hat{\tilde{s}}_t(i)\frac{\alpha(i)s(i)}{\int_0^l \alpha(i)s(i)di}di\right) =$$
$$\int_0^l \psi\left(\hat{q}_t \mathbb{I}_{i<l_t} + \hat{q}_{t+\tau(i)}\mathbb{I}_{i>l_t}\right)\frac{\alpha(i)s(i)}{\int_0^l \alpha(x)s(x)dx}di =$$
$$\psi\hat{q}_t - \psi\int_{l_t}^l \left(\hat{q}_t - \hat{q}_{t+\tau(i)}\right)\frac{\alpha(i)s(i)}{\int_0^l \alpha(x)s(x)dx}di$$

where the last line uses (A.3) annd (A.2).

Then, given our guess that l_t increases monotonically along the equilibrium path, from (12) we see that ω_t decreases monotonically along the equilibrium path. This implies that $\hat{q}_t > \hat{q}_{t+\tau(i)} > 0$ for all i and all t. So, we can show that the term inside the integral is of second order:

$$0 \leq \int_{l_t}^l \left(\hat{q}_t - \hat{q}_{t+\tau(i)}\right)\frac{\alpha(i)s(i)}{\int_0^l \alpha(x)s(x)dx}di \leq \int_{l_t}^l \hat{q}_t\frac{\alpha(i)s(i)}{\int_0^l \alpha(x)s(x)dx}di \leq \frac{\max_{i\in(l_t,l)}\alpha(i)s(i)l}{\int_0^l \alpha(x)s(x)dx}\hat{l}_t\hat{q}_t \approx 0.$$

Replacing this expression back in (A.4), we obtain the Kolmogorov-Forward equation for $\hat{y}_t$ shown in the lemma,

$$\frac{\partial \hat{y}_t}{\partial t} = -\delta\hat{y}_t + \frac{\theta-1}{\kappa\eta+\theta}\delta\psi\hat{q}_t. \tag{A.8}$$

To show the Kolmogorov-Backward equation satisfied by $\hat{q}_t$, we differentiate with respect to time and obtain

$$\frac{\partial q_t}{\partial t} = -\omega_t + (\rho+\delta)q_t.$$

Then, using that $p_t = (y_t)^{\frac{1}{1-\theta}}$ and log-linearizing, we obtain the equation shown in the lemma,

$$\frac{\partial \hat{q}_t}{\partial t} = \frac{1}{\theta-1}\hat{y}_t + (\rho+\delta)\hat{q}_t. \tag{A.9}$$

To complete the proof, we need to derive the policy functions, show the equilibrium is saddle-path stable, and verify that l_t increases monotonically along the equilibrium path (which we guessed in order to derive the 2X2 dynamic system in $\hat{q}_t, \hat{y}_t$).

Let us guess that the policy functions are given by $\frac{\partial \hat{y}_t}{\partial t} = -\lambda\hat{y}_t$ and $\hat{q}_t = \zeta\hat{y}_t$. Replacing in the 2X2 dynamic system, we obtain the expressions in the proposition for λ and ζ:

$$-\lambda = -\delta + \frac{\theta-1}{\kappa\eta+\theta}\delta\psi\zeta$$
$$-\zeta\lambda = \frac{1}{\theta-1} + (\rho+\delta)\zeta$$

Notice that the second equation immediately yield the expression for ζ. To get the ex-

pression for λ, notice that substituting the expression for ζ into the first equation implies that

$$(\delta - \lambda)(\rho + \delta + \lambda) + \frac{\psi\delta}{\kappa\eta + \theta} = 0$$

Then, the solutions for λ is

$$\lambda_{12} = -\frac{\rho}{2} \pm \sqrt{\left(\frac{\rho}{2}\right)^2 + \delta\left((\rho + \delta) + \frac{\psi}{\kappa\eta + \theta}\right)}$$

Because the term inside the square root is always positive, two solutions always exist. Furthermore, one of the solutions is negative and the other one is positive. This implies that the equilibrium is saddle-path stable. Furthermore, the positive solution is the speed of convergence of equilibrium variables.

Finally, the equilibrium threshold is $\hat{l}_t = \hat{l}_0 e^{-\lambda t}$. Then, if $\hat{l}_0 < 0$, this implies that l_t increases monotonically along the equilibrium path, which verifies our initial guess and completes the proof of the proposition.

A.5 Proof of Proposition 2

Using the definitions y_t and q_t together with Proposition 1, we have

$$\begin{aligned}
\Delta log(y_t) &= (\theta - 1)\left(\Delta log(A) - \Delta log(\omega) - \hat{\omega}_t\right) \\
&= (\theta - 1)\left(\Delta log(A) - \left(\Delta log(\omega) + \hat{\omega}_0 e^{-\lambda t}\right)\right) \\
\Delta log(q_t) &= \Delta log(q) + \hat{q}_t \\
&= \frac{1}{\rho + \delta}\Delta log(\omega) + \frac{1}{\rho + \delta + \lambda}\hat{\omega}_0 e^{-\lambda t}
\end{aligned}$$

We next derive the long-run change $\Delta log(\omega)$ and the short-to-long-run change $\hat{\omega}_0$

Long-run. In this case the skill distribution is given by (16), so that the equilibrium threshold solves

$$A^{\theta-1}\sigma(l)^\theta \int_0^l \alpha(i)(\alpha(i))^{\frac{\psi}{\rho+\delta}} di = \int_l^1 \alpha(i)\sigma(i)\left(\alpha(i)\frac{\sigma(i)}{\sigma(l)}\right)^{\frac{\psi}{\rho+\delta}} di$$

Consider a log-linear approximation around the final stationary equilibrium:

$$(\theta - 1)\Delta log(A) + \left(\left(\theta + \frac{\psi}{\rho + \delta}\right)\frac{1}{\eta} + \kappa\right)\Delta log(l) = 0$$

Thus,

$$\Delta log(l) = -\frac{\eta}{\left(\theta + \frac{\psi}{\rho+\delta}\right) + \eta\kappa}(\theta - 1)\,\Delta log(A)$$

From equation (12), $\Delta log(\omega) = -\frac{1}{\eta}\Delta log(l)$ and, therefore,

$$\Delta log(\omega) = \frac{1}{\left(\theta + \frac{\psi}{\rho+\delta}\right) + \eta\kappa} (\theta - 1)\, \Delta log(A) \tag{A.10}$$

Short-to-Long We start by deriving the change in the skill distribution using (16): $\hat{s}_0(i) = \hat{s}_0(l)$ if $i < l$ and $\hat{s}_0(i) = \hat{s}_0(l) - \frac{\psi}{\rho+\delta}\Delta log(\omega)$ if $i > l$. Along the transition, the change in the assignment threshold is determined by (14) given the change in the skill distribution:

$$\left(\frac{\theta}{\eta} + \kappa\right) \hat{l}_0 = -\frac{\psi}{\rho+\delta}\Delta log(\omega)$$

Then,

$$\hat{\omega}_0 = \frac{1}{\theta + \kappa\eta} \frac{\psi}{\rho+\delta}\Delta log(\omega)$$

Dynamic responses We now use the derivations above to show that

$$\Delta log(y_t) = \frac{1}{\theta + \kappa\eta} \left((1 + \kappa\eta) + \frac{(\theta - 1)}{\theta + \kappa\eta + \frac{\psi}{\rho+\delta}} \frac{\psi}{\rho+\delta} (1 - e^{-\lambda t}) \right) (\theta - 1) dlog(A)$$

$$\Delta log(q_t) = \frac{1}{\theta + \kappa\eta + \frac{\psi}{\rho+\delta}} \frac{1}{\rho+\delta} \left(1 + \frac{\lambda - \delta}{\delta} e^{-\lambda t} \right) (\theta - 1)\, \Delta log(A)$$

where the last line uses the solution to λ from Proposition 1.

A.6 Proof of Proposition 3 and Proposition 4

1. Long-run adjustment

$$\Delta log(y_\infty) = \left(1 - \frac{\theta - 1}{\theta + \kappa\eta + \frac{\psi}{\rho+\delta}} \right) (\theta - 1) \Delta log(A)$$

$$\Delta log\,(q_\infty) = \frac{1}{\rho+\delta} \left(\frac{1}{1-\theta} \Delta log(y_\infty) + \Delta log(A) \right)$$

Then, it is straightforward to see that $\Delta log(y_\infty)$ is increasing in both η and ψ, while the opposite holds for $\Delta log(q_\infty)$.

2. Short-run adjustment

$$\begin{aligned}\Delta log(y_0) &= \left(1 - \frac{\theta - 1}{\theta + \kappa\eta}\right)(\theta - 1)\Delta log(A) \\ \Delta log\left(q_0\right) &= \frac{1}{\theta + \kappa\eta + \frac{\psi}{\rho+\delta}} \frac{\lambda}{\delta} \frac{1}{\rho + \delta}(\theta - 1)\Delta log(A) \\ &= \frac{1}{\theta + \kappa\eta} \frac{1}{\rho + \lambda}(\theta - 1)\Delta log(A)\end{aligned}$$

The first line shows that $\Delta log(y_0)$ is increasing in η and is independent of ψ. Since λ is decreasing in η, the second line shows that $\Delta log(q_0)$ is decreasing in η. Since λ is increasing in ψ, the third line shows that $\Delta log(q_0)$ is decreasing in ψ.

3. Persistence

$$\begin{aligned}\int_0^\infty |\hat{y}_t| dt &= -\frac{1}{\lambda}\hat{y}_0 = \frac{1}{\lambda} \frac{\frac{\psi}{\rho+\delta}}{\theta + \kappa\eta + \frac{\psi}{\rho+\delta}} \frac{\theta - 1}{\theta + \kappa\eta} \left(\theta - 1\right) \Delta log(A) \\ \int_0^\infty \hat{q}_t dt &= \frac{1}{\lambda}\hat{q}_0 = \frac{1}{\theta + \eta\kappa + \frac{\psi}{\rho+\delta}} \frac{\lambda - \delta}{\lambda} \frac{1}{\delta} \frac{1}{\rho + \delta} \left(\theta - 1\right) \Delta log(A)\end{aligned}$$

The second line shows that $\int_0^\infty \hat{q}_t dt$ is increasing in ψ around $\psi = 0$ (since then $\lambda = \delta$) and is decreasing in η because λ is decreasing in η and increasing in ψ.

The first line shows that $\int_0^\infty |\hat{y}_t| dt$ is increasing in ψ around $\psi = 0$ since $\frac{\partial\left(\frac{1}{\lambda}\frac{\frac{1}{\rho+\delta}}{\theta+\kappa\eta+\frac{\psi}{\rho+\delta}}\right)}{\partial\psi}$ is bounded. To show that it is decreasing in η, we show that:

$$\begin{aligned}\frac{\partial log(\frac{1}{\lambda}\frac{\frac{\psi}{\rho+\delta}}{\theta+\kappa\eta+\frac{\psi}{\rho+\delta}}\frac{\theta-1}{\theta+\kappa\eta})}{\partial\eta} &= \frac{1}{\lambda}\frac{1}{\rho+2\lambda}\frac{\psi\delta\kappa}{(\theta+\kappa\eta)^2} - \frac{\kappa}{\theta+\kappa\eta+\frac{\psi}{\rho+\delta}} - \frac{\kappa}{\theta+\kappa\eta} \\ &= -\left(\left(1 - \underbrace{\frac{\lambda-\delta}{\lambda}\frac{\rho+\delta+\lambda}{\rho+2\lambda}}_{<1 \text{ because } \lambda>\delta}\right)\frac{1}{(\theta+\kappa\eta)} + \frac{1}{\theta+\kappa\eta+\frac{\psi}{\rho+\delta}}\right)\kappa < 0\end{aligned}$$

A.7 Proof of Proposition 5

We have that, because of the envelope theorem,

$$\begin{aligned}U_\tau &= \int \tilde{s}_\tau(i) V_t(i) di - \frac{1}{\psi}\int \tilde{s}_\tau(i) log\left(\frac{\tilde{s}_\tau(i)}{\bar{s}(i)}\right) di \\ &\approx \int s(i)(V_\tau(i) - V(i)) di + U_\infty\end{aligned}$$

Then,

$$\begin{aligned}
U_\tau - U_\infty &= \int_\tau^\infty e^{-(\rho+\delta)(t-\tau)} \int s(i) log\left(\frac{\alpha(i)\max(\omega_t\sigma(i),1)}{P_t}\right) didt \\
&- \int_0^\infty e^{-(\rho+\delta)t} \int s(i) log\left(\frac{\alpha(i)\max(\omega\sigma(i),1)}{P}\right) didt \\
&\approx \int_l^1 s(i)di \left(\int_\tau^\infty e^{-(\rho+\delta+\lambda)(t-\tau)}\hat{\omega}_\tau dt\right) - \left(\int_\tau^\infty e^{-(\rho+\delta+\lambda)(t-\tau)}\hat{P}_\tau dt\right) \\
&= -\left(\frac{y_\infty}{1+y_\infty}\frac{1}{1-\theta}\hat{y}_0 - \int_l^1 s(i)di\hat{\omega}_0\right)\frac{1}{\rho+\delta+\lambda}e^{-\lambda\tau} \\
&= -\left(\frac{y_\infty}{1+y_\infty} - \int_l^1 s(i)di\right)\hat{q}_\tau
\end{aligned}$$

Also,

$$\begin{aligned}
U_\infty - U_{0^-} &\approx \left(\int_l^1 s(i)di\right)\frac{1}{\rho+\delta}\Delta log\left(\omega_\infty\right) + \frac{y_\infty}{1+y_\infty}\frac{1}{\theta-1}\frac{1}{\rho+\delta}\Delta log(y_\infty) \\
&= \left(\int_l^1 s(i)di\right)\frac{1}{\rho+\delta}\Delta log\left(\omega_\infty\right) + \frac{y_\infty}{1+y_\infty}\frac{1}{\rho+\delta}\left(\Delta log(A) - \Delta log(\omega_\infty)\right) \\
&= \frac{y_\infty}{1+y_\infty}\frac{1}{\rho+\delta}\Delta log(A) - \left(\frac{y_\infty}{1+y_\infty} - \int_l^1 s(i)di\right)\Delta log\left(q_\infty\right)
\end{aligned}$$

Then,

$$\begin{aligned}
\Delta\bar{U} &= U_\infty - U_{0^-} + r\int_0^\infty e^{-rt}(U_\tau - U_\infty)d\tau \\
&\approx U_\infty - U_{0^-} - \left(\frac{y_\infty}{1+y_\infty} - \int_l^1 s(i)di\right) r\int_0^\infty e^{-rt}\hat{q}_\tau \\
&= \frac{y_\infty}{1+y_\infty}\frac{1}{\rho+\delta}\Delta log(A) - \left(\frac{y_\infty}{1+y_\infty} - \int_l^1 s(i)di\right)\Delta\bar{\Omega}
\end{aligned}$$

Finally, using Proposition 2,

$$\begin{aligned}
\Delta\bar{\Omega} &= r\int_0^\infty e^{-rt}\Delta log(q_\tau)d\tau \\
&= \Delta log(q_\infty) + r\int_0^\infty e^{-rt}\hat{q}_\tau d\tau \\
&\approx \Delta log(q_\infty) + \frac{r}{r+\lambda}\hat{q}_0 \\
&\approx \Delta log(q_\infty) + \frac{r\lambda}{r+\lambda}\int_0^\infty \hat{q}_\tau d\tau
\end{aligned}$$

A.8 Proof of Proposition 6

We start by pointing out that, by the definition in Proposition 1, $\lambda^1 = \lambda^2 = \delta$ because $\psi^1 = \psi^2 = 0$. Thus, Proposition 2 immediately implies that both parametrizations must

satisfy the condition that

$$\Delta \log(q_T) = \frac{\theta - 1}{(\theta + \kappa\eta^P)(\rho + \delta)} \Delta \log A^P \tag{A.11}$$

where $P = 1$ for the first approach or $P = 2$ for the second approach.

Notice also that the combination of Propositions 1 and 5 implies that

$$\Delta\bar{\Omega} = \Delta \log(q_T) - \hat{q}_0 \left(e^{-\lambda T} - 1 + \frac{\lambda}{r + \lambda} \right)$$

and, therefore,

$$\Delta\bar{\Omega} = \Delta \log(q_T) \left(\frac{1 + \frac{\lambda - \delta}{\delta} \frac{r}{r+\lambda}}{1 + \frac{\lambda - \delta}{\delta} e^{-\lambda T}} \right). \tag{A.12}$$

This expression implies that, because $\lambda^1 = \lambda^2 = \delta$, both parametrizations entail

$$\Delta\bar{\Omega}^P = \Delta \log(q_T) \tag{A.13}$$

for $P = 1, 2$.

We now use these expressions two establish the two parts of the proposition.

Part 1. In the first approach, we set $\Delta \log A^1 = \log A$. So, by equation (A.11), we must set

$$\kappa\eta^1 = \frac{\theta - 1}{\rho + \delta} \frac{\Delta \log(A)}{\Delta \log(q_T)} - \theta,$$

which is positive as long as $\frac{\Delta \log(A)}{\Delta \log(q_T)} > \frac{\theta(\rho+\delta)}{\theta-1}$.

By taking the ratio between the expressions in (A.12) and (A.13),

$$\frac{\Delta\bar{\Omega}^1}{\Delta\bar{\Omega}} > 1 \iff e^{-\lambda T} > \frac{r}{r + \lambda} \iff T < T^1 \equiv \frac{1}{\lambda} \log \left(\frac{r + \lambda}{r} \right).$$

The expression of $\Delta\bar{U}$ in Proposition 5 immediately implies that $\Delta\bar{\Omega}^1 > \Delta\bar{\Omega} \iff \Delta\bar{U}^1 < \Delta\bar{U}$ whenever $y_\infty > e_\infty$.

Part 2. In the second approach, we set $\eta^2 = \eta$. So, by equation (A.11), we must set .

$$\Delta \log A^2 = \Delta \log(q_T) \frac{(\theta + \kappa\eta)(\rho + \delta)}{\theta - 1}.$$

Expressions in (A.12) and (A.13) also hold in this case, so the same steps used above guarantee that $\Delta\bar{\Omega}^2 > \Delta\bar{\Omega}$ if, and only if, $T < T^1$. To establish the result, it is sufficient to show that $\Delta \log A^2 \leq \Delta \log A$ because, by Proposition 5, $\Delta\bar{\Omega}^2 > \Delta\bar{\Omega}$ and $\Delta \log A^2 \leq \Delta \log A$ imply that $\Delta\bar{U}^2 < \Delta\bar{U}$.

We now show that $\Delta \log A^2 \leq \Delta \log A$. By combining Proposition 2 and equation (A.11), we have that

$$\Delta \log A^2 = \frac{(\theta + \kappa\eta)}{\left(\theta + \kappa\eta + \frac{\psi}{\rho+\delta}\right)} \left(1 + \frac{\lambda - \delta}{\delta} e^{-\lambda T} \right) \Delta \log A$$

and, therefore,

$$\Delta \log A^2 \leq \frac{(\theta + \kappa\eta)}{\left(\theta + \kappa\eta + \frac{\psi}{\rho+\delta}\right)} \frac{\lambda}{\delta} \Delta \log A.$$

So, $\Delta \log A^2 \leq \Delta \log A$ if

$$F(\psi) \equiv \frac{(\theta + \kappa\eta)}{\left(\theta + \kappa\eta + \frac{\psi}{\rho+\delta}\right)} \frac{\lambda(\psi)}{\delta} \leq 1$$

with $\lambda(\psi)$ defined in Proposition 1.

This condition always holds because $\lambda(0) = \delta$, $F(0) = 1$ and $sign\left(\frac{\partial F(\psi)}{\partial \psi}\right) < 0$. To see this, we use the expression for $\lambda(\psi)$ in Proposition 1 to show that

$$\begin{aligned}
sign\left(\frac{\partial F(\psi)}{\partial \psi}\right) =& sign\left(\frac{\partial \lambda(\psi)}{\partial \psi}\left(\theta + \kappa\eta + \frac{\psi}{\rho+\delta}\right) - \frac{\lambda}{\rho+\delta}\right) \\
=& sign\left(\frac{1}{2\lambda+\rho}\frac{\delta}{\theta+\kappa\eta}\left(\theta + \kappa\eta + \frac{\psi}{\rho+\delta}\right) - \frac{\lambda}{\rho+\delta}\right) \\
=& sign\left(\frac{1}{2\lambda+\rho}\left(\delta + \frac{1}{\rho+\delta}\frac{\psi}{\theta+\kappa\eta}\right) - \frac{\lambda}{\rho+\delta}\right) \\
=& sign\left(\frac{1}{2\lambda+\rho}\left(\delta + \frac{1}{\rho+\delta}\left[\left(\lambda+\frac{\rho}{2}\right)^2 - \left(\frac{\rho}{2}\right)^2 - \delta(\rho+\delta)\right]\right) - \frac{\lambda}{\rho+\delta}\right) \\
=& sign\left(\frac{1}{2\lambda+\rho}\left(\frac{1}{\rho+\delta}\left[\left(\lambda+\frac{\rho}{2}\right)^2 - \left(\frac{\rho}{2}\right)^2\right]\right) - \frac{\lambda}{\rho+\delta}\right) \\
=& sign\left(\frac{1}{2\lambda+\rho}\left[\left(\lambda+\frac{\rho}{2}\right)^2 - \left(\frac{\rho}{2}\right)^2\right] - \lambda\right) \\
=& sign\left(\frac{\lambda+\rho}{2\lambda+\rho} - 1\right).
\end{aligned}$$

A.9 Proof of equations (23)–(24)

Proof of equation (23). We first use a first-order approximation to write the log-change in relative high-tech employment in terms of changes in the high-tech employment share:

$$\Delta \log\left(e_t^{old}\right) = \log\left(\frac{e_t^{old}}{e_{0^-}^{old}}\right) \approx \frac{1}{(1-e_{H,\infty})e_{H,\infty}}\left(e_{H,t}^{old} - e_{H,0^-}^{old}\right)$$

where $e_{H,t}^{old} = \int_{l_t}^{1} s_0(i)di$.

Since $\Delta\left(\frac{1}{(1-e_{H,\infty})e_{H,\infty}}\right)\left(e_{H,t}^{old} - e_{H,0^-}^{old}\right)$ is a second order term, we get the approximation:

$$\Delta \log\left(e_t^{old}\right) \approx \frac{1}{(1-e_{H,0^-})e_{H,0^-}}\left(e_{H,t}^{old} - e_{H,0^-}^{old}\right)$$

Notice that

$$e^{old}_{H,t} - e^{old}_{H,0^-} = \int_l^{l_{0^-}} s_0(i)di + \int_{l_t}^{l} s_0(i)di$$

By approximating these expressions around l,

$$\begin{aligned} e^{old}_{H,t} - e^{old}_{H,0^-} &\approx s_0(l)l\left(\Delta\log(l) - \hat{l}_t\right) \\ &\approx (s_0(l)l)\,\eta\Delta\log(\omega_t) \\ &\approx (s_0(l_{0^-})l_{0^-})\,\eta\Delta\log(\omega_t) \\ &\approx (1 - e_{H,0^-})\eta\Delta\log(\omega_t) \end{aligned}$$

where the third equality follows from the fact that $\Delta\left(s_0(l)l\right)\Delta\log(\omega_t)$ is a second order term, and the last equality follows from normalizing the initial skill distribution to be uniform (which implies $s_0(l_{0^-})l_{0^-} = 1 - e_{H,0^-}$).

Combining the two expressions,

$$\Delta\log\left(e^{old}_t\right) \approx \frac{1}{e_{H,0^-}}\eta\Delta\log(\omega_t)$$

Using the demand expression in (2),

$$\Delta\log\left(e^{old}_t\right) \approx \frac{1}{e_{H,0^-}}\eta\left(-\frac{1}{\theta-1}\log y_t + \Delta\log A\right)$$

Using the expression for the evolution of y_t in Proposition 2,

$$\Delta\log\left(e^{old}_t\right) \approx \frac{1}{e_{H,0^-}}\frac{\eta}{\theta+\kappa\eta}\left(-1-\kappa\eta - \frac{\psi}{\chi}(\theta-1)(1-e^{-\lambda t}) + (\theta+\kappa\eta)\right)\Delta\log A$$

$$\Delta\log\left(e^{old}_t\right) \approx \frac{1}{e_{H,0^-}}\frac{\eta}{\theta+\kappa\eta}\left(1 - \frac{\psi}{\chi}(1-e^{-\lambda t})\right)(\theta-1)\Delta\log A,$$

which is identical to (23).

Proof of equation (24). We first use a first-order approximation to write the log-change in relative high-tech employment in terms of changes in the high-tech employment share:

$$\begin{aligned} \log(\frac{e^{young}_t}{e^{young}_{0^-}}) - \log(\frac{e^{old}_t}{e^{old}_{0^-}}) \approx& \frac{1}{1-e_{H,\infty}}\left(\frac{e^{young}_{H,t} - e^{young}_{H,0^-}}{e_{H,\infty}} - \frac{e^{old}_{H,t} - e^{old}_{H,0^-}}{e_{H,\infty}}\right) \\ =& \frac{1}{(1-e_{H,\infty})e_{H,\infty}}\left(\left(e^{young}_{H,t} - e^{old}_{H,t}\right) - \left(e^{young}_{H,0^-} - e^{old}_{H,0^-}\right)\right) \\ =& \approx \frac{1}{(1-e_{H,\infty})e_{H,\infty}}\left(e^{young}_{H,t} - e^{old}_{H,t}\right) \end{aligned}$$

where the last equality follows from the fact that before the shock old and young make identical choices, $e^{young}_{H,0^-} = e^{old}_{H,0^-}$.

Using the definition of employment shares for each generation,

$$e_{H,t}^{young} - e_{H,t}^{old} \approx \frac{1}{1-(1-\tilde{x}_0)e^{-\delta t}} \left(\tilde{x}_0 e^{-\delta t} \int_{l_t}^{1} s_0(i)di + \delta \int_0^t e^{\delta(\tau-t)} \int_{l_t}^{1} \tilde{s}_\tau(i) di d\tau \right) - \int_{l_t}^{1} s_0(i)di$$
$$\approx \frac{1}{1-(1-\tilde{x}_0)e^{-\delta t}} \left(\delta \int_0^t e^{\delta(\tau-t)} \int_{l_t}^{1} \left(\tilde{s}_\tau(i) - s_0(i)\right) di d\tau \right)$$

Thus,

$$\log\left(\frac{e_t^{young}}{e_{0^-}^{young}}\right) - \log\left(\frac{e_t^{old}}{e_{0^-}^{old}}\right) \approx \frac{1}{(1-e_{H,\infty})e_{H,\infty}} \frac{1}{1-(1-\tilde{x}_0)e^{-\delta t}} \left(\delta \int_0^t e^{\delta(\tau-t)} \int_{l_t}^{1} \left(\tilde{s}_\tau(i) - s_0(i)\right) di d\tau \right) \tag{A.14}$$

We now consider the following approximation:

$$\int_{l_t}^{1} \tilde{(}\tilde{s}_\tau(i) - s_0(i))di \approx \int_{l}^{1} s(i)(\hat{\tilde{s}}_\tau(i) - \hat{s}_0(i))di$$

Then, we derive $\hat{s}_0(i)$ using the expression for the stationary skill distribution

$$s_0(i) = \frac{\bar{s}(i)\alpha(i)^{\frac{\psi}{\rho+\delta}} \left(\omega_{0^-}\sigma(i)\right)^{\frac{\psi}{\rho+\delta}\mathbb{I}_{i>l_{0^-}}}}{\int_0^{l_{0^-}} \bar{s}(j)\alpha(j)^{\frac{\psi}{\rho+\delta}} dj + \int_{l_{0^-}}^{1} \bar{s}(j)\alpha(j)^{\frac{\psi}{\rho+\delta}} \left(\omega_{0^-}\sigma(j)\right)^{\frac{\psi}{\rho+\delta}} dj}$$
$$\Longrightarrow$$
$$\hat{s}_0(i) \approx -\left(\mathbb{I}_{i>l} - \int_l^1 s(j)dj \right) \frac{\psi}{\rho+\delta} \Delta \log(\omega)$$

Using Corollary 4,

$$\int_{l_t}^{1} (\tilde{s}_\tau(i) - s_0(i))di \approx e_{H,\infty}\left(1-e_{H,\infty}\right) \left(\psi\hat{q}_\tau + \frac{\psi}{\rho+\delta}\Delta\log(\omega) \right)$$
$$= e_{H,\infty}\left(1-e_{H,\infty}\right)\psi\left(\hat{q}_\tau + \Delta\log(q)\right)$$

We now apply this expression into (A.14):

$$\log\left(\frac{e_t^{young}}{e_{0^-}^{young}}\right) - \log\left(\frac{e_t^{old}}{e_{0^-}^{old}}\right) \approx \frac{\psi}{1-(1-\tilde{x}_0)e^{-\delta t}} \left(\delta \int_0^t e^{\delta(\tau-t)} \left(\hat{q}_\tau + \Delta\log(q)\right) d\tau \right)$$
$$\approx \frac{\psi}{1-(1-\tilde{x}_0)e^{-\delta t}} \left(\delta \int_0^t e^{\delta(\tau-t)} \hat{q}_0 e^{-\lambda\tau} d\tau + (1-e^{-\delta t})\Delta\log(q) \right)$$
$$\approx \frac{\psi}{1-(1-\tilde{x}_0)e^{-\delta t}} \left(\frac{\delta}{\lambda-\delta}(e^{-\delta t} - e^{-\lambda t})\hat{q}_0 + (1-e^{-\delta t})\Delta\log(q) \right)$$

Notice that Proposition 2 implies that

$$\Delta\log(q) = \frac{1}{\chi}\left(\theta-1\right)\Delta\log A$$

$$\Delta \log(q_0) = \frac{1}{\chi}\left(1 + \frac{\lambda - \delta}{\delta}\right)(\theta - 1)\,\Delta \log A$$

$$\hat{q}_0 = \Delta \log(q_0) - \Delta \log(q) = \frac{1}{\chi}\frac{\lambda - \delta}{\delta}(\theta - 1)\,\Delta \log A$$

Thus,

$$\begin{aligned}\log(\frac{e_t^{young}}{e_{0^-}^{young}}) - \log(\frac{e_t^{old}}{e_{0^-}^{old}}) &\approx \frac{\psi}{\chi}\frac{1}{1-(1-\tilde{x}_0)e^{-\delta t}}\left((e^{-\delta t} - e^{-\lambda t}) + (1 - e^{-\delta t})\right)(\theta - 1)\log A \\ &\approx \frac{\psi}{\chi}\frac{1 - e^{-\lambda t}}{1-(1-\tilde{x}_0)e^{-\delta t}}(\theta - 1)\Delta \log A,\end{aligned}$$

which is equivalent to (24).

Appendix B Empirical Application

B.1 Cognitive Intensity and Use of New Technologies Across Occupations

This section analyzes the types of tasks required by cognitive-intensive occupations. Figure B1 reports the correlation between the occupation's intensity in cognitive skills and the share of individuals in that occupation reporting they intensely perform each of the listed tasks. The top tasks performed in cognitive-intensive occupations are directly related to technological innovations recently introduced in the workplace: working with internet, in particular, and with computers, more generally. On the other extreme, individuals employed in the least cognitive-intensive occupations tend to perform routine tasks associated with manufacturing and repairing. The results in Figure B1 are consistent with the evidence establishing the heterogeneous impact of new technologies on different tasks performed by workers – e.g., Autor, Levy, and Murnane (2003), Spitz-Oener (2006), Autor and Dorn (2013), and Akerman, Gaarder, and Mogstad (2015).

We then investigate whether these new technologies affected worker generations differently conditional on their occupation. We consider two generations: a young generation aged less than 40 years and an old generation aged more than 40 years.[30] Figure B2 shows that, while internet and computer usage are biased towards cognitive-intensive occupations, there were only small differences in the usage of these new technologies across worker cohorts employed in the same occupation in 2012. These results are complement the finding in Spitz-Oener (2006) that there were small between-cohort differences in the change of the task content of German occupations in the 1990s.

B.2 Sample statistics

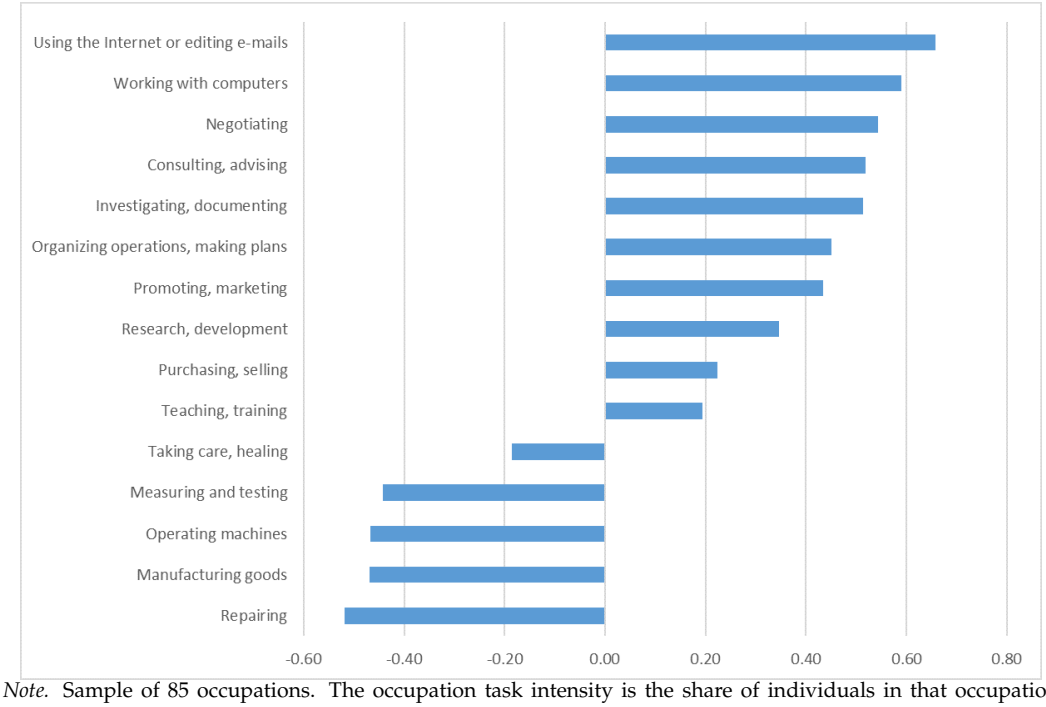
This section reports the summary statistics of our baseline sample. We begin with illustrating the increase in inequality, measured by the standard deviation of log wages, in our sample. Figure B3 compares the overall change in inequality together with the between district-generation-occupation component, measured using the residual log-wage dispersion from a mincer regression including dummies for the district-generation-occupation estimated on the sample in each year. Between 1997-2012, overall inequality in our sample increased by about 8.5 log points. As the figure illustrates, the between district-generation-occupation component explains about half of the increase in inequality during this period. In results available on request, we attest that, separately, these characteristics do not account for the inequality rise.[31]

Table B1 presents summary statistics underlying the FDZ microdata used in our empirical analysis. They illustrate the evolution of the number of employees, ages and log-wage of the baseline generations used in estimation.

[30] Results are similar if we define young generations to include workers who are less than 30, 35 or 45 years old.

[31] We also attest that the explanatory power of the between district-generation-occupation component is similar to that of the between establishment component of log-wage variance, which Card, Heining, and Kline (2013) point as the main driver of the inequality increase in Germany during this period. Notice that this is not mechanical because there are nearly 50 times as many establishments as district-occupation-generation triples in our sample.

Figure B1: Cross-occupation correlation between cognitive intensity and perfomance of different tasks



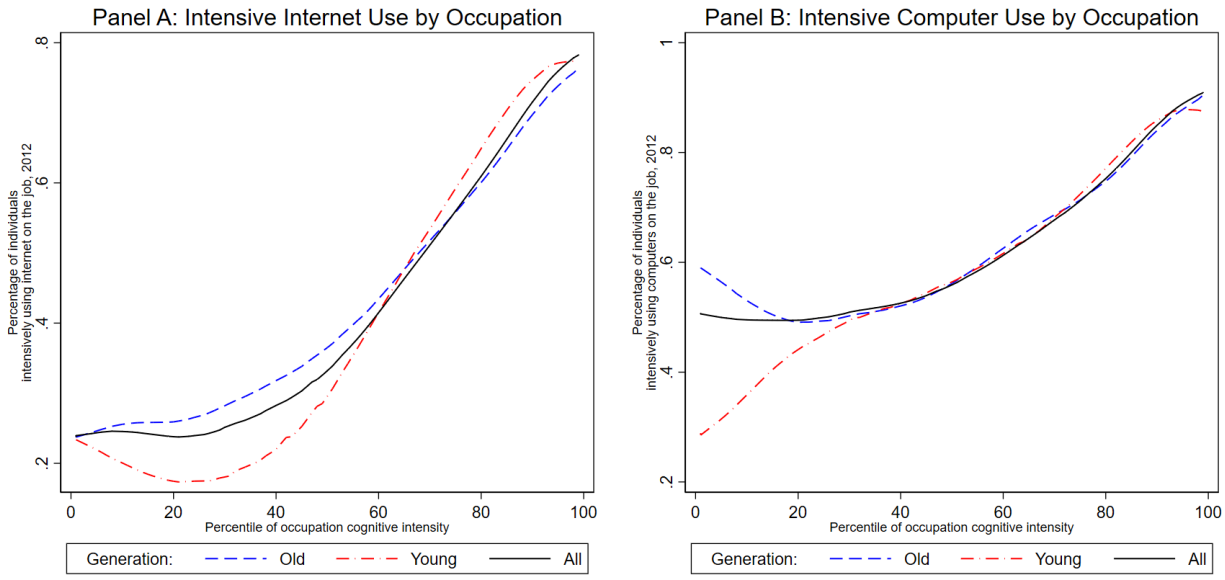
Note. Sample of 85 occupations. The occupation task intensity is the share of individuals in that occupation reporting to intensively perform the task in the 2012 Qualification and Working Conditions Survey. The occupation cognitive-skill intensity is the share of time spent on cognitive-intensive tasks in the BERUFNET dataset (2011-2013).

Table B1: Summary Statistics: German Microdata

	1995	2014
Number of observations		
Born before 1960 ("Old")	185,751	96,045
Born after 1960 ("Young")	251,451	538,590
Mean log wage		
Born before 1960 ("Old")	4.54	4.42
Born after 1960 ("Young")	4.15	4.54
Mean age		
Born before 1960 ("Old")	44.86	60.53
Born after 1960 ("Young")	28.22	39.56

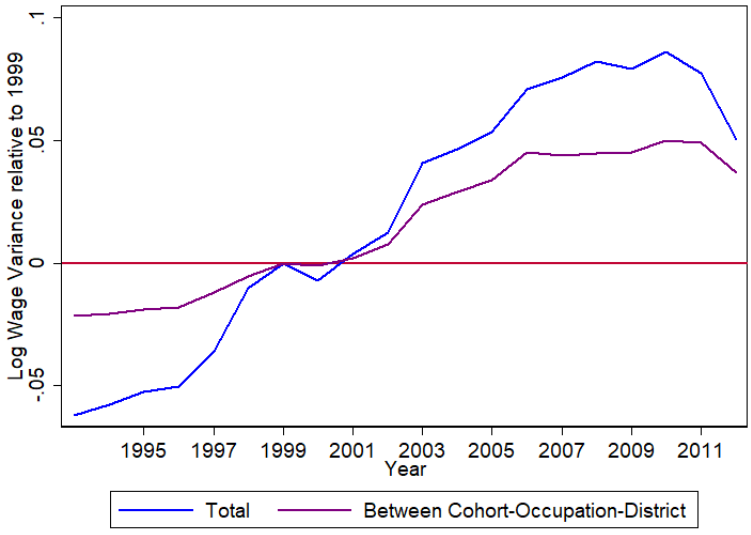
Note. Sample of male workers in LIAB data, living in West Germany, employed full-time with a positive wage in 120 occupations. Generations as defined in the table.

Figure B2: Internet and Computer Usage by Occupation: Within- and Between-Generation



Note. Sample of 85 occupations in Working Condition Survey. For each occupation, we compute the share of individuals reporting intensive internet and computer usage on their job. Young generations defined as workers aged below 40 years and Old generations defined as all workers aged above 40 years. The occupation cognitive-skill intensity is the share of time spent on cognitive-intensive tasks in the BERUFNET dataset (2011-2013). Figure reports the lowess smooth fit.

Figure B3: Aggregate Trends in Log Wage Variance



Note. Estimation of the aggregate standard deviation of log wages on the full LIAB sample and the residual dispersion in log wages from a mincer regression including district-occupation-generation dummies. Estimates are changes in dispersion relative to 1999.

B.3 Impact of New Technologies on Cognitive-intense Occupations

This section investigates the robustness of the results in Table 1 and Figures 6-7.

We first investigate the impact of cognitive-intensity on occupation employment growth with a more flexible specification that allows for different coefficients for different levels of cognitive-intensity. As is clear from Table B2, the results in Table 1 are driven largely by an increase in employment for all generations in the most cognitive intensive occupations (above the 60th percentile of cognitive intensity). This increase is substantially stronger for the young generation. Some evidence of polarization is also evident for the young generation, as they also disproportionately enter the least cognitive intensive occupations.

Appendix Table B3 investigates the robustness of the estimates of equation (25) reported in Table 1. Panel A of Table B3 reports similar results when we include occupation-level controls for import and export exposure and the growth in the fraction of migrants. Panel B shows that results are also robust to restricting the sample to native-born German males only. Panel C presents results where the "Young" generation is defined alternatively as those born after 1965 or 1955. As expected, when the definition of the young generation is further restricted to include only more recent cohorts, the coefficient on "Young" is stronger. The opposite happens if we relax the young definition to include older cohorts. Panel C also shows that results are similar if the "Young" generation is defined as those aged below 40 in each year (as in Figure 5).

Appendix Tables B4 and B5 investigate the robustness of the estimates of equations (26)-(27) reported in Figures 6-7. Table B4 investigates how our baseline set of controls affects estimates. The three panels of Table B4 present estimates for the entire post-shock period of the sample (1999-2014, Panel A), the period during which DSL was rolled out across German regions (1999-2007, Panel B), and the period before the shock (1996-1999, Panel C). Each panel includes the results of our baseline specification, as well as alternative specifications in which (i) we drop only the pre-trend control, and (ii) we augment baseline controls with district-year or district-generation-year fixed effects.

Consider first the impact of the pretrend control in the second row of each panel. This control increases the magnitude and the precision of the estimates coefficients in the period of 1999-2007 and 1999-2014. However, it has the opposite impact on the pre-shock period of 1996-1999. In this pre-shock period, there are marginally significant negative responses. Once those are taken into account, the impact of broadband internet adoption on more cognitive intensive occupations is stronger.

Turning to the specification including district-year or district-generation-year fixed effects, we can see that results are remarkably similar to our baseline estimates. This is reassuring as this specification includes a restrictive set of controls that absorb all potential confounding shocks that affect each district-generation pair in a year. For instance, they account for any pre-existing variation that might have lead to differential DSL access speeds in the district. As a result, identification in this specification comes purely through the differential effect of the cognitive intensity of the occupation for its response to the DSL access shock.

Table B5 investigates the robustness of the baseline estimates in Figures 6-7 to the sample specification. The two panels present estimates for the entire post-shock period of the sample (1999-2014, Panel A), and the period during which DSL was rolled out across German regions (1999-2007, Panel B). All specifications include the baseline set of controls.

The second row of each panel show that results are similar if we restrict the sample to only include workers born in Germany. This suggests that immigration does not seem to affect our baseline results.

We consider next several alternative definitions of the young generation based on (i) cohorts groups born after 1955, 1965 or 1970, and (ii) age groups aged below 35, 40 or 45 in each year. For all definitions, the coefficient on the cognitive intensity of the occupation for young workers is positive and strongly significant, while that for the old generation is insignificant and close to zero. As before, the coefficient in column (2) is stronger when we restrict the young generation to cohorts born in more recent years. In contrast, the coefficient is weaker when we expand the young generation to also include cohorts born in less recent years. Similar patterns arise when we define the young generation based on a lower or higher age cutoff in each year.

The last row of each panel reports estimates when we restrict the sample by excluding workers employed in establishments belonging to the top 25 percentile of establishment sizes. This exercise accounts for the likelihood that the largest establishments in Germany acquired DSL earlier through specialized private connections. In this case, we would expect adjustment in these establishments to have occurred earlier, biasing our results to zero. In line with this intuition, estimated coefficients are stronger than the baseline for all workers in column (1) and for the young-old gap in column (4). This indicates that our instrument seems to generate variation in the roll-out of broadband internet that mostly affected the occupation composition of small establishments across German districts.

Finally, Table B6 presents the Sanderson-Windmeijer F-statistics (Sanderson and Windmeijer, 2016) for the first-stage of equation (27) in the period of 1999-2014. As discussed in Section 3, this equation has multiple endogenous variables since it include DSL access interacted with occupation cognitive intensity and worker generation dummies. To test for weak instruments in this setting, we use the Sanderson-Windmeijer F-statistic, which checks whether any of our endogenous variables is weakly instrumented, as well as whether there are sufficiently many strong instruments to instrument the multiple endogenous variables. The table makes clear that we obtain uniformly high first-stage SW F-statistics in all specifications.

Table B2: Cognitive intensity and labor market outcomes across occupations in Germany: Percentiles specification

Dependent variable:	Employment Growth			Real Payroll Growth		
	All	Young	Old	All	Young	Old
Percentile of Cognitive Intensity	(1)	(2)	(3)	(4)	(5)	(6)
Low: below percentile 30	-0.012	0.286**	-0.946***	0.220*	0.590***	-0.875***
	(0.128)	(0.137)	(0.129)	(0.132)	(0.140)	(0.135)
Medium: percentiles 30-60	-0.054	-0.046	0.031	-0.086	-0.036	-0.014
	(0.194)	(0.205)	(0.208)	(0.195)	(0.208)	(0.202)
High: above percentile 60	0.812***	1.038***	0.531***	0 .816***	1.099***	0 .592***
	(0.156)	(0.166)	(0.016)	(0.158)	(0.169)	(0.157)

Note. Sample of 120 occupations. The table reports the estimate for the dependent variable over the time period 1995-2014. Occupations have been classified into 100 percentiles based on cognitive intensity, and separate coefficients estimated for percentiles below 30, 30-60 and above 60. Young generation defined as all workers born after 1960 and Old generation as all workers born before 1960. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B3: Cognitive intensity and labor market outcomes across occupations in Germany: Robustness

Dependent variable:	Employment Growth		
	All	Young	Old
	(1)	(2)	(3)
Panel A: Alternative control set, 1995-2014			
Controls for immigration and trade	1.426***	1.807***	0.841***
	(0.261)	(0.279)	(0.252)
Panel B: Alternative sample definition, 1995-2014			
Native-born Males Only	1.396***	1.807***	0.778***
	(0.226)	(0.235)	(0.231)
Panel C: Alternative generation definition, 1995-2014			
Young: Born after 1965	1.488***	2.137***	0.857***
	(0.225)	(0.299)	(0.246)
Young: Born after 1955	1.488***	1.639***	0 .967***
	(0.225)	(0.268)	(0.290)
Young: Aged Below 40 in each year	1.488***	1.748***	0.773***
	(0.225)	(0.294)	(0.246)

Note. Sample of 120 occupations, sample periods as defined in the table. Table reports the estimated coefficient on the occupation's cognitive intensity in equation (25). Each row defines a separate robustness exercise. The row "Controls for immigration and trade" includes a set of baseline controls: growth in occupational exposure to exports during the sample period, growth in occupational exposure to imports during the sample period, and growth in the fraction of immigrants in the occupation during the sample period. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B4: Impact of early DSL adoption on more cognitive-intensive occupations: Alternative control sets

Dependent variable:	Employment Growth			
	All	Young	Old	Between
Control Set	(1)	(2)	(3)	(4)
Panel A: 1999-2014				
Baseline	0.240***	0.482***	-0.065	0.546**
	(0.085)	(0.154)	(0.193)	(0.287)
No Pretrend Control	0.177**	0.292***	-0.026	0.319
	(0.087)	(0.114)	(0.189)	(0.222)
District-Year Effects	0.149**	0.475***	-0.035	0.510*
	(0.067)	(0.160)	(0.203)	(0.302)
Panel B: 1999-2007				
Baseline	0.077*	0.223***	-0.138	0.361**
	(0.043)	(0.092)	(0.116)	(0.177)
No Pretrend Control	0.015	0.137	-0.200	0.337
	(0.061)	(0.085)	(0.127)	(0.149)
District-Year Effects	0.093	0.234**	-0.134	0.368*
	(0.060)	(0.098)	(0.125)	(0.019)
Panel C: 1996-1999				
Baseline	-0.002	0.011	-0.019	0.029
	(0.026)	(0.030)	(0.031)	(0.049)
No Pretrend Control	-0.109*	-0.141*	-0.074	-0.068
	(0.065)	(0.077)	(0.084)	(0.061)
District-Year Effects	0.012	0.011	-0.022	0.034
	(0.032)	(0.032)	(0.032)	(0.050)

Note. Sample of 2 cohorts, 120 occupations and 323 districts. Sample periods as defined in the table. Column (1) reports the estimated coefficient on interaction between the occupation's cognitive intensity and district DSL access in equation (26). Columns (2)-(3) report the estimated coefficients on interaction between the occupation's cognitive intensity, generation dummies and district DSL access in equation (27). Column (4) reports the difference between the coefficients in columns (3) and (2). Generations are the baseline generations with young workers those born after 1960. All regressions are weighted by the district population size in 1999 and include a set of baseline district-level controls as well as occupation-year and generation-year fixed effects. Each row defines a separate robustness exercise. "District-Year Effects" are estimated as district-year fixed effects in column (1) and as district-year-generation fixed effects in columns (2)-(4). Standard errors clustered at the district-level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B5: Impact of early DSL adoption on more cognitive-intensive occupations: Sample selection

Dependent variable:	Employment Growth			
	All	Young	Old	Between
Sample Definition	(1)	(2)	(3)	(4)
Panel A: 1999-2014				
Baseline	0.240***	0.482***	-0.065	0.546**
	(0.085)	(0.154)	(0.193)	(0.287)
Native-born Males Only	0.223***	0.446***	0.074	0.372**
	(0.078)	(0.143)	(0.144)	(0.208)
Young: born after 1970		0.714***	-0.048	0.789**
		(0.182)	(0.242)	(0.371)
Young: born after 1965		0.612***	-0.171	0.783***
		(0.157)	(0.201)	(0.303)
Young: born after 1955		0.573***	-0.298	0.871**
		(0.196)	(0.233)	(0.355)
Young: Aged < 35 in each year		0.612***	0.059	0.553***
		(0.139)	(0.139)	(0.203)
Young: Aged < 40 in each year		0.529***	0.076	0.453**
		(0.164)	(0.163)	(0.237)
Young: Aged < 45 in each year		0.445***	0.159	0.286
		(0.170)	(0.198)	(0.266)
Small Establishments Only	0.309***	0.466***	-0.128	0.594**
	(0.083)	(0.140)	(0.183)	(0.286)
Panel B: 1999-2007				
Baseline	0.077*	0.223***	-0.138	0.361**
	(0.043)	(0.092)	(0.116)	(0.177)
Native-born Males Only	0.054	0.145*	0.037	0.108
	(0.045)	(0.092)	(0.105)	(0.171)
Young: born after 1970		0.449***	-0.070	0.518*
		(0.120)	(0.191)	(0.289)
Young: born after 1965		0.298***	-0.168	0.465**
		(0.092)	(0.118)	(0.183)
Young: born after 1955		0.203**	-0.155	0.358***
		(0.087)	(0.115)	(0.167)
Young: Aged < 35 in each year		0.118**	0.095	0.022
		(0.066)	(0.095)	(0.130)
Young: Aged < 40 in each year		0.195**	0.030	0.165
		(0.086)	(0.104)	(0.166)
Young: Aged < 45 in each year		0.206***	0.091	0.115
		(0.090)	(0.111)	(0.174)
Small Establishments Only	0.129*	0.263**	-0.170	0.434**
	(0.070)	(0.103)	(0.120)	(0.181)

Note. Sample of 2 cohorts, 120 occupations and 323 districts. Sample periods as defined in the table. Column (1) reports the estimated coefficient on interaction between the occupation's cognitive intensity and district DSL access in equation (26). Columns (2)-(3) report the estimated coefficients on interaction between the occupation's cognitive intensity, generation dummies and district DSL access in equation (27). Column (4) reports the difference between the coefficients in columns (3) and (2). All regressions are weighted by the district population size in 1999 and include a set of baseline district-level controls, pretrend controls, occupation-year and generation-year fixed effects. Each row defines a separate sample selection exercise: (i) baseline sample restricted to only Germans ("Native-born"), (ii) different definitions of young workers based on year of birth or age cutoff in each year, and (iii) baseline sample restricted to workers employed in establishments below the 75th percentile of all establishment sizes ("Small Establishments Only"). Standard errors clustered at the district-level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B6: First-stage SW F-statistics for estimation of equation (27) reported in Figure 7

Instrumented Variable	1997	2007	2014
*Panel A:*Log Change in Occupation Employment			
Young Generation*DSL Access	18.74	18.78	19.04
Old Generation* DSL Access	19.04	17.57	20.45
Young Generation*DSL Access*Cognitive Intensity	21.95	20.48	19.77
Old Generation*DSL Access*Cognitive Intensity	21.31	18.57	22.32
*Panel B:*Log Change in Occupation Payroll			
Young Generation*DSL Access	18.89	18.67	18.89
Old Generation* DSL Access	19.18	17.73	21.16
Young Generation*DSL Access*Cognitive Intensity	21.76	20.37	19.46
Old Generation*DSL Access*Cognitive Intensity	21.14	18.16	21.48

Note. Sample of 2 cohorts, 120 occupations and 323 districts. Table reports the Sanderson-Windmeijer F-statistic for each endogenous regressor when estimating equation (27).

B.4 Calibration

This appendix discusses in detail the calibration of the model. As discussed in Section 6.1, we calibrate $\theta = 3$ and $\rho = 0.02$. We also select the distribution of innate ability to normalize the initial skill distribution to be uniform: $s_0(i) \equiv 1$. In addition, we abstract from differences across skills in non-cognitive productivity: $a(i) \equiv 1$.

We now describe the procedure to calibrate all remaining parameters of the model using the estimates in Section 5. For all parameters, we assume that the shock starts with the roll-out of broadband internet in 2003. We then select parameters to match the estimates for the period of 2008 to 2014 in which we find statistically significant response in the relative payroll and relative employment of cognitive-intensive occupations.

Generation size: δ and $\tilde{x}_0$. We first set $\tilde{x}_0$ to match the 60% share of young workers in the national population in 1997. We then select δ to match the incline of 25 p.p. in the share of young workers in population between 1997 and 2014. Specifically, we select x and δ such that

$$\hat{\delta} = \frac{1}{2014 - 1997} \log(0.40/0.15)$$

$$x = -\frac{1}{\delta} \log 0.4$$

We obtain $\delta = 0.0574$. This says that the expected work life of a worker after turning 40 years is 18 further years.

Speed of Adjustment: λ. Proposition 2 implies that it is possible to write the impulse response function of relative output as

$$\Delta log(y_t) = \alpha_0 + \alpha_1 e^{-\lambda t}$$

where $\alpha_0 > 0$, $\alpha_1 < 0$, and $\lambda > 0$.

We select the parameter λ to match the growth in the estimates response of relative payroll of more cognitive-intensive occupations:

$$\hat{\lambda} = arg \min_{\lambda} \sum_{t=2008}^{2014} \left[(\hat{\beta}_t^y - \hat{\beta}_{2007}^y) - \alpha_1 e^{-\lambda(t-2007)} \right]^2 \tag{B.1}$$

where $\hat{\beta}_t^y$ is the estimated coefficient of (26) reported in Panel B of Figure 6.

The minimization problem in (B.1) yields $\hat{\lambda} = 0.135$. Figure B5 shows the fit of the calibrated model

Long-run skill supply elasticity: ψ. To calibrate ψ, we first construct the parameter

$$\hat{\alpha} = \hat{\delta} \left[\left(\frac{\rho}{2} + \hat{\lambda} \right)^2 - \left(\frac{\rho}{2} \right)^2 - \hat{\delta}(\rho + \hat{\delta}) \right]^{-1}$$

Our baseline calibration implies that $\hat{\alpha} = 3.484$.

Proposition 1 implies that

$$\kappa\eta = \psi\hat{\alpha} - \theta \tag{B.2}$$

Using expression (24), we have that

$$\Delta \log e_t^{young} - \Delta \log e_t^{old} = \frac{\psi}{\chi} \frac{1 - e^{-\lambda t}}{1 - (1 - \tilde{x}_0) e^{-\delta t}} (\theta - 1) \Delta \log A.$$

From Proposition 2,

$$(\theta - 1)\, \Delta log(A) = \Delta \log(y_t) \left(\frac{1 + \kappa\eta}{\theta + \kappa\eta} + \frac{\psi}{\chi} \frac{\theta - 1}{\theta + \kappa\eta} (1 - e^{-\lambda t}) \right)^{-1} \tag{B.3}$$

where $\chi = (\theta + \kappa\eta)(\rho + \delta) + \psi$.

Combining these two expressions, we get that

$$\frac{\Delta \log e_t^{young} - \Delta \log e_t^{old}}{\Delta \log y_t} = \frac{1 - e^{-\lambda t}}{1 - (1 - \tilde{x}_0) e^{-\delta t}} \left(\frac{1 + \kappa\eta}{\theta + \kappa\eta} \frac{\chi}{\psi} + \frac{\theta - 1}{\theta + \kappa\eta} (1 - e^{-\lambda t}) \right)^{-1}$$

Using the expression for $\kappa\eta$ in (B.2),

$$\frac{\Delta \log e_t^{young} - \Delta \log e_t^{old}}{\Delta \log y_t} = \frac{1 - e^{-\lambda t}}{1 - (1 - \tilde{x}_0) e^{-\delta t}} \left((\rho + \delta) \frac{1 + \psi\alpha - \theta}{\psi} + 1 - \frac{\theta - 1}{\psi\alpha} e^{-\lambda t} \right)^{-1}.$$

We then define the function:

$$F^{\psi}(\psi, t) \equiv \frac{1 - e^{-\hat{\lambda} t}}{1 - (1 - \tilde{x}_0) e^{-\hat{\delta} t}} \left((\rho + \hat{\delta}) \frac{1 + \psi\hat{\alpha} - \theta}{\psi} + 1 - \frac{\theta - 1}{\psi\hat{\alpha}} e^{-\hat{\lambda} t} \right)^{-1}.$$

We select the parameter ψ to match the ratio of the between-generation employment response and the payroll response:

$$\hat{\psi} = arg \min_{\psi} \sum_{t=2008}^{2014} \left[\frac{\hat{\beta}_t^{young} - \hat{\beta}_t^{old}}{\hat{\beta}_t^{y}} - F^{\psi}(\psi, t) \right]^2 \tag{B.4}$$

where $\hat{\beta}_t^y$ is the estimated coefficient of (26) reported in Panel B of Figure 6, and $\hat{\beta}_t^{young} - \hat{\beta}_t^{old}$ is the between-generation employment response obtained from the estimation of (27) reported in Panel B of Figure 7.

The minimization problem in (B.4) yields $\hat{\psi} = 0.345$. Figure B5 shows the fit of the calibrated model.

Short-run skill supply elasticity: η. The combination of (23) and (B.3) implies that

$$\frac{\Delta \log e_t^{old}}{\Delta \log y_t} \approx \frac{\eta}{e_{H,0^-}} \frac{1 - \frac{\psi}{\chi}(1 - e^{-\lambda t})}{1 + \kappa\eta + \frac{\psi}{\chi}(\theta - 1)(1 - e^{-\lambda t})}.$$

Using the expression for $\kappa\eta$ in (B.2),

$$\frac{\Delta \log e_t^{old}}{\Delta \log y_t} \approx \frac{\eta}{e_{H,0^-}} \frac{1 - \frac{(\theta - 1)(1 - e^{-\lambda t})}{\alpha(\rho + \delta) + 1}}{1 + \psi\alpha - \theta + \frac{(\theta - 1)(1 - e^{-\lambda t})}{\alpha(\rho + \delta) + 1}}.$$

We then define

$$F^{\eta}(\eta, t) \equiv \frac{\eta}{e_{H,0^-}} \frac{1 - \frac{\theta-1}{\hat{\alpha}\left(\rho+\hat{\delta}\right)+1}\left(1 - e^{-\hat{\lambda}t}\right)}{1 + \hat{\psi}\hat{\alpha} - \theta + \frac{\theta-1}{\hat{\alpha}\left(\rho+\hat{\delta}\right)+1}\left(1 - e^{-\hat{\lambda}t}\right)}.$$

where $(\hat{\delta}, \hat{\lambda}, \hat{\psi})$ are the calibrated parameters above and $e_{H,0^-}$ is the initial share of employment in cognitive-intensive occupations.

We select the parameter η to match the ratio of the employment response of old workers and the payroll response:

$$\hat{\eta} = arg \min_{\eta} \sum_{t=2008}^{2014} \left[\frac{\hat{\beta}_t^{old}}{\hat{\beta}_t^{y}} - F^{\eta}(\eta, t) \right]^2 \tag{B.5}$$

where $\hat{\beta}_t^{y}$ is the estimated coefficient of (26) reported in Panel B of Figure 6, and $\hat{\beta}_t^{old}$ is the employment response for old workers obtained from the estimation of (27) reported in Panel A of Figure 7.

The negative point estimates reported in Panel A of Figure 7 imply that the minimization problem in (B.5) yields $\hat{\eta} < 0$. Since the employment response of old generations is small and nonsignificant, we assume that they are identical to zero, which yields $\hat{\eta} = 0$. Hence, we calibrate $\eta = 0$ and evaluate the model predictions under alternative specifications of this parameter.

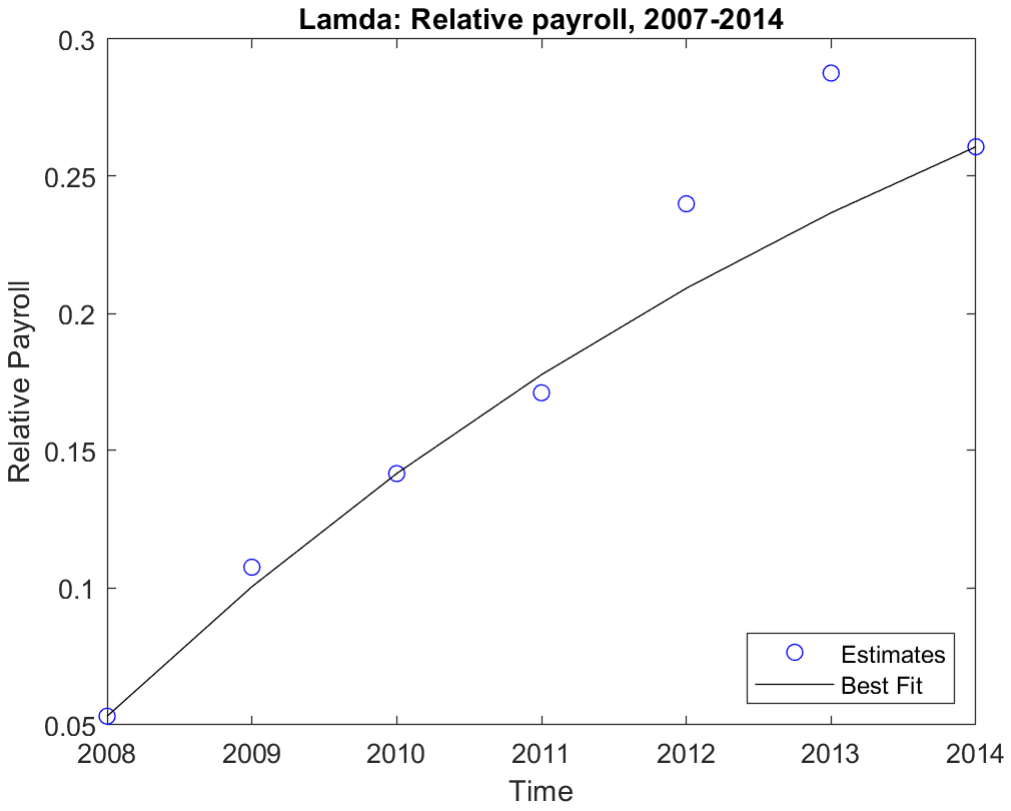


Figure B4: Calibration of λ

Note. Blue dots represent the point estimates of β_t reported in Panel B of Figure 6. Black solid curve represents the bet fit line with $\lambda = 0.135$ obtained from the solution of (B.1).

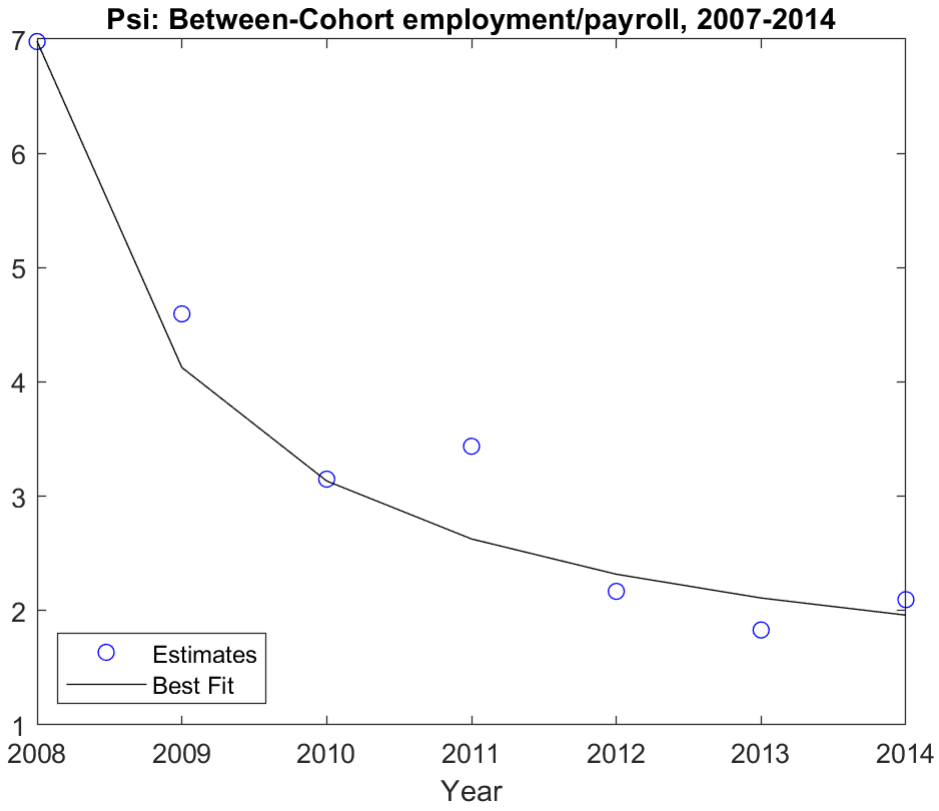


Figure B5: Calibration of ψ

Note. Blue dots represent the point estimates of $\frac{\hat{p}_t^{young} - \hat{p}_t^{old}}{\hat{p}_t^{y}}$ using the estimates reported in Panel B of Figures 6 and 7. Black solid curve corresponds to $F^{\psi}(\hat{\psi}, t)$ with $\hat{\psi} = 0.354$ obtained from the solution of (B.4).

Appendix C Additional Theoretical Results

C.1 Microfoundation of the Production Functions in (4)–(5)

Consider two firms: high-tech ($k = H$) and low-tech ($k = L$). Assume that the output of firm k at time t aggregates per-worker output $y_{kt}(i)$,

$$Y_{kt} = \int_0^1 y_{kt}(i) s_{kt}(i) di,$$

where $s_{kt}(i)$ is the quantity demanded of workers of type i at time t by firm k.

The output of workers of type i depends on their skills to perform cognitive and noncognitive tasks, $\{a_C(i), a_{NC}(i)\}$, as well as how intensely each task is used the firm's production process:

$$y_{kt}(i) = a_C(i)^{\beta_k} a_{NC}(i)^{1-\beta_k},$$

where β_k denotes the production intensity of firm k on cognitive tasks.

In our model, *technology-skill specificity* arises whenever firms are heterogeneous in terms of task intensity and workers are heterogeneous in terms of their task bundle. To see this, suppose that firm H's technology uses cognitive tasks more intensely than firm L's technology, $\beta_H > \beta_L$, and that a worker of type i is able to produce a higher cognitive-noncognitive task ratio than a worker of type j, $a_C(i)/a_{NC}(i) > a_C(j)/a_{NC}(j)$. In this case, i has a higher relative output with the cognitive-intensive technology H than j, $y_{Ht}(i)/y_{Lt}(i) > y_{Ht}(j)/y_{Lt}(j)$, and, therefore, type i is more complementary to the cognitive-intensive technology H than type j.

To map this setting to the production functions in (4)–(5), we assume that high-tech production is more intensive in cognitive tasks than low-tech production, $\beta_H > \beta_L$. We also assume that types differ in terms of their skill bundle and, without loss of generality, impose that high-i types are relatively better in performing cognitive-intensive tasks.

1. High-tech technology H uses cognitive tasks more intensely than Low-tech technology L: $\beta_H > \beta_L$.

2. Define $\sigma(i) \equiv \left(\frac{a_C(i)}{a_{NC}(i)}\right)^{\beta_H - \beta_L}$ and $\alpha(i) \equiv a_C(i)^{\beta_L} a_{NC}(i)^{1-\beta_L}$. Assume that high-$i$ types have higher cognitive-noncognitive task ratio: $\sigma(i)$ is increasing in i.

C.2 Extensions

TBA