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

October 10, 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, Northwestern, NYU, Notre-Dame, Stockholm (IIES), UCLA, UCSD, U of Munich, USC Marshall, UT Austin, WashU/St. Louis Fed, 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 cognitive-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 technology-specific wages, which determine how the skill distribution evolves over time and, ultimately, the actual equilibrium path of technology-specific wages 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 technology-specific wages (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 reminiscent of those in Lucas and Moll (2014) and Perla and Tonetti (2014) which also reduce the analysis of the equilibrium dynamics of a high-dimensional

[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.

endogenous 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 reallocate 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 and, therefore, the differences in skill heterogeneity across generations. Along the transition, q falls and relative output increases because the supply of skills expands as younger generations replace older generations.

This result shows that the impact of new technologies on the economy may significantly change over time due to the endogenous evolution of the skill distribution across generations. It provides a micro-foundation for the idea that supply elasticities tend to be lower at shorter horizons compared to longer horizons, a form of Samuelson's LeChatelier principle. However, as opposed to *naive* theories that use a reduced-form supply elasticity, such micro-foundation allows us to: (i) speak to why is it that some technological transitions are different than others, (ii) appropriately compute the welfare and distributional implications of new technologies, and (iii) connect observables to parameters governing technology-skill specificity and the skill investment cost. The rest of the paper explores each of these points.

Our third result presents comparative static exercises with respect to changes in technology-skill specificity and the cost of skill investment. These exercises show why technological transitions may look different. Therefore, they speak to past episodes (or future transitions) with differences in the nature of technologies, skills, and educational institutions. We establish 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. This implies that the more relevant margin of adjustment in economies with higher technology-skill specificity is not the reallocation of workers within a generation but the changes in the supply of skills that occur across generations. Importantly, a lower cost of skill investment (higher ψ) reinforces the impact of technology-skill specificity on the persistence of the adjustment.

Our fourth and last theoretical result analyzes the average and distributional welfare consequences of a skill-biased technological innovation. Our notion of average welfare is the average across all generations of ex-ante expected utility. Our notion of inequality is the average across all generations in lifetime welfare inequality. We establish that both measures depend not only on the long- or short-run responses of labor market outcomes but also on their persistence.

Taken together, these results indicate which economies should cause researchers to exercise more caution when extrapolating from observed changes at shorter horizons: those with a more back-loaded, persistent adjustment due to higher technology-skill specificity or lower costs of skill investment. Such extrapolations will miss most of the inequality decline and the output increase that happens at longer horizons. Therefore, they will understate the average welfare benefits and overstate the lifetime welfare inequality increases. Finally, these results also illustrate the risks of directly extrapolating from past episodes where the nature of technologies and skills involved, as well as educational institutions, may have differed.

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 cognitive-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

[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).

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.

In sum, the three pieces of evidence suggest that cognitive-skill specificity is high and that the supply of cognitive skills is elastic at longer horizons. As discussed before, 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 parameterize our model to match our estimated dynamic responses for Germany. We consider a skill-biased innovation that increases the cognitive-intensive technology's employment share from 20 percent to 50 percent across long-run equilibria. This is about the range of cognitive-intensive employment shares across developed economies. We find that the consumption equivalent average welfare increase across all generations is 46 percent and the lifetime welfare inequality increase is 39 percent. We then compare these figures to those obtained by calculations that ignore the adjustment across generations. If we assumes that changes observed at impact are permanent, we find a lower average welfare increase of 31 percent and a larger inequality increase of 76 percent. If we assume that changes observed in the long-run were permanent and happened at impact, we instead find a higher average welfare increase of 55 percent and a lower inequality increase of 30 percent. These biases are much smaller in economies with lower skill-specificity or higher cost of skill investment because the economy's adjustment is much less persistent due to the smaller differences in the skill-distribution across generations.

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 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

[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)).

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) document 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 Sections 3 and 4, we analyze the dynamic adjustment to skill-biased technological innovations. The welfare consequences implied by our theory are evaluated in Section 5. Section 6 links our theory to responses in observable outcomes for different generations of workers. It then presents our empirical analysis. Section 7 shows our quantitative analysis. Section 8 concludes.

2 A Model of Skilled-biased Technological Transitions

We consider a closed economy in continuous time. 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 inputs, and A_t is a shifter of the relative productivity of the high-tech input (as in Katz and Murphy (1992)).

Conditional on the price of intermediate inputs, the cost minimization problem of firms producing the final good 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 final output is

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

Assignment of skills to technologies. 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 their differential productivity in high-tech production. Specifically, we assume that the production function of the low-tech input 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 D.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 bundles 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

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

Equation (7) determines the labor income of workers of type i. It is the result of a maximization problem in which skill types select the technology they will work with. As discussed in detail below, the implied assignment of skill types to technology plays a central role in determining the economy's adjustment to technological shocks. Equation (7) illustrates that such an assignment depends on the endogenous price ω_t defining the relative value of one unit of effective labor employed in high-tech production, as well as the exogenous function $\sigma(i)$ defining the differential productivity of type i in high-tech production. Without loss of generality, we assume that $\sigma(i)$ is increasing; that is, we order types such that high-i types have a higher relative productivity in high-tech production. Recent papers have considered a similar structure of endogenous sorting of workers to different technologies – e.g., Acemoglu and Autor (2011), Costinot and Vogel (2010), Adão (2016).

In our theory, ω_t is a natural measure of inequality as it is the endogenous relative wage rate of skill types employed in different technologies conditional on their productivity. In what follows, we will refer to ω_t as the relative technology-specific wage or, sometimes, simply as the relative wage. However, it is important to notice movements in ω_t are not perfectly aligned with movements in the relative labor income of high-tech employees. As pointed out by Heckman and Honore (1990), the endogenous assignment problem in (7) implies that high-tech relative labor income may change due to changes in the "selection" of skill types employed in high-tech production – that is, changes in the average $\sigma(i)$ and $\alpha(i)$ of types employed with the H technology. Adão (2016) shows that, depending on the shape of $\alpha(i)$, these selection forces may amplify or offset the impact of ω_t on the average income of high-tech employees.

Skill investment. We consider an overlapping generations setting in which the birth and death of workers follows a Poisson process with rate δ.[6] 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 future path for the wage distribution $\{w_s(i)\}_{s>t}$, 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 exogenous distribution of innate ability, $\bar{s}_t(i)$. A worker's type is then fixed

[6] In Appendix Section E, we consider an extension where population grows because the birth and death rates differ.

during their lifetime.

Formally, we assume that the cost of the lottery is proportional to the Kullback-Leibler divergence between the lottery $\tilde{s}_t(i)$ and the baseline distribution $\bar{s}_t(i)$:

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

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-type 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 generations 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 high- to low-tech production from (4)–(5).

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

1. *Given ω_t, the technology-skill assignment is given by workers self-selection decisions according to (7).*
2. *Given $\{\omega_t\}_{t\geq 0}$, the skill lottery is given by the skill investment decisions of new generations according to (9).*
3. *Given $s_0(i))$ and $\{\tilde{s}_t(i)\}_{t\geq 0}$, the skill distribution satisfies the Kolmogorov-Forward equation (10).*
4. *The ideal price index is given by (3).*

5. *For all $t \geq 0$, the technology-skill assignment, the skill distribution, the relative value of output, and the relative wage satisfy the market clearing conditions* (11).

Discussion. A number of comments on the assumptions and their economic interpretation are in order. There are admittedly three strong assumptions that we make for simplicity and tractability. The first is that $\bar{s}_t(i)$ is exogenous. The second is that only new incoming generations can invest in skills in response to changes in relative wages. Old generations may only respond to such wage changes by moving across technologies but not by changing their skill-type. The third is that skill investment has an uncertain outcome represented by the skill lottery $\tilde{s}(i)$ whose cost takes the particular functional form in (9).

As can be seen from (10), the first and second assumptions imply that the flow of new workers to a particular point in the skill distribution is independent of the current skill distribution. This simplifies the law of motion of the skill-distribution and allows us to characterize its dynamics in general equilibrium. In Appendix Section E, we relax both assumptions. First, we endogenize $\bar{s}_t(i)$ by considering an extension where workers can "learn from others." Specifically, we assume that $\bar{s}_t(i)$ depends on the current skill distribution in a way that makes it is easier for workers to target skills that are already abundant. Second, we allow old generations to re-optimize their skill investments as well. We show that what is important for our main results is that the cost of skill investment is lower for younger generations when compared to that of older generations. It is not essential that this cost is infinite for older generations, as in our baseline specification.

Our preferred economic interpretation of these assumptions is that changes in relative wages may induce older workers in the labor market to switch towards sectors or occupations that require similar skills and may thus entail minimal re-training. Older workers face a high cost to fundamentally change career paths by acquiring completely different skills. Yet, for younger workers, such skill investments are less costly due to lower opportunity cost or higher ability to learn new skills.[7] For tractability, we collapse these investments that in reality occur either through formal schooling or on-the-job into a one-time decision upon entering the market.

Regarding the third assumption, we make it purely for reasons of tractability. Different than in theories of uni-dimensional human capital investment, workers in our theory can direct their investments to target specific skill types in the continuum of existing skills. Yet, mathematically, this *directed skill investment* problem is in principle substantially more complex. As we will see below, the uninsurable uncertainty in the skill investment of ex-ante identical individuals simplifies the *directed skill investment* problem. This allows us to analyze the consequences of technology-skill specificity for the economy's dynamic adjustment to technological innovations.

Our preferred interpretation for the uncertainty of the skill-type realization is that in-

[7] In line with this interpretation, Lee and Wolpin (2006) show that older workers exhibit much lower mobility across occupations and sectors than younger workers.

dividuals with different unobservables may have heterogeneous returns to education and on-the-job training. This is in fact consistent with the evidence in Carneiro, Heckman, and Vytlacil (2011). Our model treats this heterogeneity in unobservables through the uncertainty of the type realization.

Our analysis is also simplified by the choice of the entropy cost function of skill investment. This function has a long tradition in macroeconomics. It has been used to tractably compare the distance between distributions – e.g., in frameworks with rational inattention (Sims (2003)) and model uncertainty (Hansen and Sargent (2008). As discussed later, 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.[8] In our theory, the combination of a continuum of skill types, continuous time, and the entropy cost function implies that the dynamic adjustment of all outcomes is smooth along the equilibrium path. This allows us to sharply characterize the economy's transitional dynamics following technological innovations in general equilibrium.

2.1 Static and Dynamic Equilibrium Conditions

We now proceed to derive equilibrium conditions in two steps. First, we consider static conditions that, given the skill distribution $s_t(i)$ at time t, determine the assignment of workers and the relative wage and value output of the high-tech input, $\{\omega_t, y_t\}_t$. Second, we consider dynamic conditions that, given the path of the relative wage $\{\omega_t\}_{t\geq 0}$, determine the optimal skill lottery chosen by entering generations $\{\tilde{s}_t(i)\}_{t\geq 0}$ and thus the evolution of the skill distribution $\{s_t(i)\}_{t\geq 0}$.

Static equilibrium conditions. The endogenous sorting decision in (7) determines the assignment of skill types to technologies. It 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. Since $\sigma(i)$ is increasing, the assignment is described by a threshold l_t characterizing the type that is indifferent between working with any of the two technologies. The following lemma formalizes this discussion.

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*

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

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

[8] With a discrete number of skill types, our specification yields skill choices that are isomorphic to those implied by a discrete-choice problem a la McFadden et al. (1973). It can thus be seen as a generalization of this framework when there is a continuous of available choices.

of workers across technologies. The slope of $\sigma(l_t)$ determines the strength of the *comparative advantage* in high-tech production of skill types slightly above l_t compared to that of skill type l_t. Thus, as shown by Acemoglu and Autor (2011) and Costinot and Vogel (2010), it essentially determines how much relative wages must change to induce the reallocation of skill types above l_t. Accordingly, the inverse elasticity of $\sigma(i)$ controls the mass of skill types that reallocate across technologies in response to changes in the relative wage. Formally, (12) implies that

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

where $l_t(\omega_t)$ is the implicit function defined by (12). Since the economy's skill distribution does not adjust instantaneously, the inverse elasticity of $\sigma(i)$ plays the role of short-run skill supply across technologies. In the rest of the paper, we refer to the elasticity of $\sigma(i)$ (i.e., $1/\eta$) as the technology-skill specificity.

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

$$x_t(l_t, s_t) = \frac{\int_{l_t}^{1} \sigma(i)\alpha(i)s_t(i)di}{\int_{0}^{l_t} \alpha(i)s_t(i)di}. \tag{13}$$

The threshold l_t is then determined by the market clearing condition in (11). Whenever l_t is high, equation (12) implies that ω_t is low and, therefore, the relative demand for input H is high. In this case, however, the relative high-tech supply is low as only a small share of types are employed in high-tech production. Whenever l_t is low, the opposite is true. In equilibrium, relative demand and supply are equalized. The following lemma formalizes the existence of a unique equilibrium threshold for l_t for any given distribution $s_t(i)$.

Lemma 2 (Equilibrium Threshold) *Given $s_t(i)$ and A_t, there is a unique equilibrium threshold l_t which is the solution to*

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

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 (Optimal 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*

optimal 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 a multinomial logit function over 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 $\log(W(i)) = \frac{\log(\alpha(i)\max\{\omega\sigma(i),1\})}{\rho+\delta}$ is the present discounted log-wage 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.[9] 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, ψ governs the long-run skill supply across technologies, which we formally define as

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

In the rest of the paper, we refer to $1/\psi$ as the cost of adjusting skill investment, which is inversely related to the long-run skill supply across technologies.

2.2 Skill-distribution Dynamics: a q-theory of skill investment

We now combine the static and dynamic equilibrium conditions to solve for the equilibrium path of the skill-distribution as well as all other equilibrium variables, given an arbitrary initial skill distribution $s_0(i)$ and a constant path for $\{A_t, \bar{s}_t(i)\}_{t\geq 0}$.

In principle, this involves solving a a complex infinite-dimensional fixed-point problem. To see this, consider a conjectured path for the relative technology-specific wages ($\{\omega_t\}_{t\geq 0}$). This path determines the skill-investment decisions of new generations being born from (15) and, as such, a path for the skill-distribution from (10) given an initial skill distribution. Furthermore, the path of relative wages determines a path for the assignment of workers to the two technologies ($\{l_t\}_{t\geq 0}$) from the indifference condition (12). Taken together, the

[9] Notice that the long-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.

skill-distribution and the assignment threshold determine the relative supply of the high-tech input ($\{x_t\}_{t\geq 0}$). In an equilibrium, the relative supply of the high-tech input needs to be equal to its relative demand at the the conjectured path for relative wages, i.e., they need to be consistent with market-clearing.

Our first result approximates the solution of this fixed-point problem by considering a log-linear expansion around the stationary equilibrium. It establishes that the approximate equilibrium of this economy can be represented as a q-theory of skill investment, where q refers to the present discounted value of the log-relative wage or, as we call it from now on, lifetime welfare inequality:[10]

$$log(q_t) \equiv \int_t^{\infty} e^{-(\rho+\delta)(s-t)} log(\omega_s) ds$$

Specifically, we show that one need not keep track of the whole skill-distribution in order to solve for the equilibrium path of relative wages and the assignment threshold. This significantly reduces the dimensionality of the fixed-point problem. Instead, the approximate equilibrium path for the skill distribution is only a function of q_t an the assignment threshold threshold l_t. The equilibrium dynamics of these two variables are in turn described by a simple system of linear differential equations.

Letting " $\hat{}$ " denote variables in log-deviations from the stationary equilibrium, the following proposition presents the system of differential equations that, given $\hat{l}_0$,[11] determines the equilibrium path of $\{\hat{q}_t, \hat{l}_t\}$ when $\{A_t, \bar{s}_t(i)\}_{t\geq 0}$ are constant over time. The corollary then characterizes the approximate dynamics of the skill-distribution $\hat{s}_t(i)$, the optimal skill lottery $\hat{\bar{s}}_t(i)$ and relative value of output $\hat{y}_t$, given the equilibrium path for $\hat{q}_t, \hat{l}_t$ and the initial skill-distribution $s_0(i)$.

Proposition 1 (q-theory of Skill Investment) *Suppose that $\{A_t, \bar{s}_t(i)\}_{t\geq 0}$ are constant over time.*

1. *Given initial condition $\hat{l}_0$ and terminal condition $\lim_{t\to\infty} \hat{l}_t = 0$, the equilibrium dynamics of $\{\hat{q}_t, \hat{l}_t\}$ are described by the system of differential equations*

$$\frac{\partial \hat{l}_t}{\partial t} = -\delta \hat{l}_t + \frac{\eta\psi}{\theta + \kappa\eta} \delta \hat{q}_t \tag{17}$$

$$\frac{\partial \hat{q}_t}{\partial t} = (\rho + \delta)\hat{q}_t + \frac{1}{\eta}\hat{l}_t, \tag{18}$$

   where $\kappa > 0$ is a constant.

2. *The equilibrium $\{\hat{q}_t, \hat{l}_t\}_{t\geq 0}$ is saddle-path stable and given by:*

$$\hat{l}_t = \hat{l}_0 e^{-\lambda t} \quad \text{and} \quad \hat{q}_t = \zeta \hat{l}_t \tag{19}$$

[10] Remember that $log(\omega_t)$ is a natural measure of welfare inequality at a given point in time since it captures the relative log-wage of skill types employed in different technologies conditional on their productivity.

[11] Note that an initial l_0 is determined by the initial skill-distribution from the static equilibrium condition (14).

where

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

Proof. See Appendix A.3. ■

The first part of the proposition presents a system that is a rather standard one in macroeconomics. The assignment threshold, $\hat{l}_t$, is a state variable whose law of motion needs to be solved backward. The present discounted value of relative technology-specific wages, $\hat{q}_t$, is a control variable whose law of motion needs to be solved forward. The system is in fact mathematically isomorphic to the q-theory of capital investment (Tobin, 1969, Hayashi, 1982). In our model, however, $\hat{q}_t$ is the present discounted value of the relative wage of the high-tech input. In other words, it is the shadow price of the human capital "asset" associated with having one additional unit of high-tech good. Whenever this price is higher, the incentives to invest in high-i skills are stronger. As entering generations replace old ones at rate δ, this increases the relative supply of high-tech input and decreases its relative price over time.

As in the seminal q-theory, parameters governing the costs of adjustment in the economy (i.e., δ and ψ) affect the sensitivity of changes in the assignment threshold $\frac{\partial \hat{l}_t}{\partial t}$ to $\hat{q}_t$. However, our model features both imperfect substitution of human capital across technologies and heterogeneous skills. Thus, the impact of q_t on the evolution of l_t also depends on the degree of technology-skill specificity (as measured by η) and the substitutability of inputs (as measured by θ).

The second part of the proposition shows that (locally) the equilibrium exists and is unique—a consequence of saddle-path stability. Given an initial condition $\hat{l}_0$, both $\hat{l}_t$ and $\hat{q}_t$ converge at a constant rate of λ to the stationary equilibrium. The expressions in (19) show that, whenever $\hat{l}_0 < 0$, the assignment threshold l_t increases and q_t decreases along the equilibrium path. In this case, the high value of q_0 at time zero (i.e., $\hat{q}_0 < 0$) induces investment in high-i types, which increases the supply of the high-tech good and, consequently, reduces its relative value. The decline in ω_t over time implies that only relatively higher i types find it profitable to sort themselves into the high-tech sector, raising the threshold l_t.

Having characterized the equilibrium dynamics of $\hat{q}_t$ and $\hat{l}_t$, the following corollary characterizes the approximate dynamics of the remaining supply-side variables: the skill-distribution, the optimal skill-lottery, and the relative value of high-tech output.

Corollary 1 (Supply Dynamics) *Given $\hat{s}_0(i)$ and the equilibrium path for $\{\hat{l}_t, \hat{q}_t\}_{t\geq 0}$, the equilibrium dynamics of the skill distribution $\hat{s}_t(i)$, the optimal lotteries $\hat{s}_t(i)$, and the value of relative*

high-tech output $\hat{y}_t$ are approximated by:

$$\hat{\hat{s}}_t(i) = \left(\mathbb{I}_{i>l} - \int_l^1 s(i)di \right) \psi \hat{q}_t + o_t(i) \tag{20}$$

$$\hat{s}_t(i) = \hat{s}_0(i)e^{-\delta t} + \int_0^t e^{\delta(\tau - t)} \hat{\hat{s}}_\tau(i) d\tau \tag{21}$$

$$\hat{y}_t = (\theta - 1)\frac{1}{\eta}\hat{l}_t \tag{22}$$

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

Proof. See Appendix A.4 ■

Corollary 1 links the equilibrium path of the skill distribution and the assignment threshold to the joint dynamics of $\{\hat{q}_t, \hat{l}_t\}$. The change in the optimal skill lottery in (20) along the transition depends centrally on the evolution of the relative return of skills employed in the high-tech production. In particular, changes in $\hat{q}_t$ affect the relative lifetime earnings of types employed in the H-technology. The parameter ψ controls the sensitivity of the optimal skill investment to such changes. The overall skill distribution in (21) 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 value of relative high-tech output is driven by changes in relative wages $\hat{\omega}_t = -\frac{1}{\eta}\hat{l}_t$. The sensitivity of such wage changes to changes in the assignment threshold's depends on the degree of technology-skill specificity controlled by the short-run skill supply elasticity η. We obtain the expression in (22) because the demand equation in (2) implies that $\hat{\omega}_t = \frac{1}{\theta - 1}\hat{y}_t$.

One important implication of Proposition 1 is that the approximate equilibrium $\{\hat{q}_t, \hat{l}_t\}_{t \geq 0}$ can be solved for without keeping track of the evolution of the whole skill distribution $s_t(i)$. Given $\{\hat{q}_t, \hat{l}_t\}_{t \geq 0}$, Corollary 1 immediately yields the transitional dynamics of $s_t(i)$. This significantly reduces the dimensionality of state-space of the equilibrium's fixed-point problem. We achieve this by noticing that the dynamics of $s_t(i)$ only depend on $log(Q_t(i)) = \int_t^\infty e^{-(\rho+\delta)(s-t)} \max\{log(\omega_s \sigma(i)), 0\} ds$ via the optimal skill-lottery. Yet, as long as relative wages are not too far from their stationary level along an equilibrium path, most worker types never switch sectors. The only workers that do are those who are marginal. This implies that the present discounted utility of skill-types far above the assignment threshold only changes over time because the present discounted value of the relative wage changes over time. That is, we have that $log(Q_t(i)) = log(q_t) + \frac{log(\sigma(i))}{\rho + \delta}$. Furthermore, the market-clearing condition (14) determining the assignment l_t only contains integrals of $s_t(i)$. The equilibrium characterization shows that, because of the continuum of skill types, the effect of worker switches are of second order when evaluating such integrals. This then implies that the approximate dynamics of l_t depend only on $log(q_t)$ and not on the dynamics of the full skill-distribution. Finally, since relative wages are fully determined by l_t alone via the

indifference condition (12), then $log(q_t)$ only depends on the future path of l_t. This then gives rise to the system of linear differential equations in (17)-(18).

3 The Adjustment to Skill-biased Technological Innovations

We now analyze the dynamic adjustment of our economy to a permanent, unanticipated 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 use the results from the previous section to characterize in closed-form the dynamic responses of q_t, l_t and y_t, as well as the evolution of the skill distribution $s_t(i)$. The dynamic responses show that the impact of new technologies on the economy may significantly change over time due to the endogenous evolution of the skill distribution across generations. The economy's adjustment is thus shaped by a form of Samuelson's LeChatelier principle: the adjustment in the skill investment decisions of incoming cohorts implies that the elasticity of relative output supply increases over time.

Our results formally establish the role that changes in the skill distribution across generations play in shaping the economy's adjustment to technological innovations. It provides a micro-foundation for the idea that supply elasticities are lower at shorter horizons compared to longer horizons. In the rest of the paper, we show how this micro-foundation is useful for understanding the determinants and implications of skill-biased technological innovations, as well as point to the risks involved in using *naive* theories to extrapolate from either past transitions or observed changes at short horizons.

3.1 Dynamic responses of equilibrium outcomes

We assume that immediately prior to the shock at time $t = 0^-$ the economy is in a stationary equilibrium. We let $\Delta \log(A) > 0$ be the relative productivity shock, and denote log-changes in equilibrium outcomes as $\Delta \log(q_t) \equiv \log(q_t/q_{0^-})$, $\Delta \log(y_t) \equiv \log(y_t/y_{0^-})$, and $\Delta \log(l_t) \equiv \log(l_t/l_{0^-})$.

Proposition 2 (Dynamic responses) *Given a skill-biased technological innovation $\Delta \log(A)$, the dynamic responses $\Delta \log(l_t)$, $\Delta \log(q_t)$ and $\Delta \log(y_t)$ are approximated by:*

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

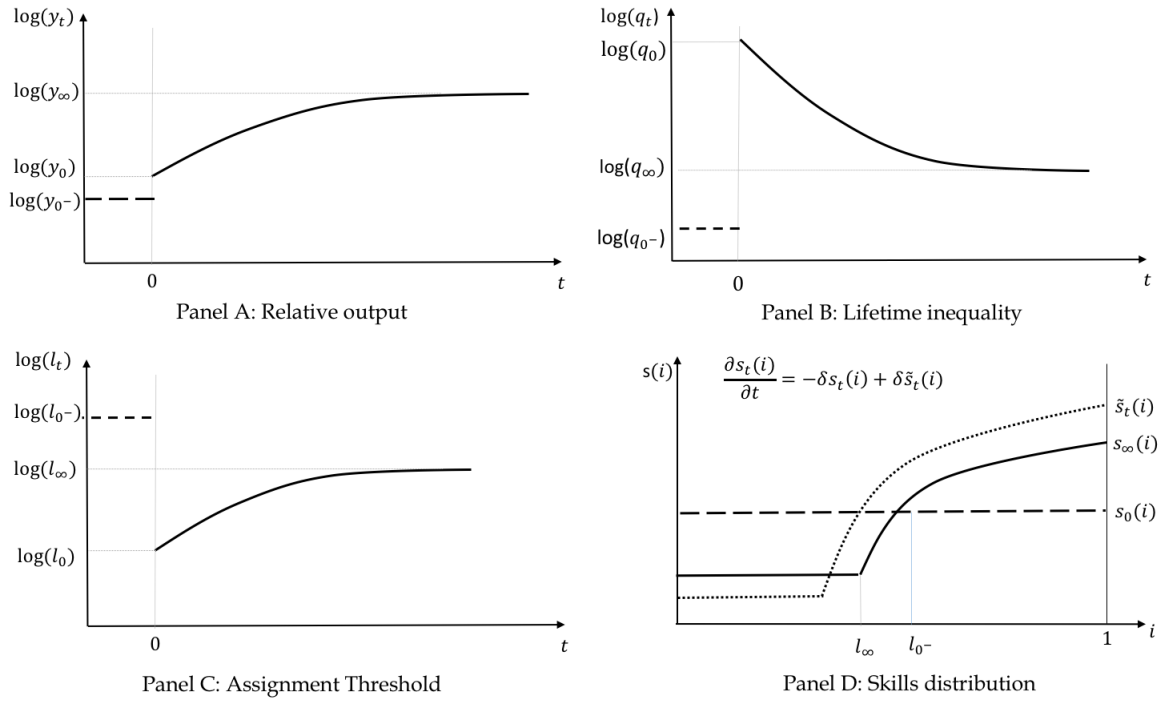
where $\chi \equiv \left(\theta + \kappa\eta + \frac{\psi}{\rho + \delta}\right)(\rho + \delta)$.

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, whereas the assignment threshold l_t falls. However, along the transition, q_t falls, and both l_t and y_t increase at rate λ. To derive these expressions, first note that the results in Proposition 1 and Corollary 1 immediately yield the transitional dynamics right after impact given an initial condition for $\hat{l}_0 = \log(l_0/l_\infty)$. Then, to characterize the full dynamic responses, we derive expressions for the short- and long-run changes implied by the technological shock when $s_0(i)$ is the stationary distribution before the shock.

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 Corollary 1's characterization of the evolution of the skill distribution. Figure 1 provides a graphical illustration.

Figure 1: The economy's adjustment to a skill-biased technological shock



Panel A: Relative output

Panel B: Lifetime inequality

Panel C: Assignment Threshold

Panel D: Skills distribution

The productivity increase causes an increase in the demand for workers in the high-tech production. In equilibrium, the relative wage 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).

3.2 LeChatelier Principle in Action: Supply and Demand Framework

To provide further intuition for the impulse response functions above, we connect our theory to a supply and demand framework. At each point in time, the relative output and wage in sector H solve the following system of equations:

$$\Delta log(x_t) = (\theta - 1)\Delta log(A) - \theta \Delta log(\omega_t), \tag{23}$$

$$\Delta log(x_t) = \varphi_t \Delta log(\omega_t). \tag{24}$$

The first expression is the "relative demand equation" in (2). As discussed above, it is the cornerstone of the canonical model in Katz and Murphy (1992) and its extensions reviewed by Acemoglu and Autor (2011). The demand equation relates changes in relative demand, $\Delta log(x_t)$, to changes in relative productivity, $\Delta log(A)$, and relative wages, $\Delta log(\omega_t)$. In our setting, θ plays the role of the elasticity of substitution between the output of skill types employed with different technologies.

The second expression is the "relative supply equation" linking changes in relative output supply, $\Delta log(x_t)$, to changes in relative wages, $\Delta log(\omega_t)$. The parameter φ_t is the elasticity of relative output supply, which is a function of the degree of skill-technology specificity and the cost of adjusting skill investment. Specifically, Proposition 2 implies that

$$\tilde{\varphi}_t = \frac{\kappa\eta + \frac{\psi}{\rho+\delta}\left(1 - \frac{\theta}{\theta+\kappa\eta}e^{-\lambda t}\right)}{1 + \frac{\psi}{\rho+\delta}\frac{1}{\theta+\kappa\eta}e^{-\lambda t}}. \tag{25}$$

At every point in time, the relative supply elasticity is positive, $\varphi_t > 0$. Importantly, it is straight forward to show that this elasticity is increasing over time. Thus, our theory exhibits a form of Samuelson's LeChatelier principle: the elasticity of relative output supply is higher over longer horizons. The higher elasticity in the long-run is a direct consequence of the adjustment in the skill investment decisions of incoming cohorts in response to the skill-biased technological innovation. As these new cohorts slowly take over the labor force, the supply of skills to high-tech production slowly increases, implying a higher elasticity of relative output supply over longer horizons.

We can now return to the impulse response functions in Figure 1. Since φ_t increases over time, the same shock will generate stronger relative output responses and smaller relative wage increases over longer horizons. Thus, as in Samuelson's LeChatelier principle, the initial impact of the shock becomes weaker over time.

This discussion illustrates the risks of extrapolating from observed responses in the economy over any given horizon. To make this point clear, consider a researcher who knows θ and obtains φ_T and $\Delta \log A$ from the estimated impact of a technological shock on relative output and wages at horizon T.

Now suppose that the researcher uses these estimates to analyze the impact of technological innovations. It is clear that the time-varying nature of the reduced-form parameter φ_t implies that predictions will be biased for any period other than T. Specifically, the researcher's predictions will overestimate (underestimate) inequality changes and underestimate (overestimate) relative output changes for any period after (before) horizon T. The magnitude of the bias depends on the change in φ_t over time. In our theory, this is a function of the parameters governing technology-skill specificity (η) and the skill investment cost (ψ). Section 5 shows that failing to account for the transitional dynamics in φ_t also introduces bias in the quantification of the average and distributional consequences of technological innovations.

This researcher also faces another type of bias. Whenever the nature of the technology or the underlying flexibility of skill investment is different, the parameters η and ψ will be different, implying that the relative supply elasticity φ_T will also be different. This type of concern is important even if the researcher matches the dynamics of φ_t from a particular episode as these parameters affect the entire path of φ_t. Section 4 investigates how η and ψ affect the transitional dynamics through their impact on φ_t at different horizons.

4 Determinants of Skill-biased Technological Transitions

In this section, we analyze how parameters governing technology-skill specificity and the skill investment cost affect the economy's adjustment to a skill-biased technological innovation. This comparative statics exercises show why is it that some technological transitions are different than others. They speak to past episodes (or future transitions) with differences in the nature of technologies, skills, and educational institutions involved. Furthermore, they highlight which classes of economies are associated with larger in differences in supply elasticities at longer horizons and should thus give researchers more caution when extrapolating from observed changes at shorter horizons.

4.1 Comparative Statics with respect to Technology-Skill Specificity (η)

Consider first how the economy's impulse response functions respond to changes in technology-skill specificity (i.e., how different skill types are in terms of relative productivity in the high-tech sector). In our theory, specificity is inversely related to the short-run skill supply elasticity η. Thus, this exercise speaks to differences in the transitional dynamics across events in which skills of incumbent workers were more or less easily transferable for use in the new improved technology.

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

Proposition 3 (Comparative statics with respect to η) *Assume that $\theta > 1$. Then,*

1. *Short-run adjustment*

$$\frac{\partial \Delta log(y_0)}{\partial \eta} > 0, \quad \frac{\partial |\Delta log(l_0)|}{\partial \eta} > 0, \quad \frac{\partial \Delta log(q_0)}{\partial \eta} < 0$$

2. *Long-run adjustment*

$$\frac{\partial \Delta log(y_\infty)}{\partial \eta}, \quad \frac{\partial |\Delta log(l_\infty)|}{\partial \eta} > 0, \quad \frac{\partial \Delta log(q_\infty)}{\partial \eta} < 0$$

3. *Persistence*

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

Proof. See Appendix A.6. ■

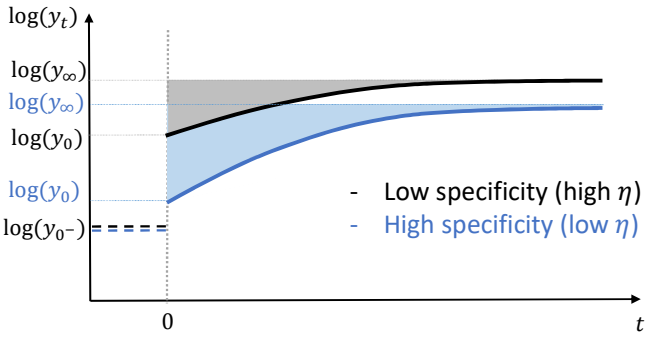
To fix ideas, Figure 2 illustrates the results in Proposition 3 with the impulse response functions of two economies.[12] The black lines show the responses of an economy with a high value of η (i.e., high short-run skill supply elasticity or low technology-skill specificity). The blue lines show the 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, when technology-skill specificity is higher (lower η), a smaller mass of workers reallocate across technologies in response to the shock (as can be seen from $\frac{\partial |\Delta log(l_0)|}{\partial \eta} > 0$). As a result, the increase in relative wages (and lifetime inequality) on impact is larger and the increase in relative output smaller. Then, the larger increase in relative wages (current and future) implies that younger entering generations have stronger incentives to invest in the skills that became more valuable. As a consequence, there are larger differences in skill heterogeneity across generations. This then implies that the transitional dynamics of y_t and q_t are less persistent, as measured by the cumulative impulse response (e.g., $\int_0^\infty \hat{q}_t dt$), because larger changes in the skill distribution take place as younger generations replace older generations. Finally, while the smaller changes in the skill distribution could have implied a smaller (larger) overall long-run adjustment in relative output (lifetime inequality), it turns that the larger (smaller) short-run response dominates. Thus, the long-run adjustment in relative output (lifetime inequality) is larger (smaller).

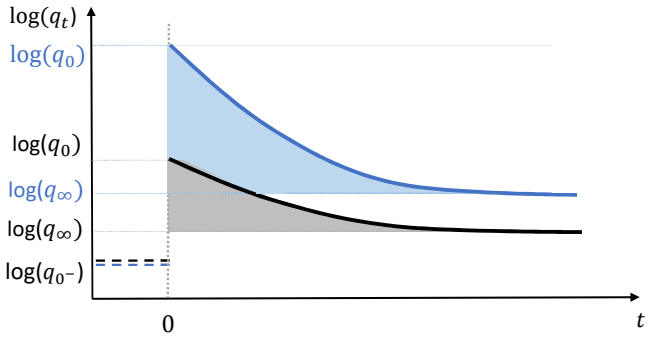
These results show that economies with higher technology-skill specificity feature overall smaller responses in relative output and larger responses in lifetime inequality at all

[12] The figure shows the case where the threshold's cumulative impulse response increases with η.

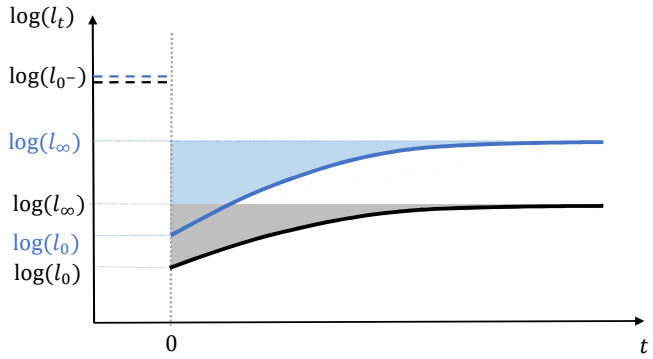
Figure 2: Comparative statics with respect to η



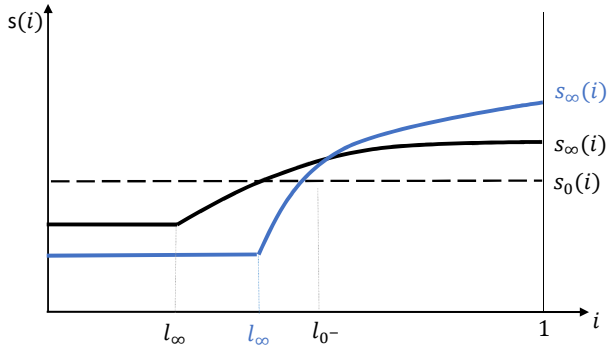
Panel A: Relative output



Panel B: Lifetime inequality



Panel C: Assignment Threshold



Panel D: Skills distribution

horizons, as well as a slower and more persistent adjustment path to the new long-run equilibrium.

4.2 Comparative Statics with respect to Skill Investment Cost (ψ)

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. Thus, it speaks to differences across historical episodes where young workers may have found it easier to invest in skills in high demand due to, for example, better educational systems, the availability of vocational training, or better opportunities to learn on the job.

Proposition 4 (Comparative statics with respect to ψ)

1. *Short-run adjustment*

$$\frac{\partial \Delta log(y_0)}{\partial \psi} = 0, \quad \frac{\partial |\Delta log(l_0)|}{\partial \psi} = 0, \quad \frac{\partial \Delta log(q_0)}{\partial \psi} > 0$$

2. *Long-run adjustment*

$$\frac{\partial \Delta log(y_\infty)}{\partial \psi} > 0, \quad \frac{\partial |\Delta log(l_\infty)|}{\partial \psi} < 0, \quad \frac{\partial \Delta log(q_\infty)}{\partial \psi} < 0$$

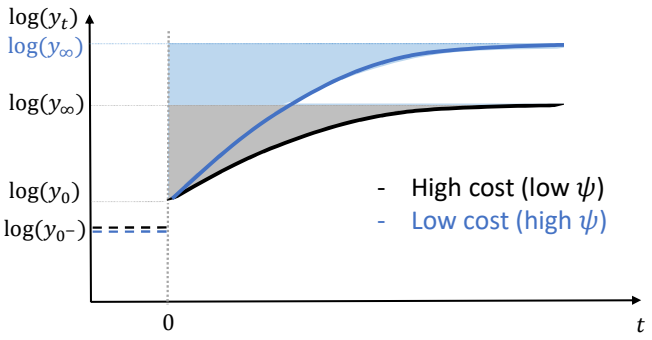
3. *Persistence*

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

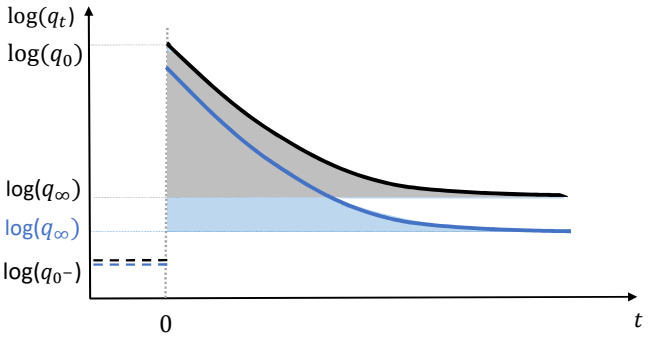
Proof. See Appendix A.6. ■

Figure 3 illustrates Proposition 4 with a graphical representation of the impulse response functions of two economies. 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, since skill investment decisions do not change if $1/\psi$ is zero.

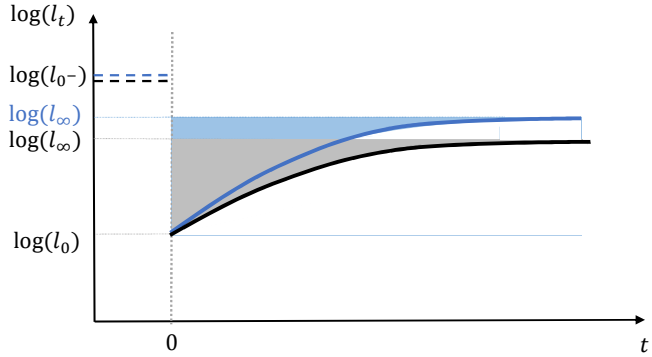
Figure 3: Comparative statics with respect to ψ



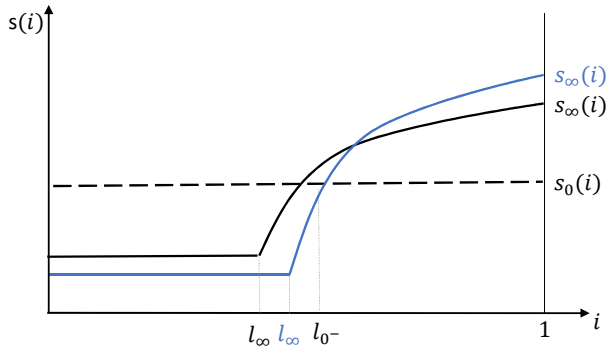
Panel A: Relative output



Panel B: Lifetime inequality



Panel C: Assignment Threshold



Panel D: Skills distribution

The first part of Proposition 4 indicates that, in the short-run, both economies exhibit identical responses in relative output and worker allocation. This follows from the fact that ψ does not affect the self-selection decisions of generations born before the shock. However, a

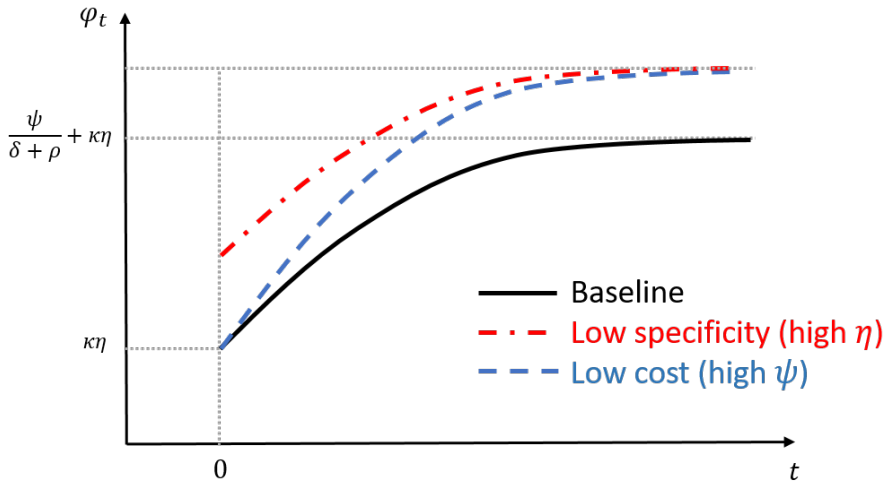
higher ψ attenuates the short-run increase in lifetime inequality because future relative wages fall by more due to the larger increase in the future supply of high-i skills (as can be seen from Lemma 1). The latter also implies that relative output (lifetime inequality) increases more (less) in the long-run. Finally, because of the larger change in the skill distribution along the transition, the persistence of both lifetime inequality and relative output are higher when ψ is higher. Thus, economies with a larger long-run skill supply elasticity (lower cost of skill investment) exhibit slower, more persistent adjustment in relative output and wages.

4.3 Discussion

The results above show that the dynamic adjustment differs in economies with lower levels of technology-skill specificity (i.e., higher η) or skill investment cost (i.e., higher ψ). Increasing either η or ψ yield similar qualitative implications in the long-run. In both cases, there is a stronger long-run increase in relative output and a weaker long-run increase in lifetime inequality. However, while a higher η reduces the adjustment's persistence by front-loading changes in all outcomes, a higher ψ increases the persistence of the transitional dynamics by back-loading the response in relative output and relative wage.

To understand these differences, it is useful to return to the supply-demand representation of the economy's adjustment introduced in Section 3.2. The different dynamic implications of changing η or ψ arises because the two parameters shape different horizons of the reduced-form elasticity of relative output supply. As illustrated in Figure 4, higher values of η and ψ increase the elasticity of relative supply in the long-run. However, the timing of the increase in φ_t differs when the economy has a higher η or a higher ψ. Specifically, increasing η flattens the path of φ_t, but increasing ψ steepens the evolution φ_t.

Figure 4: Effect of η and ψ on the elasticity of relative output supply (φ_t)



Intuitively, the different implications for the path of φ_t follow directly from the impact of η and ψ on between-generation skill differences. A lower cost of adjusting skill investment

(i.e., higher ψ) implies that, in response to the shock, it is easier for new generations to adjust their skills, amplifying the distance between the skill distribution of new and old generations. In contrast, a lower level of technology-skill specificity (i.e., higher η) makes it easier for skill-types to reallocate across technologies in response to the shock. This reduces the relative wage change and, therefore, the incentives of young workers to modify their skill investment compared to that of old generations.

This discussion highlights that the shape and persistence of skill-biased technological transitions varies across economies with different levels of technology-skill specificity or skill investment cost. They establish which economies should cause researchers to exercise more caution when extrapolating from observed responses at short horizons because increases in relative output and decreases in relative wages are back-loaded. This is the case precisely when technology-skill specificity is high and the skill investment cost is low. These results also indicate that researchers should be cautious when making predictions based on lessons from historical episodes in which the nature of technologies, skills, or education institutions were different. Such features are likely to affect technology-skill specificity and/or skill investment cost, causing the transition to be different.

5 Welfare Analysis of Skill-biased Technological Transitions

We now evaluate the welfare consequences of skill-biased technological innovations. We show how the economy's transitional dynamics shape the innovation's impact on average welfare and lifetime inequality of different worker generations.

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. Given our log-utility assumption, we obtain the consumption-equivalent utility by multiplying the ex-ante utility by $(\delta + \rho)$. In our setting, the ex-ante utility of a 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 consumption-equivalent utility of cohort τ as

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

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 wage is the only endogenous component of the relative earnings of skill types employed in different technologies. This implies that it will be useful to define the lifetime welfare inequality for cohort τ as the consumption-equivalent of the present discounted value of the relative wage, $(\delta + \rho) \log q_\tau$. We again aggregate different worker cohorts by defining the economy's discounted lifetime welfare inequality at t as

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

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 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 induced changes in average welfare $\Delta\bar{U} \equiv \bar{U}_0 - U_{0^-}$ and lifetime inequality $\Delta\bar{\Omega} \equiv \bar{\Omega}_0 - log(q_{0^-})$.

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

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

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.

Proof. See Appendix A.7. ■

The change in average welfare, $\Delta\bar{U}$, combines the impact of the shock on both the present value of 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 $\frac{y_\infty}{1+y_\infty}\Delta \log(A)$, and (ii) the endogenous relative wage change captured by the term $\frac{y_\infty}{1+y_\infty}\Delta\bar{\Omega}$. In addition, the change in average welfare responds to the change in the average lifetime earnings of workers. Again, since only the H-specific wage changes, the change in average lifetime earnings is the product of the high-tech employment share, $\frac{e_\infty}{1+e_\infty}$, and the change in lifetime inequality, $\Delta\bar{\Omega}$.

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, between-generation differences 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.

Discussion. Taken together, the results above highlight the importance of accounting for the persistence of the transitional dynamics when evaluating the welfare consequences of technological innovations.

To provide further intuition for this point, we illustrate the bias in welfare calculations that ignore the economy's transitional dynamics. We again rely on the supply-demand representation of Section 3.2, and consider the same researcher that has estimates of θ and φ_T. For any shock $\Delta \log(A)$, such estimates permit the calculation of the predicted change in relative wage at time T, $\Delta \log \omega_T$. By ignoring that φ_t changes over time, the researcher assumes that this relative wage change is permanent and mistakenly computes the shock-induced change in lifetime inequality as $\Delta\bar{\Omega}_T^{Static} = (\delta+\rho)\Delta \log q_T = \Delta \log \omega_T$.

It is clear from the dynamics of $\log q_t$ in Figure 1 that, in general, $\Delta\bar{\Omega}_T^{Static}$ is different from the true value of $\Delta\bar{\Omega}$ in Proposition 5. The direction and magnitude of the bias depends on the particular horizon T used in estimation: $\Delta\bar{\Omega}_T^{Static} < \Delta\bar{\Omega}$ when T is large, but $\Delta\bar{\Omega}_T^{Static} > \Delta\bar{\Omega}$ when T is low. To make this point explicit, consider the extreme case of $T = 0$, which

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

yields the highest possible value of $\Delta\bar{\Omega}_0^{Static} = \Delta \log \omega_0$. In this case, the bias is

$$\Delta\bar{\Omega}_0^{Static} - \Delta\bar{\Omega} = \left(1 + \frac{\delta + \rho}{r + \lambda}\right) \lambda^2 \int_0^\infty \hat{q}_t dt > 0.$$

The fact that the bias is proportional to the persistence of inequality is intuitive. Longer periods of inequality decline following the shock amplify the difference between the relave wage at $t = 0$ and the present discounted value of the inequality change, $\Delta\bar{\Omega}$. For any given $\log(A)$, Proposition 5 immediately implies that the researcher underestimates the average welfare gain of the new technology. Again, the size of this negative bias is increasing in the persistence of the adjustment.[14]

6 Application: Cognitive-biased Technological Transitions

Our theoretical results establish that technology-skill specificity and skill investment cost play a central role in the economy's dynamic adjustment to skill-biased technological innovations. In this section, we empirically assess the importance of these mechanisms in shaping recent cognitive-skill-biased technological transitions. We start by connecting the parameters of technology-skill specificity and skill investment cost to the employment adjustment of old and young generations respectively. We then explore this insight to provide three pieces of evidence indicating that these two separate mechanisms affected the transitional dynamics following 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, for recent cognitive-biased innovations, technology-skill specificity is high, but the skill supply elasticity is larger at longer horizons.

[14] In Appendix D.2, we consider researchers taking alternative approaches that ignore the the dynamics of the skill distribution across generations. In particular, we consider a researcher using a static assignment model (i.e., a version of our model with $\psi = 0$) to analyze the impact of a new technology in an economy that behaves according to our theory with $\psi > 0$. Similar biases in the welfare calculation arise when the researcher uses different procedures to match observed changes in relative output and/or relative wages.

6.1 Observable Predictions: Within- and Between-Generation Adjustment of Labor Market Outcomes

This section links the model's predictions to observable labor market outcomes. There are three important challenges to design empirical specifications based on the impulse response functions shown in Section 3. First, as discussed in Section 2, the endogenous skill-technology assignment implies that the relative average labor income is different from the relative wage ω_t. This difference arises from changes in the "selection" of skill types employed in high-tech production implied by changes in ω_t – that is, changes in the average $\sigma(i)$ and $\alpha(i)$ of types employed with the H technology.[15] Second, as in the q-theory of capital investment, q_t is a forward-looking variable whose measurement requires knowledge of the entire equilibrium path of ω_t. So, to construct q_t, we would need to observe ω_t along the entire transition to the new stationary equilibrium. Finally, the direct measurement of the skill distribution $s_t(i)$ and the technology-skill assignment l_t require taking an explicit stance on observable attributes that determine worker skills in different activities (e.g., college graduation or occupation history). The empirical analysis is misspecified whenever the chosen attributes do not completely determine the relative productivity of workers in the two technologies.

Given these challenges, our empirical analysis focuses on observable responses in relative payroll and relative employment across technologies. Specifically, we investigate whether relative payroll y_t slowly increases following the shock. As stated in Proposition 2, the evolution of y_t converges at rate λ. In addition, we use responses in relative employment to indirectly investigate the underlying responses in the skill distribution, the relative wage, and the assignment threshold. As we now show, the evolution of relative employment for different worker generations is a function of the two main mechanisms in our theory: the degree of technology-skill specificity, as captured by η, and the skill investment cost, as captured by ψ.

Consider the same one-time permanent change in A at $t = 0$. We define 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-

[15] Recent empirical applications of assignment models use functional forms for $\sigma(i)$ and $\alpha(i)$ that yield identical distributions of labor income across technologies – for examples, see Hsieh et al. (2013) and Burstein, Morales, and Vogel (2016).

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.[16]

Relative employment of old generation: Technology-skill specificity. In Appendix A.8, we show that the change in the relative employment of old generations is

$$\Delta \log e_t^{old} \approx \frac{\eta}{\theta + \kappa\eta} \frac{1}{e_H} \left(1 - \frac{\psi}{\chi}(1 - e^{-\lambda t})\right) (\theta - 1)\Delta \log A, \tag{26}$$

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

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 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 (26) shows that the magnitude of the increase in relative employment of older generations is decreasing in the degree of technology-skill specificity (i.e., increasing in η). To see this more clearly, consider the relative employment response of old workers at $t = 0$:

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

This expression indicates that, conditional on the shock size, a high η induces a large short-run change in relative employment for older generations. In this case, skills in the economy are easily transferable across technologies, so any change in the relative wage induces a large reallocation of workers.

Relative employment of young generation: Skill investment cost. Turning to the employment response among young generations, Appendix A.8 also establishes that

$$\Delta \log e_t^{young} \approx \Delta \log e_t^{old} + \frac{\psi}{\chi} \frac{1 - e^{-\lambda t}}{1 - (1 - \tilde{x}_0)e^{-\delta t}} (\theta - 1)\Delta \log A. \tag{27}$$

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 approxi-

[16] 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 after the shock, 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.

mated 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 (26) indicates that the between-generation difference in the response of relative employment is decreasing in the skill investment cost (i.e., it is increasing in ψ). To see this more clearly, consider the between-generation difference 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 their skills more in response to changes in future relative wages, which amplifies between-generation differences in employment.

6.2 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. In our theory, we will interpret these activities as those corresponding to high-tech production. 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

[17] In Appendix B.2, 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,

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]

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 occupations 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 6.1, the different employment responses for young and old workers suggest that the relative supply of cognitive-intensive occupations is more elastic in the long-run than in the short-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 investigate the impact of cognitive-biased technologies on the German labor market.

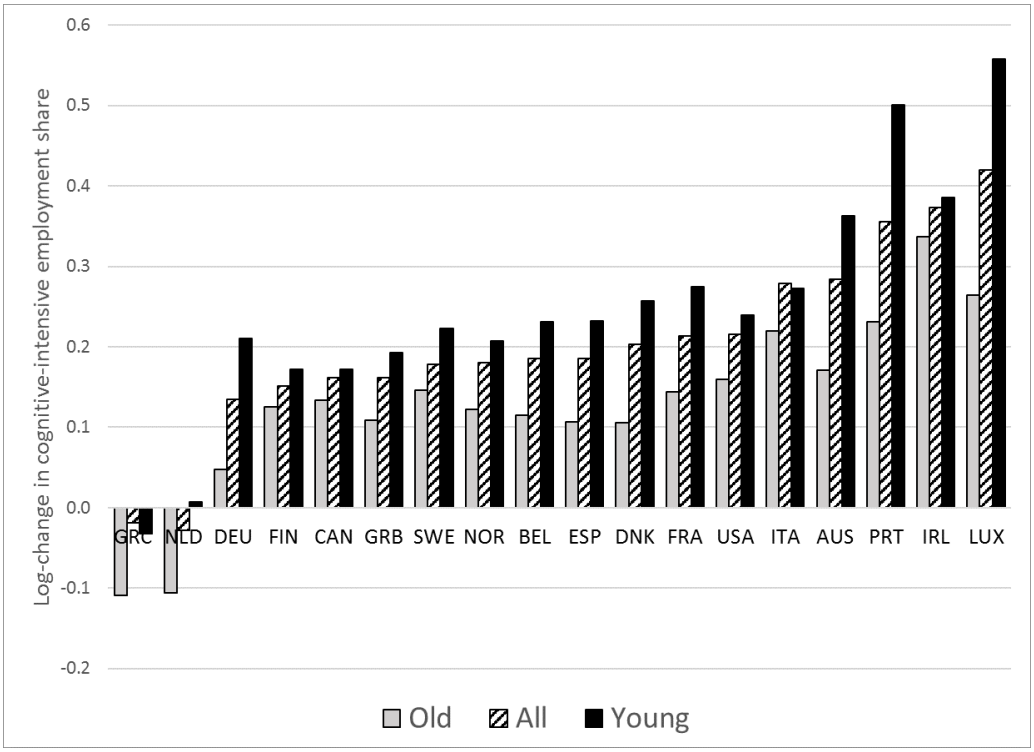
6.3 Cognitive-Intensive Employment Growth and New Technologies: Evidence from Germany

We next study how the German economy adjusted to recent cognitive-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

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: 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.

on occupations with a higher cognitive intensity over time.

6.3.1 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.[20] We first use individual-level information to construct yearly series of employment and payroll 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.[21] We use the BERUFNET dataset discussed previously to define each

[20] Appendix B.1 lists the steps involved in constructing our sample.

[21] We construct our data using the district of the establishment of the main job of each individual in any given year. Since

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 employment 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.

6.3.2 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 (26)–(27), we estimate the following linear regression for each worker generation g and year t:

$$\log Y^g_{o,t} - \log Y^g_{o,1995} = \beta^g_t \bar{C}_o + \epsilon^g_{o,t} \tag{28}$$

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.[22]

Table 1 reports the estimation of equation (28) 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

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.

[22] 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. *** $p < 0.01$

relative changes in the average earnings of those employed in cognitive-intensive occupations. As discussed above, 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). So, the difference between columns (4) and (1) do not correspond to the response of the relative wage per efficiency unit of more cognitive-intensive occupations. In fact, the small responses in relative average earnings for both young and old are consistent with the strong selection forces that arise from assignment models with a Frechet distribution of technology-specific ability, as in Hsieh et al. (2013) and Burstein, Morales, and Vogel (2016).

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 relative supply of cognitive-intensive occupations is more elastic 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 small relative employment response for old generations suggests that skills are very specific to occupations with the same cognitive content.

Robustness Appendix B.4 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 rel-

ative 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.

6.3.3 Dynamic Adjustment to 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 6.1 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.[23] 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 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

$$Y_{io,t} - Y_{io,1999} = (\alpha_t + \beta_t \bar{C}_o)\, DSL_i + \delta_{o,t} + X_{io,t}\gamma_t + \epsilon_{io,t}, \tag{29}$$

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

[23] As shown Akerman, Gaarder, and Mogstad (2015), broadband internet expanded the relative demand for more educated workers in non-routine jobs inside firms. In Appendix B.2, we show that this new technology is disproportionately used by individuals employed in more cognitive-intensive occupations.

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.[24]

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

$$Y^g_{io,t} - Y^g_{io,1999} = \sum_{c \in \{\text{young, old}\}} (\alpha^c_t + \beta^c_t \bar{C}_o)\, 1_{[g=c]} DSL_i + \delta_{o,t} + \zeta_{g,t} + X^g_{io,t}\gamma^g_t + \epsilon^g_{io,t}, \tag{30}$$

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 $X^g_{io,t}$.

We are mainly interested on the impact of broadband internet adoption on the relative outcome of cognitive-intensive occupations: β_t in (29) for the all workers, and β^g_t in (30) 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^g_t is the equivalent difference among workers of generation g.

The consistent estimation of equations (29)–(30) 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

[24] 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 caused by 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 in 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.

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.[25] 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) 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 (29) 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 (30), 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 (29)–(30) 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.[26] That is, we begin by estimating the

[25] 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.

[26] As discussed above, equations (29)–(30) 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

following linear regression for year t:

$$DSI_{i,t} = Z_i \rho_t + X_i \gamma_t + \epsilon_{i,t} \tag{31}$$

where X_i is the vector of district-level controls used in the estimation of (29)–(30).

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.

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

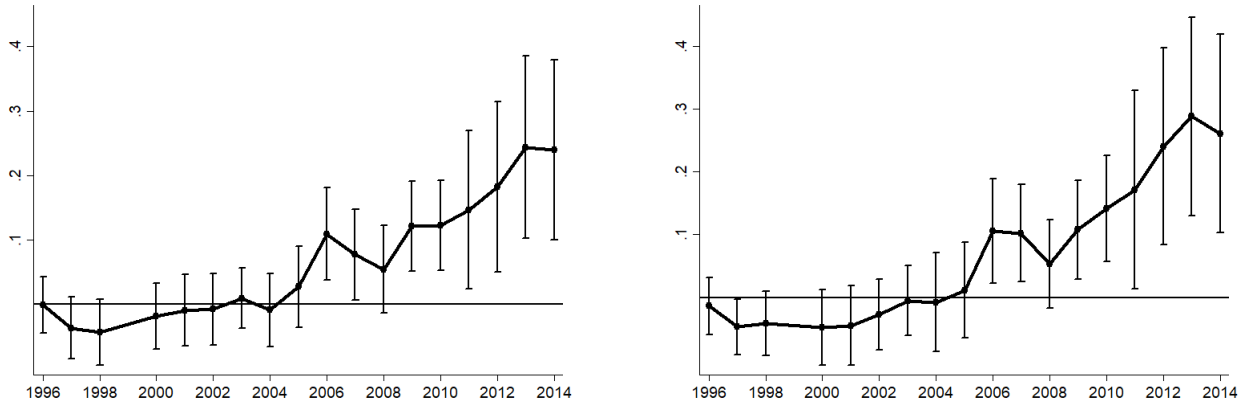
Note. Sample of 323 districts in West Germany. All regressions are weighted by the district population size in 1999. Baseline controls include the following district variables in 1999: college graduate population share, manufacturing employment share, immigrant employment share and workforce age composition.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

We now turn to the estimation of β_t in (29). Figure 6 reports the estimates of broadband internet expansion on the relative employment (Panel A) and the relative payroll (Panel B) of 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.

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

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 (29) 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.

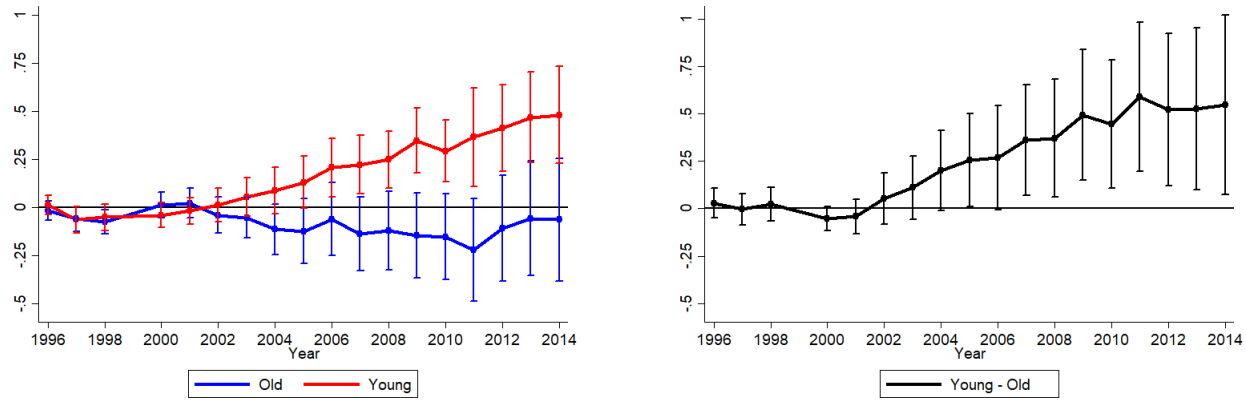
We now turn to the estimation of employment responses for each generation: β_t^{old} and β_t^{young} in (30). Panel A of Figure 7 reports the estimates for each year between 1996 and 2014. Prior to 2003, regions with early DSL expansion did not experience differential growth in the relative outcomes 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 technology-skill specificity is very high (i.e., η is low). 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 more suitable for cognitive-intensive jobs. This suggests that cost of skill investment for young workers is moderate (i.e., ψ is positive).

Robustness 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

the table, we obtain uniformly high first-stage SW F-statistics in all specifications.

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 (30) 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.

coefficients are less precise and slightly smaller in magnitude. We also demonstrate that results are similar when controlling for district-generation-year fixed-effects. This is reassuring as this restrictive set of controls absorbs 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 when 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 recent years.

In addition, we consider a sample that includes only 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.

7 A Numerical Illustration of the Theory

We conclude the paper by using the evidence in the preceding section to analyze how economies adjust to cognitive-biased innovations. Our goal is not to provide a full quantitative account of such technological transitions, but rather to numerically illustrate our theoretical insights. In particular, we are interested in giving a sense of how large are the impacts of technology-skill specificity and skill investment cost on the adjustment persistence and the welfare changes following technological shocks. Furthermore, by presenting the full non-linear equilibrium dynamics, the numerical exercise also demonstrates that our theoretical insights are not entirely driven by the first-order approximations.

We map the H technology to cognitive-intensive occupations, and use the empirical impulse responses of Section 6 to parameterize the model. We first externally calibrate the discount rate ρ to match an annual interest rate of 2% and the demand elasticity of substitution to $\theta = 3$. We then select the parameters governing production technologies $(\alpha(i), \sigma(i))$ and the skill distribution dynamics (δ, ψ) to match the estimates in Figures 6 and 7. The decline in the share of the old workers in total employment from 1997 to 2014 implies $\delta = 0.057$, i.e, an expected working life-span of about 18 years after age 35. The small response in the cognitive-intensive employment of old generations yields an η close to zero, and the large young-old gap in the relative employment response implies $\psi = 0.35$. Appendix C.1 describes the matching procedure in detail, along with the model's goodness of fit.

We use the parameterized model to study the consequences of a cognitive-biased innovation that increases the employment share in the cognitive-intensive technology from 20% to 50%.[27] We focus on the impact of the shock on average welfare ($\Delta\bar{U}$) and lifetime welfare inequality ($\Delta\bar{\Omega}$).[28] Appendix C.2 shows the dynamics of the skill distribution and several other labor market outcomes.

Table 3 shows that, for our baseline parameterization, the increase in average welfare across all generations (in consumption equivalent units) is 46% and the increase in lifetime welfare inequality is 39%. These large effects follow from the substantial shock size necessary to induce the reallocation of more than one-fourth of the economy's labor force. The remaining rows of Table 3 compare these figures to those obtained with two calculations that ignore the adjustment 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. As discussed in Section 5, these calculations are equivalent to those that would be obtained by researchers using a reduced-form supply-demand model that ignores changes in supply

[27] 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 cognitive-biased shock that generates convergence in cognitive-intensive employment shares across such countries.

[28] Our analysis specifies the discount rate of social welfare to $r = \rho + \delta$, so that the social discounting of future generations is identical to the discounting of worker's future utility.

Table 3: Changes in Average Welfare and Lifetime Welfare Inequality

	Baseline ($\eta \approx 0, \psi = 0.35$)		Low specificity ($\eta \approx 0.75, \psi = 0.35$)		High cost ($\eta \approx 0, \psi \approx 0.05$)	
	$\Delta\bar{U}$	$\Delta\bar{\Omega}$	$\Delta\bar{U}$	$\Delta\bar{\Omega}$	$\Delta\bar{U}$	$\Delta\bar{\Omega}$
True	46%	39%	44%	29%	35%	66%
Short-run	31%	76%	40%	45%	31%	76%
Long-run	55%	30%	47%	24%	38%	62%
$\frac{\int_0^\infty \hat{q}_t dt}{1/\delta}$	1.6		0.8		0.8	

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 20% to 50% between stationary equilibria. 'True' corresponds to the measures that fully 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.

elasticity over time. 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. The biases arise because of the slow, persistent adjustment in the skill-distribution. For example, as the last line shows, the inequality persistence is 1.6 expected lifetimes. 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.

Finally, the remaining columns of Table 3 consider alternative economies where the adjustment is less persistent, either due to a lower degree of technology-skill specificity (higher η) or a higher skill investment cost (lower ψ). We pick the counterfactual degree of technology-skill specificity to match existing estimates of the dispersion in occupation-specific skills. In particular, we target typical estimates of 0.75 obtained from wage dispersion in a cross-section of workers.[29] In addition, we pick the counterfactual skill investment cost to match the same counterfactual persistence of lifetime inequality (0.8 lifetimes) as in the counterfactual with respect to the degree of technology-skill specificity. Both counterfactual exercises show that the biases from the short- and long-run welfare calculations become much smaller when persistence is lower – particularly so for average welfare. Furthermore, they show the risks of extrapolating from past episodes where the nature of technologies and skills differed. For example, if we used estimates from past episodes where technology-

[29]Estimates in the literature are typically not reported in terms of η. Instead, they report the elasticity of relative occupation-level employment with respect to a change in relative occupation-level wages: $\frac{\Delta \log(e)}{\Delta \log(\omega)}$. Estimates for this elasticity are between 1 and 2 – e.g., see Hsieh et al. (2013). Given that the equilibrium employment share is 0.5 in the long-run of our economy, a reported relative employment elasticity of 1.5 (the mid-point between 1 and 2) implies $\eta = 0.75$.

skill specificity was lower (as in the second column), then we would predict a much smaller average lifetime welfare inequality increase than the true one corresponding to an economy with higher skill specificity (29% versus 39%). Alternatively, if we used past episodes with higher skill-investment costs (as in the third column), we would incorrectly predict much smaller increases in average welfare (35% versus 46%).

8 Conclusions

In this paper, we develop 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, most of the adjustment in economies with high skill-specificity follows from changes in the supply of skills that happen across generations, as opposed to from the reallocation of workers within a generation. We then 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. Parameterizing 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, 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. In particular, when the innovation affects activities with a high degree of technology-skill specificity because such transitions are slower. Third, 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.

Heckman, James J and Bo E Honore. 1990. "The empirical content of the Roy model." *Econometrica* 58 (5):1121–1149.

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-II 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.

McFadden, Daniel et al. 1973. "Conditional logit analysis of qualitative choice behavior." .

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 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) with respect to time, we get that

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

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

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

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

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

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

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

where the last line uses our guess that $l_t \leq l$ for all t.

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 $\hat{l}_t$ shown,

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

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

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

Then, using the indifference condition $\omega_t = \frac{1}{\sigma(l_t)}$ and log-linearizing, we obtain the equation shown in the lemma,

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

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.

Let us guess that the policy functions are given by $\frac{\partial\hat{l}_t}{\partial t} = -\lambda\hat{l}_t$ and $\hat{q}_t = \zeta\hat{l}_t$. Replacing in

the 2X2 dynamic system, we obtain the expressions in the proposition for λ and ζ:

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

Notice that the second equation immediately yields the expression for ζ. To get the expression 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.4 Proof of Corollary 1

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.3 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 relative value of output y_t were already derived in equations A.1 and A.2.

A.5 Proof of Proposition 2

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

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

Furthermore,

$$\Delta log(l_t) = -\eta\Delta log(\omega_t) = -\eta\left(\Delta log(\omega) + \hat{\omega}_0 e^{-\lambda t}\right)$$

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

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

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

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

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

2. Short-run adjustment

$$\begin{aligned}
\Delta log(y_0) &= \left(1 - \frac{\theta - 1}{\theta + \kappa\eta}\right)(\theta - 1)\Delta log(A) \\
\Delta log\left(q_0\right) &= \frac{1}{\theta + \kappa\eta + \frac{\psi}{\rho+\delta}} \frac{\lambda}{\delta} \frac{1}{\rho + \delta}(\theta - 1)\Delta log(A) \\
&= \frac{1}{\theta + \kappa\eta} \frac{1}{\rho + \lambda}(\theta - 1)\Delta log(A) \\
|\Delta log(l_0)| &= \frac{\eta}{\theta + \kappa\eta}\left(\theta - 1\right)\Delta log(A)
\end{aligned}$$

The first and last lines show that $\Delta log(y_0), |\Delta log(l_0)|$ are increasing in η and 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

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

The second line shows that $\int_0^\infty \hat{q}_t dt$ is decreasing in η, since λ is decreasing in η. Furthermore, $\int_0^\infty \hat{q}_t dt$ is increasing in ψ around $\psi = 0$. This is because λ is increasing in ψ, $\lambda = \delta$ when $\psi = 0$, and $\frac{\partial\left(\frac{1}{\lambda}\frac{1}{\theta+\kappa\eta+\frac{\psi}{\rho+\delta}}\right)}{\partial\psi}$ is bounded.

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:

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

Finally, $\int_0^\infty |\hat{l}_t| dt$ is increasing in ψ around $\psi = 0$, since it is proportional to $\int_0^\infty |\hat{y}_t| dt$. However, the derivative with respect to η is ambiguous. This is because the constant of

proportionality $\eta/(\theta-1)$ is increasing in η while $\int_0^\infty |\hat{y}_t|dt$ is decreasing in η.

A.7 Proof of Proposition 5

We have that, because of the envelope theorem,

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

Then,

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

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 equations (26)–(27)

Proof of equation (26). 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

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

By approximating these expressions around l,

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

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,

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

Using the demand expression in (2),

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

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

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

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

which is identical to (26).

Proof of equation (27). 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:

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

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

Using the definition of employment shares for each generation,

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

Thus,

$$\log(\frac{e_t^{young}}{e_{0^-}^{young}}) - \log(\frac{e_t^{old}}{e_{0^-}^{old}}) \approx \frac{1}{(1 - e_{H,\infty}) e_{H,\infty}} \frac{1}{1 - (1 - \tilde{x}_0) e^{-\delta t}} \left(\delta \int_0^t e^{\delta(\tau - t)} \int_{l_t}^1 \left(\tilde{s}_\tau(i) - s_0(i) \right) di d\tau \right) \tag{A.11}$$

We now consider the following approximation:

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

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 1,

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

We now apply this expression into (A.11):

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

Notice that Proposition 2 implies that

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

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

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

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 (27).

Appendix B Empirical Application

B.1 Data Construction

The raw data in the LIAB comes in the form of entire job histories of workers in the sample. Individual entries therefore contain worker information, as well as information on the start and end date of a job spell for that individual, the location (establishment), and characteristics of the job spell. We transform this data into an annual panel dataset following the steps in Card, Heining, and Kline (2013), with minor modifications. These steps are:

- Restrict sample to males in West Germany.
- Restrict ages to 15-64 years at the time of the job spell.
- Restrict sample to full-time job spells only.
- Select job-spell within a calendar year with maximum earnings.
- Deflate earnings using German CPI information from FRED (Series id: DEUCPIALLMINMEI).
- Replace daily wages with Upper Earnings Limits in Card, Heining, and Kline (2013) for daily wages above censor limit.
- Impute district of employment using the district of the establishment if the district of employment is missing.

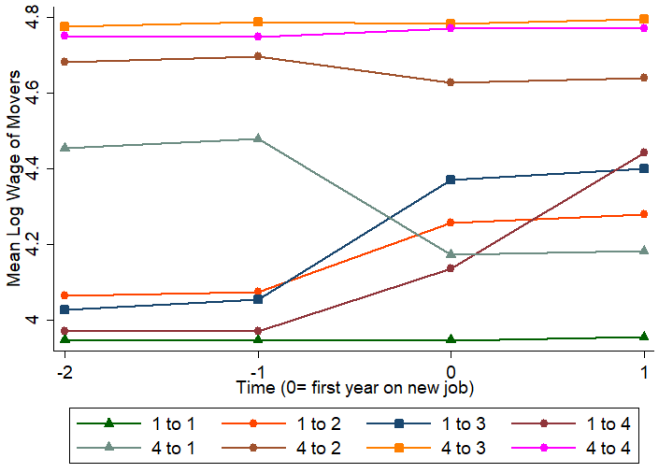
While the years represented in our data and our underlying data sample differ from Card, Heining, and Kline (2013), our panel well represents the data used in that paper. Figure B1 illustrates that the mean wage changes of job movers, classified by the mean log wages of coworkers in their old and new establishments, is similar in our data to the main findings in Card, Heining, and Kline (2013) (their Figure Vb).

We link our LIAB-based worker panel to the DSL access data from Falck, Gold, and Heblich (2014) using the district identifiers in both datasets. These data are discussed in detail in Section 6.3.1. Figure B2 illustrates the spatial variation in our main instrument which underlies the identification in Section 6.

B.2 Cognitive Intensity and Use of New Technologies Across Occupations

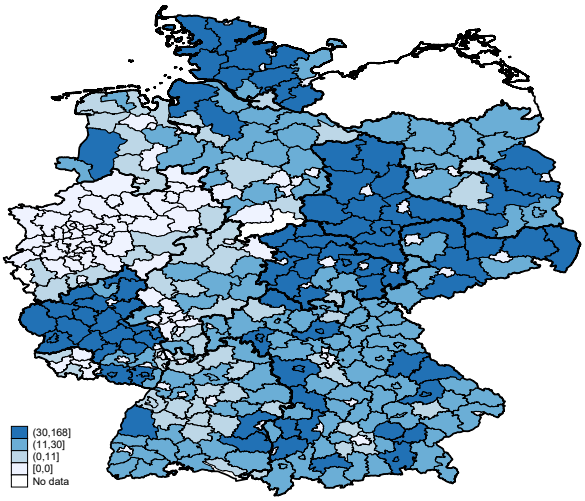
This section analyzes the types of tasks required by cognitive-intensive occupations. Figure B3 reports the correlation between the occupation's intensity in cognitive skills and the share of individuals in that occupation reporting they intensely perform each of the listed tasks. The top tasks performed 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 B3 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).

Figure B1



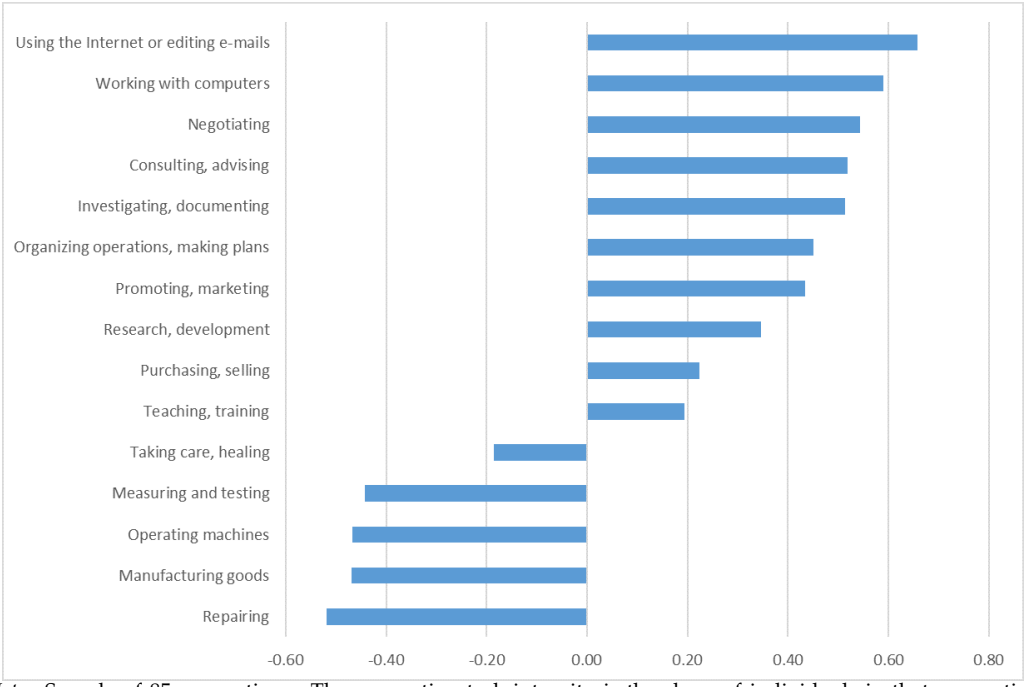
Note. Figure illustrates the mean wage changes for job movers from the fourth and first quartile of establishments in all quartiles of establishments. Movers are defined as workers who move jobs from a job they held for two years before moving, and stay in the new job for two years after moving. Quartiles are defined by the mean log wages of coworkers in the old and new establishments. The sample period is 2002-2009. $trans_t = 3$ is the year of moving.

Figure B2



Note. Figure illustrates the mean number of municipalities across districts in Germany that did not have access to an MDF within the 4000m radius, as described in Section 6.3.1.

Figure B3: 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).

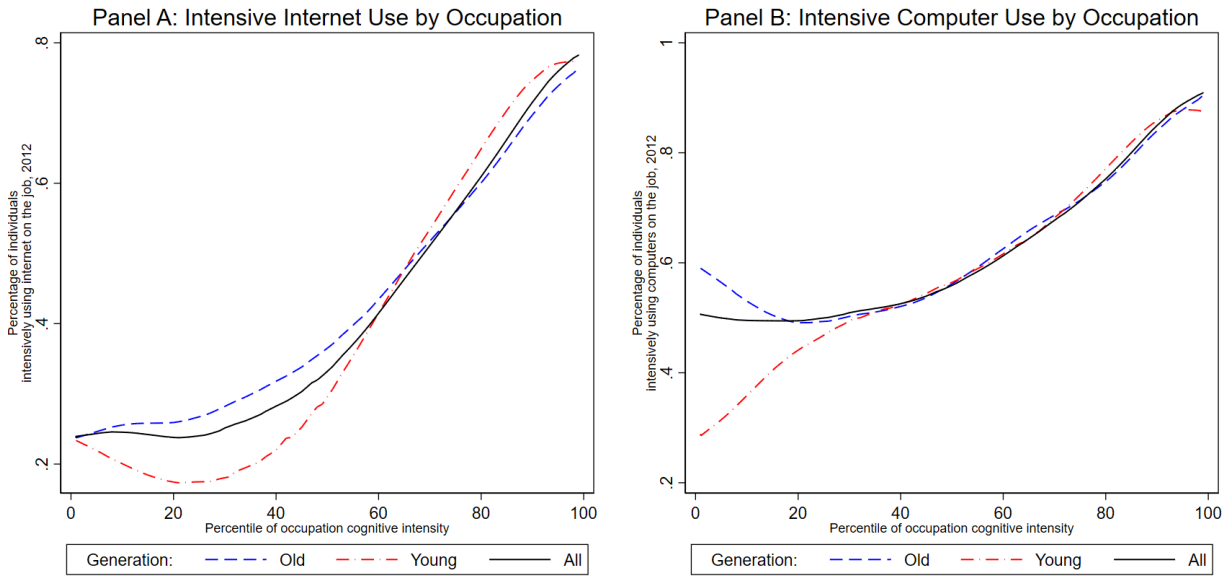
We then investigate whether these new technologies affected worker generations differently conditional on their occupation. We consider two generations: a young generation aged less than 40 years and an old generation aged more than 40 years.[30] Figure B4 shows that, while internet and computer usage 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.3 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 B5 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

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

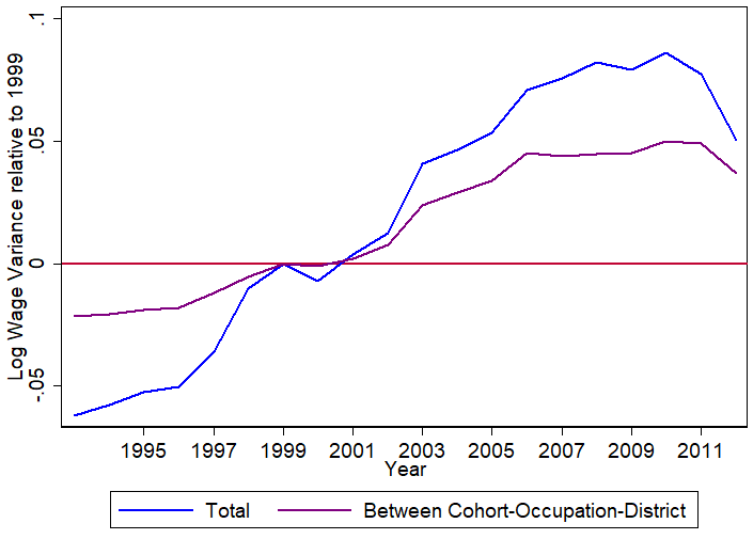
Figure B4: Internet and Computer Usage by Occupation: Within- and Between-Generation



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.

inequality rise.[31]

Figure B5: 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.

[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.

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.

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.

B.4 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 (28) 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 (29)-(30) 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 (30) 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	Between
	(1)	(2)	(3)	(4)
Panel A: Alternative control set, 1995-2014				
Controls for immigration and trade	1.426***	1.807***	0.841***	0.966***
	(0.261)	(0.279)	(0.252)	(0.376)
Panel B: Alternative sample definition, 1995-2014				
Native-born Males Only	1.396***	1.807***	0.778***	1.029***
	(0.226)	(0.235)	(0.231)	(0.340)
Panel C: Alternative generation definition, 1995-2014				
Young: Born after 1965	1.488***	2.137***	0.857***	1.280***
	(0.225)	(0.299)	(0.246)	(0.387)
Young: Born after 1955	1.488***	1.639***	0 .967***	0.671*
	(0.225)	(0.268)	(0.290)	(0.395)
Young: Aged Below 40 in each year	1.488***	1.748***	0.773***	0.975**
	(0.225)	(0.294)	(0.246)	(0.383)

Note. Sample of 120 occupations, sample periods as defined in the table. Columns (1)–(3) report the estimated coefficient on the occupation's cognitive intensity in equation (28). Column (4) reports the difference between the coefficients in columns (3) and (2). 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 (29). Columns (2)-(3) report the estimated coefficients on interaction between the occupation's cognitive intensity, generation dummies and district DSL access in equation (30). 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 (29). Columns (2)-(3) report the estimated coefficients on interaction between the occupation's cognitive intensity, generation dummies and district DSL access in equation (30). 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 (30) 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 (30).

Appendix C Numerical Analysis

C.1 Parameterization by impulse response matching

This appendix discusses in detail the parameterization of the model. As in Section 6, 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.[32] 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 6 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 6.1, 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 parametrization procedure next, along with an analysis of the model fit.[33]

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. We also select the distribution of innate ability to normalize the initial skill distribution to be uniform: $s_0(i) \equiv 1$.

[32] 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.

[33] 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 1) and η matches the short-run employment change.

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:

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

where $\hat{\beta}_t^y$ is the estimated coefficient of (29) reported in Panel B of Figure 6.

The minimization problem in (C.1) yields $\hat{\lambda} = 0.135$. Figure B7 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

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

Using expression (27), 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,

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

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 (C.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:

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

where $\hat{\beta}_t^y$ is the estimated coefficient of (29) 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 (30) reported in Panel B of Figure 7.

The minimization problem in (C.4) yields $\hat{\psi} = 0.345$. Figure B7 shows the fit of the calibrated model.

Short-run skill supply elasticity: η. The combination of (26) and (C.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 (C.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

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

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:

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

where $\hat{\beta}_t^y$ is the estimated coefficient of (29) reported in Panel B of Figure 6, and $\hat{\beta}_t^{old}$ is the employment response for old workers obtained from the estimation of (30) reported in Panel A of Figure 7.

The negative point estimates reported in Panel A of Figure 7 imply that the minimization problem in (C.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.

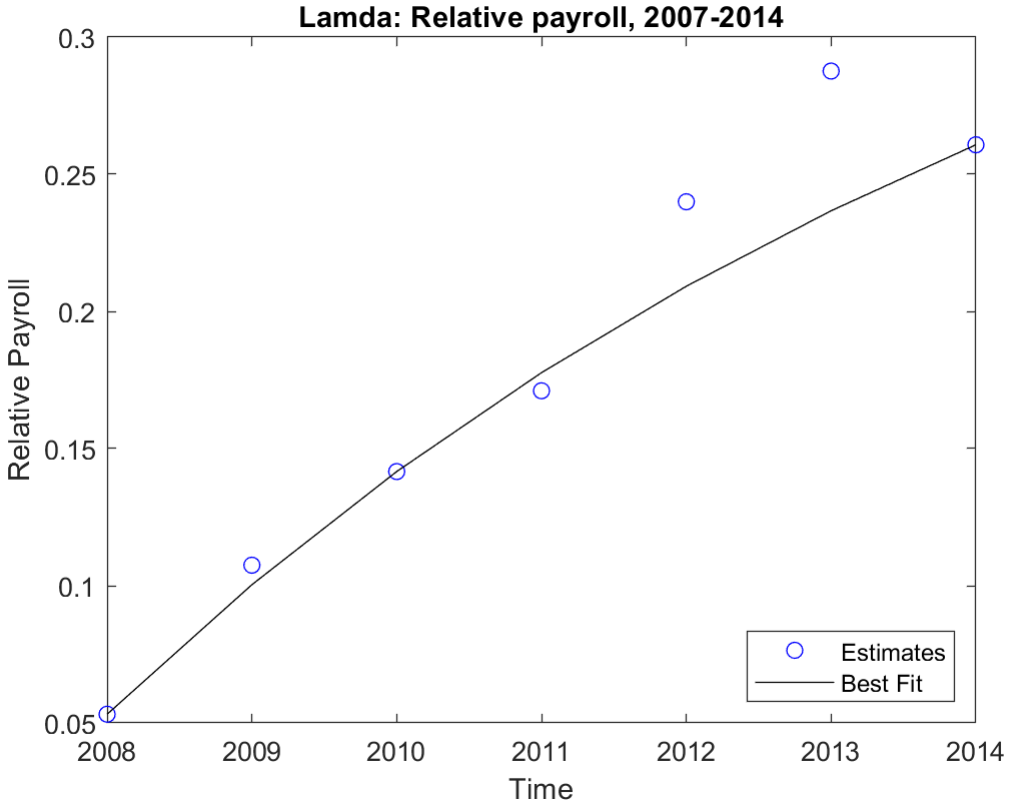


Figure B6: Calibration of λ

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 (C.1).

C.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

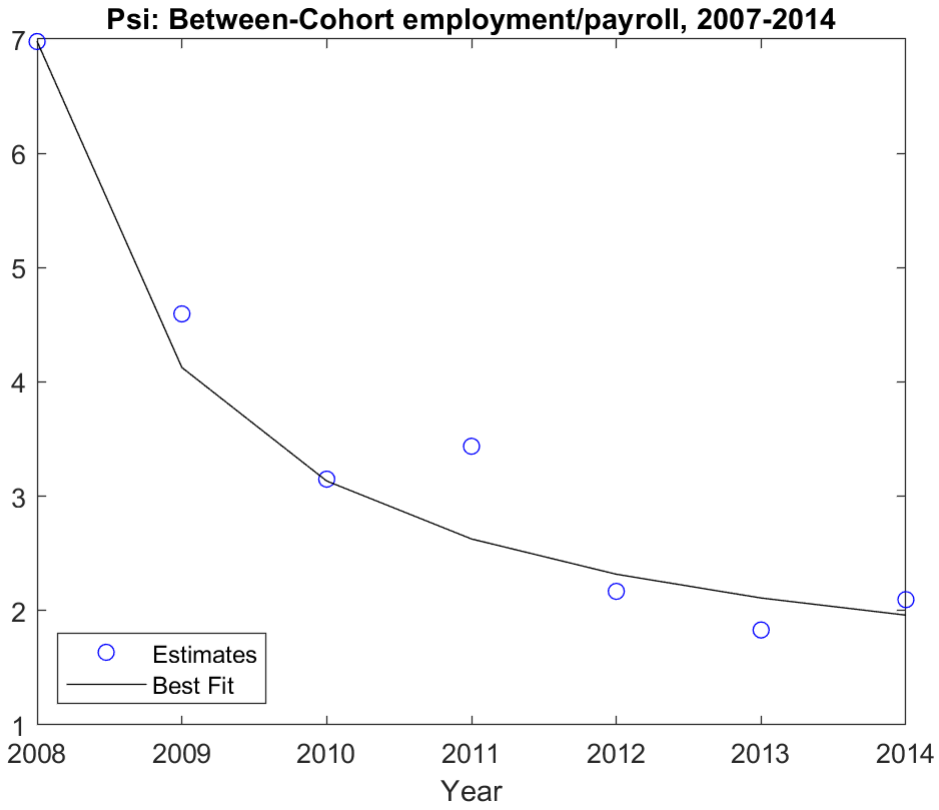


Figure B7: Calibration of ψ

Note. Blue dots represent the point estimates of $\frac{\hat{\beta}_t^{young} - \hat{\beta}_t^{old}}{\hat{\beta}_t^{\forall}}$ 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 (C.4).

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.

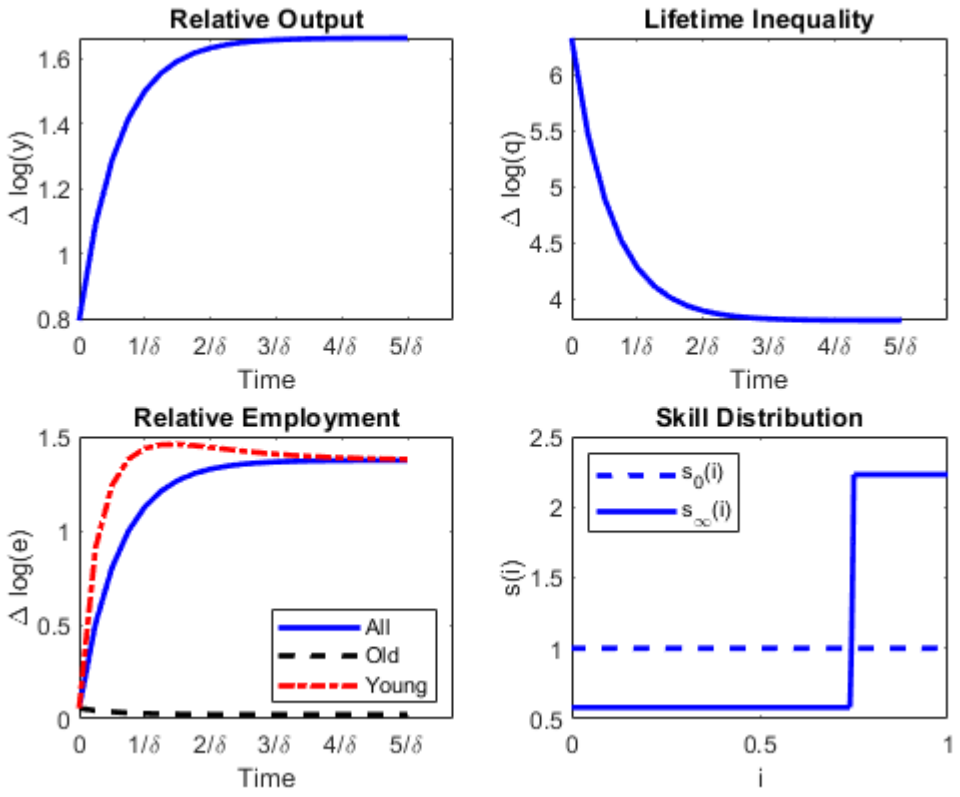
Figure B8 presents the predicted impulse response functions of labor market outcomes. Consider first response at impact. Given that our theory abstracts from several additional sources of dynamics, it would be wrong to interpret the impact adjustment as happening instantaneously in reality. We view this 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, $1/\delta \approx 18yrs$). 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.

Figure B8: 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 Appendix C.1.

Appendix D Additional Results

D.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.

D.2 Welfare Consequences of Adjustment Across Generations

This section investigates 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 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 $\psi^2 = 0$.

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 D.2.1. ■

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 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 4, 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).

D.2.1 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

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

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,

$$\Delta\bar{\Omega} = \Delta \log(q_T) \left(\frac{1 + \frac{\lambda - \delta}{\delta} \frac{r}{r+\lambda}}{1 + \frac{\lambda - \delta}{\delta} e^{-\lambda T}} \right). \tag{D.2}$$

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

$$\Delta\bar{\Omega}^P = \Delta \log(q_T) \tag{D.3}$$

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 (D.1), 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 (D.2) and (D.3),

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

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 (D.1), we must set .

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

Expressions in (D.2) and (D.3) 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 (D.1), 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

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

Appendix E Extensions

TBA