NBER WORKING PAPER SERIES

TECHNOLOGICAL TRANSITIONS WITH SKILL HETEROGENEITY ACROSS GENERATIONS

Rodrigo Adão
Martin Beraja
Nitya Pandalai-Nayar

Working Paper 26625
http://www.nber.org/papers/w26625

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2020

We are grateful to Masao Fukui, Agustin Gutierrez and Michelle Lam for excellent research assistance. We thank Christopher Tonetti for his superb discussion as well as all other participants at the NBER's EF&G Fall meeting. For their valuable suggestions, we thank Daron Acemoglu, David Autor, Fernando Alvarez, Arnaud Costinot, Ariel Burstein, V.V. Chari, Dave Donaldson, Oded Galor, Erik Hurst, Greg Kaplan, Pablo Kurlat, Frank Levy, Simon Mongey, Fabrizio Perri, Venky Venkateswaran, Gianluca Violante, and Fabrizio Zilibotti. We also thank seminar participants at Dartmouth, Duke, MIT, Northwestern, NYU, Notre-Dame, Philly Fed, 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. This paper previously circulated under the title "Skill-Biased Technological Transitions." The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Technological Transitions with Skill Heterogeneity Across Generations
Rodrigo Adão, Martin Beraja, and Nitya Pandalai-Nayar
NBER Working Paper No. 26625
January 2020
JEL No. C0,E0,F0,J0,O0

ABSTRACT

Why are some technological transitions particularly unequal and slow to play out? We develop a theory to study transitions after technological innovations driven by worker reallocation within a generation and changes in the skill distribution across generations. The economy's transitional dynamics have a representation as a q-theory of skill investment. We exploit this in two ways. First, to show that technology-skill specificity and the cost of skill investment determine how unequal and slow transitions are by affecting the two adjustment margins in the theory. Second, to connect these determinants to measurable, short-horizon changes in labor market outcomes within and between generations. We then empirically analyze the adjustment to recent cognitive-biased innovations in developed economies. Strong responses of cognitive-intensive employment for young but not old generations suggest that cognitive-skill specificity is high and that the supply of cognitive skills is more elastic for younger generations. This evidence indicates that cognitive-biased transitions are slow and unequal because they are mainly driven by changes in the skill distribution across generations. Naively extrapolating from observed changes at short horizons leads to too pessimistic views about their welfare and distributional implications.

Rodrigo Adão
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
rodrigo.adao@chicagobooth.edu

Martin Beraja
Department of Economics
MIT
50 Memorial Drive
Cambridge, MA 02142
and NBER
martinberaja@gmail.com

Nitya Pandalai-Nayar
Department of Economics
University of Texas at Austin
3.142 BRB, 2225 Speedway
Austin, TX 78713
and NBER
npnayar@utexas.edu

An online appendix is available at https://www.nber.org/data-appendix/w26625/

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 (for a review, see Acemoglu and Autor, 2011). If an economy's adjustment margins vary as horizons lengthen, then focusing on short or long horizons alone risks missing the overall impact of technological innovations on labor markets as well as their average and distributional welfare consequences. Such concerns are particularly important when the adjustment is slow and takes many generations. How then do economies adjust to technological innovations over different horizons? Why are some technological transitions particularly unequal and slow to play out?

We begin by developing a theory to study technological transitions driven by both worker reallocation within a generation and changes in the distribution of skills across generations. The transitional dynamics of this economy can be represented as a q-theory of skill investment. We use this in two ways. First, to characterize how the nature of the technological innovation and associated skills determine the importance of the two theoretical adjustment margins and, as a result, how slow and unequal technological transitions are. Second, to connect these determinants to measurable, short-horizon changes in labor market outcomes within and between generations. Three related pieces of evidence for developed countries suggest that transitions following cognitive-biased innovations are slow and unequal because they are mostly driven by skill changes across generations – rather than worker reallocation within a generation – due to the high specificity of cognitive skills. Naively extrapolating from observed changes at short horizons therefore misses much of the adjustment at longer horizons, thus underestimating the average welfare gains and overestimating the distributional consequences of cognitive-biased innovations.

The theory has four 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 the relative technology-specific wage at a point in time, there is a threshold determining which skill types self-select into each of the two technologies. *Technology-skill specificity* – i.e., the change in productivity when skill types are assigned to different technologies – then determines how sensitive the assignment threshold is to changes in relative technology-specific wages. Third, the output of the two technologies is combined to produce a final consumption good, as in Katz and Murphy (1992), Ngai and Pissarides (2007), and Buera et al. (2011). Fourth, given future relative technology-specific wages, workers make a costly investment upon entering the labor market that determines their skill type for their lifetime, similar to Chari and Hopenhayn (1991), Caselli (1999) and Galor and Moav (2002). The *cost of skill investment* for entering workers then determines how different the skill distribution is across generations following changes in future relative wages.[1]

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

Our first result reduces the dimensionality of this fixed-point problem. It establishes that the approximate equilibrium of this economy can be represented as a q-theory of skill investment. The path for the skill distribution is only a function of two variables at each point in time: the present-discounted value of the log-relative technology-specific wage (q) and the threshold determining the assignment of skills to technologies (which plays the role of the pre-determined variable). We show that a simple system of linear differential equations characterizes the equilibrium dynamics of these two variables. Thus, we solve for the equilibrium dynamics by keeping track of these variables and not the skill distribution itself. Our approach is reminiscent of those in Perla and Tonetti (2014) and the special case with linear objectives in Lucas and Moll (2014) which characterize the dynamics of a distribution by tracking the evolution of a threshold.

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 technol-

[1]We formalize this directed skill investment problem by allowing workers to enter a skill lottery whose cost is proportional to the relative entropy between the chosen lottery and a reference distribution. The optimal lottery is the multinomial logit function over a continuum, similar to standard discrete-choice models.

ogy. On impact, marginal skill types within each generation reallocate into that technology. The increase in current and future relative wages leads younger entering generations to invest in those skills that are more complementary to the improved technology. Along the transition, as younger generations replace older ones, q falls and relative output increases because the economy's skill distribution tilts towards skills more complementary to the improved technology. To evaluate how slow the transition is, we define the discounted cumulative impulse response (DCIR). For old generations born before the innovation, the DCIR measures how different the adjustment they expect to see during their lifetime is compared to the overall (long-run) adjustment. We say that the adjustment is slower whenever they expect to miss more of the overall adjustment (i.e., the DCIR is larger). Crucially, we also show that the DCIR of q is a central determinant of the average and distributional welfare consequences of new technologies.

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 than longer horizons, a form of Samuelson's LeChatelier principle. Our micro-foundation points to two types of risks associated with ignoring dynamics induced by changes in the skill distribution across generations. The first arises when extrapolating from observed responses in the economy that span much less than a generation. Such extrapolation will overestimate inequality changes and underestimate average welfare gains. The second arises when extrapolating from past technological transitions to different contexts: a type of threat to external validity. This leads to biased predictions about the economy's dynamic adjustment whenever the nature of technology and skills, or the underlying flexibility of skill investment, significantly differs across episodes.

Our third result presents comparative static exercises that speak to why some technological transitions are particularly unequal and slow to play out. As such, these exercises help interpret differences between past or future transitions where the nature of technological innovations and associated skills differ. First, we show that an economy where technology-skill specificity is higher has a slower, more back-loaded adjustment path to the new long-run equilibrium. The q-theory analogy again delivers the intuition for this result. When technology-skill specificity is higher, there is less worker reallocation across technologies in the short-run and, therefore, the increase in lifetime inequality q is larger. This strengthens the incentives of young entering generations to invest in those skills that are more complementary to the improved technology. As a result, the adjustment is slower because transitional dynamics become more important as larger changes in the skill distribution take place. Second, we show that a lower cost of skill investment for young generations makes the adjustment slower as well, both directly and by amplifying the effects of technology-skill specificity. In both cases, the more relevant margin of adjustment is not the reallocation of workers within a generation but the changes in the skill distribution across generations.

Our fourth result connects the degree of technology-skill specificity and the cost of skill investment to observable changes in labor market outcomes within and between generations. In particular, we focus on short-run implications that can be credibly measured in most datasets. Our measurement insight is that, in the short-run, economies with higher technology-skill specificity are associated with weaker within-generation changes in the relative employment of older workers across occupations (or sectors), but stronger between-generation differences in the relative employment of younger and older workers. In contrast, a lower cost of skill investment for entering generations is also associated with larger between-generation differences in relative employment, but has no effect on the responses for older workers. Such generation-specific changes are common in empirical analysis of how economies adjust to different types of shocks.[2] By connecting them to structural parameters, our theory shows how these measurable moments in the short-run are informative about the economy's transitional dynamics and, consequently, how unequal and slow the adjustment will be.

In the second part of the paper, we explore our theory's observable predictions to provide three pieces of evidence indicating that technology-skill specificity and changes in the skill distribution across generations are relevant to understand how developed economies adjusted to recent cognitive-biased technological innovations. First, we analyze employment trends in nine broad occupation groups in eighteen developed countries. We document that, in all countries, employment growth in the three most cognitive-intensive occupations was stronger for younger workers than for older workers. Second, we use microdata to provide a more detailed investigation of these responses in Germany. Controlling for a number of confounding factors, we show that employment and payroll grew more in occupations more intensive in cognitive tasks. This effect is strong for younger generations but weak for older generations. We also explore the unique

[2]For example, see Kim and Topel (1995), Card and Lemieux (2001), Autor and Dorn (2009)

large-scale German training program to document higher growth in the number of trainees in more cognitive intensive occupations, providing direct evidence that younger generations invest more in cognitive skills. Finally, following Falck et al. (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 impulse response functions that show an increase in the relative employment and payroll of more cognitive-intensive occupations starting in 2005. The estimates are again different for older and younger generations. The impact on relative employment is small and nonsignificant for older generations, but it is positive and statistically significant for younger generations.

In sum, this evidence suggests that cognitive-skill specificity is high and that the supply of cognitive skills is elastic at longer horizons. Parameterizing our model to match the empirical impulse responses for Germany, we find that these two features make cognitive-biased transitions particularly unequal and slow. As a result, we quantify that, compared to naive extrapolations from observed changes on impact, the true average welfare (lifetime welfare inequality) increase across generations is about 50 percent higher (lower) following a relatively large cognitive-biased innovation. Had technology-skill specificity been lower, such innovation would have led to a faster and less unequal transition instead, featuring smaller between-generation differences in occupation composition changes.[3]

Related literature. Our paper is related to several strands of the literature. A long literature has analyzed structural transformation in the form of long-run reallocation driven changes in relative demand across sectors – e.g., Ngai and Pissarides (2007), Rogerson (2008), Buera et al. (2011), and Buera and Kaboski (2012). Recently, Lagakos and Waugh (2013) show that endogenous skill-sector sorting affects the process of structural transformation. Moreover, a number of papers have also emphasized between-generation differences in employment reallocation across sectors following long-run changes in sectoral productivity growth (Kim and Topel, 1995 and Hobijn et al., 2019) or schooling (Porzio and Santangelo, 2019). We make two contributions to this literature. First, we provide a tractable theory to analyze how skill heterogeneity within and across generations shapes the transitional dynamics induced by technological innovations. This allows us to

[3]Consistent with the idea that the nature of technological innovations and skills may be different across episodes, we also document that, compared to most recent years, these between-generation differences were smaller in the United States and Germany in the 1960s and 1970s.

characterize how fast the transition is, a focus we share with Gabaix et al. (2016). We use this characterization to point out which features of the economy lead to slow adjustment dynamics and large biases from welfare calculations that ignore them. Second, we estimate impulse response functions to a technological innovation in Germany and show how they discipline our theoretical mechanisms.

The only source of dynamics in our theory is the endogenous change in the distribution of skills across generations. This mechanism is consistent with recent evidence documenting the impact of labor demand shocks on young individuals' decisions of educational attainment (Atkin, 2016 and Charles et al., 2018) and field of study (Abramitzky et al., 2019, Ghose, 2019). We add to this literature by documenting that cognitive-biased innovations differentially affect young employment and training in cognitive-intensive occupations in Germany. Several papers have proposed alternative sources of dynamics to study technological transitions, including sluggish labor mobility across sectors (Matsuyama, 1992), technology diffusion across firms (Atkeson and Kehoe, 2007), firm-level investment in R&D (Atkeson et al., 2018), endogenous creation of new tasks for labor in production (Acemoglu and Restrepo, 2018), mobility costs of heterogeneous workers (Dvorkin and Monge-Naranjo, 2019), and rising wealth inequality via permanent changes in the returns to wealth following increases in automation (Moll et al., 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 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), Hsieh et al. (2013), Burstein et al. (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). 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 skill investment costs for younger generations generate slower adjustment following skill-biased innovations. We then link the two adjustment margins in our theory to observable responses of labor market outcomes within and between 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.

Our paper is also related to the literature that has estimated the distributional consequences of shocks to the demand and supply of skills – for a review, see Acemoglu and Autor (2011). Our empirical analysis follows the literature studying the impact of new technologies on the demand of skills across occupations with different task intensity – e.g., Autor et al. (2003) and Acemoglu and Restrepo (2017). As Akerman et al. (2015), we estimate the labor market consequences of broadband internet adoption. While they focus on the impact on educational composition of employment in Norwegian firms, we estimate its effect on the occupation composition of employment across German local labor markets. Similar to Autor and Dorn (2009), we find that the impact of new technologies differs for younger and older workers. Relative to this literature, our results indicate that reduced-form evidence estimated at short horizons is informative about structural parameters governing the adjustment to new technologies, but they also caution against directly extrapolating from it when technological transitions are slow.[4]

2 A Model of Skilled-biased Technological Transitions

We consider a closed economy in continuous time. There is a single final good 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.

[4] A full account of inequality trends would require analyzing not only the impact of shocks to skill demand but also how the economy adjusts to shocks to the supply of skills, as in Katz and Murphy (1992) and Goldin and Katz (2009). For instance, Card and Lemieux (2001) show that generation-specific skill supply shocks are important determinants of inequality trends. Similarly, we incorporate skill differences across generations. However, while it would be possible to study the consequences of skill supply shocks in our theory, we focus on the consequences of shocks to skill demand when skill supply responds across generations.

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

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

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 input prices, the cost minimization problem of firms producing the final good implies that the relative spending on the high-tech input is

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

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

In a competitive environment with zero profits, the final good price 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 her differential productivity in high-tech production. The production functions of L and H are respectively

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

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

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

We assume a competitive labor market with zero profit in production. In equilibrium, the wage of skill type i with the H and L technologies are respectively

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

As in Roy (1951), workers self-select across technologies to maximize labor income. Thus, the labor income of a worker with skill type i is

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

The technology-skill assignment in equation (7) 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 H production, as well as the exogenous function $\sigma(i)$ defining the differential productivity of type i in H production. Without loss of generality, assume that $\sigma(i)$ is increasing: we order types such that higher i types have higher relative productivity in high-tech production.

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.[5]

Skill investment. We now endogenize the distribution of skills by allowing workers to direct their skill investment decisions to target particular skill types. After the equilibrium definition, we discuss the interpretation of our main assumptions and why they yield a tractable theory of technological transitions.

We consider an overlapping generations setting in which the birth and death of workers follow a Poisson process with rate δ. At each point in time, workers use their labor earnings to buy the final good. Utility from consumption is the present value of the log-utility flow discounted at rate ρ. For type i born at time t, expected utility from consumption is then

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

Crucially, we next let workers invest in 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 during their lifetime.[6]

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 $\bar{s}_t(i)$, so

[5] Note that changes in ω_t are not identical to changes in the relative labor income of H employees because of endogenous changes in the "selection" of skill types in H implied by (7) (Heckman and Honore, 1990).

[6] Appendix B includes extensions where (i) workers can re-optimize their skills after the arrival of a new technology, and (ii) the innate ability density of type i is increasing in the density of workers with that type in the economy – a form of "learning from others" externality.

that workers of the cohort born at time t solve the skill investment problem:

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

The positive 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. Because only new generations choose skill lotteries, 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 suffices that relative demand and supply of the H good are equal:

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

where y_t is given by (2) and x_t is the ratio of H to L production given by (4)–(5).

Definition 1 (Competitive Equilibrium) *Given an initial skill distribution $s_0(i)$ and exogenous paths for $\{A_t, \bar{s}_t(i)\}_{t\geq 0}$, 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 wage and final price index $\{\omega_t, P_t\}_{t\geq 0}$, such that*

1. *Given $\{\omega_t\}_{t\geq 0}$, $\{G_t(i), \tilde{s}_t(i)\}_{t\geq 0}$ are determined by (7) and (9).*
2. *Given $s_0(i)$ and $\{\tilde{s}_t(i)\}_{t\geq 0}$, $\{s_t(i)\}_{t\geq 0}$ is determined by (10).*
3. *For all $t \geq 0$, the market clearing condition (11) is satisfied and P_t is given by (3).*

Discussion. A number of comments on the assumptions and their economic interpretation are in order. There are admittedly four strong assumptions that we make for simplicity and tractability.

The first is that we consider a continuum of skills. As discussed below, this implies that changes in the technology-skill assignment are smooth along the transition because any relative wage change triggers the reallocation of a positive mass of skill types. Moreover, the Roy-like skill heterogeneity yields equilibrium responses to technological innovations that do not arise in the canonical model with skills specified with observable attributes (Acemoglu and Autor, 2011). This allows studying how technology-skill specificity affects the economy's adjustment.

Regarding a skill-type's economic interpretation, Section 3.1 of the Online Additional Material provides a microfoundation of (4)–(5) where production combines individual-level output of each worker's "cognitive" and "non-cognitive" task input. A type determines the differential ability to perform cognitive tasks.

Second, we assume that $\bar{s}_t(i)$ is exogenous and only new generations invest in skills. These imply that the flow of new workers to a particular point in the skill distribution is independent of the current skill distribution (see equation (10)), allowing us to characterize the skill distribution's law of motion and the equilibrium transitional dynamics. This independence arises because skill investment decisions are independent from a worker's current skill type. By relaxing both assumptions in Appendix B, we show that our main results do not depend on either the cost of skill investment being infinite for older generations, nor on the skill investment technology being independent of the economy's skill distribution.

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

Third, we assume that skill investments yield an uncertain outcome in the form of the skill lottery $\tilde{s}_t(i)$. We make this assumption only for tractability. Different from theories of uni-dimensional human capital investment, ex-ante identical workers in our theory can direct their investments to target specific skill types. Yet, mathematically, this *directed skill investment* problem is in principle substantially more complex. As we will see below, this assumption delivers a tractable problem with a non-degenerate skill distribution as a solution. What is important for our results though is not that workers are ex-ante identical, but (again) the independence of the skill investment from a worker's current skill type.[7]

[7] It is easy to extend our model to introduce worker-groups that have ex-ante different observed attributes that only affect $\bar{s}(i)$ and ψ. The overall skill lottery would then be the average of lotteries across worker-

Our preferred interpretation for the uncertainty of the skill type realization is that individuals with different unobservables may have heterogeneous returns to education and on-the-job training. This is in fact consistent with the evidence in Carneiro et al. (2011). Our model treats this heterogeneity in unobservables through the uncertainty of the type realization.

Finally, we impose that the cost of a skill lottery takes the form of the entropy function in (9). Entropy cost functions have a long tradition in macroeconomics (Sims, 2003, Hansen and Sargent, 2008).[8] As discussed later, this function implies that the optimal skill lottery is a multinomial logit over a continuum.

One interpretation of this environment then follows from its discrete-choice analog where a worker's innate ability to acquire skills associated with skill-types follows a Type 1 extreme-value distribution (e.g., McFadden, 1973). An alternative interpretation follows from Matějka and McKay (2015). They show that a multinomial logit choice structure arises when individuals choose actions with imperfect information about their payoffs while paying an entropy cost for signal acquisition. In our theory though, having a continuum of skill types is useful when combined with continuous time because it implies that the dynamic adjustment of all outcomes is smooth along the equilibrium path.

2.1 Static and dynamic equilibrium conditions

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 types receive higher relative earnings in H 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 L with labor income of $w_t(i) = \alpha(i)$. Worker types $i > l_t$ are employed in H with labor income of $w_t(i) = \omega_t \sigma(i) \alpha(i)$. The threshold is determined by the indifference condition,*

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

groups. However, this independence would be violated if the type of a worker affects the relative cost of particular lotteries. This would be the case, for example, if there was inter-generation transmission of skills, or skill acquisition had monetary costs in an environment with credit frictions.

[8] However, the entropy-based cost is not crucial for our main results, since in later sections we take a log-linear approximation around the stationary equilibrium. What matters is the curvature of the distance metric around the stationary equilibrium, similar to investment problems with a convex cost of adjustment.

Lemma 1 links the relative wage ω_t to the allocation of skill types across technologies. Condition (12) is central to understand the impact of technological shocks on the allocation of workers across technologies. The slope of $\sigma(.)$ at the threshold determines the strength of *comparative advantage* in H of skill types slightly below l_t compared to skill type l_t. Thus, the inverse elasticity of $\sigma(i)$ controls the subset of skill types that reallocate across technologies in response to changes in the relative wage. Formally, (12) implies

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

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

We say that *technology-skill specificity* is higher when η is lower because the productivity of skills associated with higher i types decreases more when deployed to the L-technology rather than to the H-technology. As a result, a lower η implies that the induced worker reallocation following relative wage changes is smaller in the short-run (when the skill distribution is given).

The technology-skill assignment in Lemma 1 together with equations (4)–(5) imply that the relative supply of high-tech production 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 uniquely determined by market clearing in (11).

Lemma 2 (Equilibrium Threshold) *Given $s_t(i)$ and A_t, there is a unique equilibrium threshold l_t that guarantees goods market clearing,*

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

Proof. See Appendix A.1. ■

Dynamic equilibrium conditions. We now turn to the entrant's forward-looking problem of choosing their skill lottery $\tilde{s}_t(i)$ given the path of the relative wage $\{\omega_s\}_{s>t}$. The following lemma shows that the solution to the problem in (9) takes the form of a multinomial logit function over the continuum of types. In particular, the investment on high-i types is a function of the present value of the relative wage in high-tech production as captured by $Q_t(i)$.

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. ■

Note also that 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 i.

In this case, the skill distribution is a constant-elasticity function of relative income across types, where the elasticity is ψ. 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 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 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 complex infinite-dimensional fixed-point problem. To see this, consider a conjectured path for the relative wage $\{\omega_t\}_{t\geq 0}$. This path determines the skill investment decisions of new generations in (15) and, as such, the path for the skill distribution $\{s_t(i)\}_{t\geq 0}$ from (10) given $s_0(i)$. The relative wage path also determines the assignment threshold path ($\{l_t\}_{t\geq 0}$) from the indifference condition (12). Taken together, the skill distribution and the assignment threshold determine the relative supply of the high-tech input

$(\{x_t\}_{t\geq 0})$ in (13). In an equilibrium, the relative supply of the high-tech input needs to be equal to its relative demand at the conjectured path for the relative wage – 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 that of a q-theory of skill investment, where $log(q)$ refers to the present discounted value of the log-relative wage or, as we call it from now on, lifetime inequality:

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

Specifically, we show that one does not need to keep track of the whole skill distribution to solve for the approximate equilibrium path of q_t and the assignment threshold l_t. The approximate equilibrium dynamics of these two variables are fully characterized by a simple system of linear differential equations. Letting " ˆ " denote variables in log-deviations from the stationary equilibrium, the following theorem presents the system of differential equations that, given $\hat{l}_0$, determines the equilibrium path of $\{\hat{q}_t, \hat{l}_t\}_t$ when $\{A_t, \bar{s}_t(i)\}_{t\geq 0}$ are constant over time.[9] It then characterizes the skill distribution, skill lottery, and relative output.

Theorem 1 (q-theory of skill investment) *Suppose $\{A_t, \bar{s}_t(i)\}_{t\geq 0}$ is 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 κ is a positive constant.

2. *The equilibrium $\{\hat{q}_t, \hat{l}_t\}_{t\geq 0}$ is saddle-path stable with rate of convergence λ.*

3. *The equilibrium dynamics of the skill distribution $\hat{s}_t(i)$, the optimal lottery $\hat{\hat{s}}_t(i)$, and the value of relative high-tech output $\hat{y}_t$ are determined by $\{\hat{q}_t, \hat{l}_t\}_{t\geq 0}$ and $\hat{s}_0(i)$.*

Proof. See Appendix A.3. ■

The first part of the theorem presents a system that is a rather standard one in macroeconomics, with one control and one predetermined variable. The system

[9] The initial l_0 is determined by the initial skill distribution $s_0(i)$ from the static equilibrium condition (14).

is in fact mathematically isomorphic to the q-theory of capital investment. In our model, $\hat{q}_t$ is the present discounted value of the log-relative wage in high-tech production, representing the shadow price of the human capital "asset" associated with having one additional unit of the high-tech good. Whenever this price is higher, the incentives to invest in high-i skills are stronger. 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 controlled by η) and substitutability between goods (as measured by θ).

The second part shows that (locally) the equilibrium exists and is unique—a consequence of saddle-path stability. The proof further shows that, given an initial condition $\hat{l}_0$, both $\hat{l}_t$ and $\hat{q}_t$ converge at a constant rate to the stationary equilibrium, where $-\lambda$ is the negative eigenvalue of the system of differential equations.

The last part of the theorem links the equilibrium path of the optimal skill lottery, the overall skill distribution and the relative value of output to the joint dynamics of $\{\hat{q}_t, \hat{l}_t\}$. The proof shows that the change in the optimal skill lottery along the transition depends centrally on the evolution of $\hat{q}_t$. Formally,

$$\hat{\hat{s}}_t(i) = \left(\mathbb{I}_{i>l} - \int_l^1 s(i) di \right) \psi \hat{q}_t + o_t(i), \tag{19}$$

where $o_t(i)$ is such that $\int s(i) o_t(i) di = 0$. The parameter ψ crucially shapes the extent to which a generation's skill distribution tilts towards skills associated with higher i types when they face a higher q_t at birth. The overall skill distribution is then simply a population-weighted average of the skill distributions of each generation. Finally, relative output is driven by changes in the relative wage, with $1/\eta$ controlling how the latter responds to threshold changes.

Theorem 1 reduces the dimensionality of the equilibrium's fixed-point problem. It characterizes $\{\hat{q}_t, \hat{l}_t\}_{t\geq 0}$ without tracking the dynamics of the skill distribution. This is possible for three reasons. First, the dynamics of $s_t(i)$ only depend on $\log(Q_t(i))$ via the optimal skill lottery. Yet, $\log(q_t)$ suffices to determine the value of most skill types in the investment decision – as opposed to the full path of ω_t in $\log(Q_t(i))$ – because most workers never switch technologies along an equilibrium path if relative wages are close to their stationary level. Second, the market clearing condition (14) only contains integrals of $s_t(i)$. Because of the continuum of skill types, the effect of the marginal types that switch technologies are of second order when evaluating changes in these integrals. Taken together, the two observations imply that changes in $\hat{l}_t$ over time are a function of $\hat{q}_t$ and $\hat{l}_t$, as seen in (17). Finally, since condition (12) yields a mapping between ω_t and l_t, the dynamics of $\hat{q}_t$ can be written as a function of the path of l_t, as seen in (18).[10]

3 The Adjustment to Skill-biased Innovations

We now analyze the dynamic adjustment of our economy to a permanent, unanticipated increase in the relative productivity A. We refer to this shock as a skill-biased technological innovation since it increases the relative productivity of higher skill-types i sorted into the H technology. We show that the economy's adjustment may significantly change over time due to the endogenous evolution of the skill distribution across generations, a form of Samuelson's LeChatelier principle. How slow this adjustment is then crucially determines the average and distributional welfare consequences of new technologies.

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. Let $\Delta \log(A) > 0$ be the relative productivity shock. The following proposition characterizes the log-change in $\Delta \log(q_t) \equiv \log(q_t/q_{0^-})$, $\Delta \log(y_t) \equiv \log(y_t/y_{0^-})$, and $\Delta \log(l_t) \equiv \log(l_t/l_{0^-})$.

Proposition 1 (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:*

$$\begin{bmatrix} \Delta \log(l_t) \\ \Delta \log(q_t) \\ \Delta \log(y_t) \end{bmatrix} = \underbrace{\begin{bmatrix} -\eta \\ \frac{1}{\rho+\lambda} \\ 1+\kappa\eta \end{bmatrix} \frac{\theta-1}{\theta+\kappa\eta} \Delta \log(A)}_{\text{Short-run}} + \underbrace{\frac{\psi}{\chi} \begin{bmatrix} \eta \\ \frac{-1}{\rho+\delta+\lambda} \\ \theta-1 \end{bmatrix} (1-e^{-\lambda t}) \frac{\theta-1}{\theta+\kappa\eta} \Delta \log(A)}_{\text{Short- to long-run transition}} \tag{20}$$

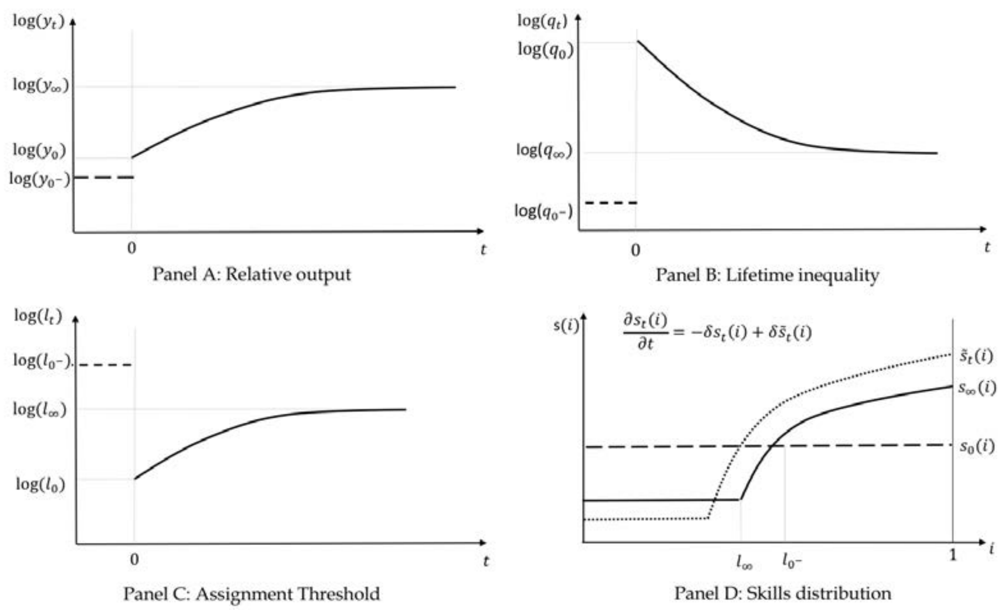
where $\chi \equiv \left(\theta + \kappa\eta + \frac{\psi}{\rho+\delta}\right)(\rho+\delta)$.

Proof. See Appendix A.4. ■

[10] We conjecture that extending the model to multiple sectors is possible while keeping the q-theory representation (albeit in vector form) if skills and sectors satisfy log-supermodularity.

Figure 1 illustrates these dynamic responses together with the dynamics of the skill distribution and skill lottery. We do so for the case where the two technologies are substitutes in production ($\theta > 1$) and $\alpha(i) = 1$.

Figure 1: The economy's adjustment to a skill-biased technological shock ($\theta > 1$)



The first term in (20) is the immediate impact of the shock represented by the responses at $t = 0$ in Figure 1. In the short-run, there are increases in both relative output ($\Delta \log(y_0) > 0$) and lifetime inequality ($\Delta \log(q_0) > 0$). The higher relative wage in the H technology induces the reallocation of skill types in the existing worker generations from the L to the H technology, as can be seen from the decline in the assignment threshold l_t. This reallocation contributes to the relative output increase.

The second term in (20) shows that, along the transition, all variables converge at rate λ. The increase in the relative lifetime wage in high-tech production causes entering worker generations to twist their skill lotteries $\tilde{s}_\tau(i)$ towards high-i types whose skills are more complementary to H production (see bottom right panel of Figure 1). This triggers changes in the economy's skill distribution $s_t(i)$ as older generations are replaced with younger generations at rate δ. Along the transition, the growing mass of high-i types employed with the H technology implies a continuing process of relative output increase and inequality decline. The rising relative high-tech output yields a decline in the consumption price index in (3). By reducing the relative wage, the arrival of more high-i types in younger generations triggers the displacement of marginal i types from H production over time (i.e., l_t increases along the transition).

In the long-run, the H technology has higher relative wage, output and employment. The increase in H employment is driven both by a skill distribution with higher mass in high-i types and a lower assignment threshold of skill types employed in H. Finally, note that the only source of dynamics in our theory is the skill investment decision of generations born after the shock. Whenever incoming generations cannot invest in skills (i.e., $\psi = 0$), the transitional dynamics term in (20) disappears and the responses in the long-run and short-run are identical.

We conclude this section by defining the discounted cumulative impulse response (DCIR). It conveniently summarizes the importance of transitional dynamics and thus relates to how slowly economies adjust to skill-biased innovations. Intuitively, it is the answer to the question: from the point of view of generations alive just before the shock, how different is the adjustment they expect to see during their lifetime compared to the long-run adjustment? We say that the economy's adjustment is *slower* when existing generations expect to miss more of the overall adjustment during their lifetime (i.e., the DCIR is larger).

Definition 2 (Discounted Cumulative Impulse Response) *For any variable z_t and innovation $\Delta \log(A)$, the discounted cumulative impulse response $DCIR(z)$ is:*

$$DCIR(z) = \left| \int_0^\infty \delta e^{-\delta t} \frac{\Delta \log(z_t)}{\Delta \log(A)} dt - \frac{\Delta \log(z_\infty)}{\Delta \log(A)} \right| .$$

Formally, the DCIR is the distance between the long-run response and the expected response of $\log(z_t)$ during the initial generations' lifetime, since all generations born before the shock have exponentially distributed death probabilities with rate δ. This is a convenient measure of the importance of transitional dynamics in our context for a number of reasons. First, it encodes not only the convergence rate λ, but also other relevant features of the impulse responses like how front-loaded they are. For instance, one could have an adjustment where the short- and long-run changes are almost identical—implying a DCIR close to zero—but the rate of convergence λ from the short- to the long-run is very low. According to the DCIR, we would intuitively say that it is a fast adjustment since almost all of the overall adjustment is completed on impact, whereas looking at λ alone suggests a slow adjustment. Second, the DCIR does not mechanically scale with the replacement rate of generations. If δ is higher, this mechanically increases λ (making the adjustment faster) but it also decreases the expected lifetime of a generation. Finally, in the next section, we show that this measure is relevant for analyzing the welfare consequences of skill-biased innovations.

3.2 Changes in average welfare and lifetime welfare inequality

We now compute the average and distributional welfare consequences of skill-biased innovations. Our welfare measure is the ex-ante expected utility of individuals born at each point in time (equation (9)). Given the log-utility assumption, the consumption-equivalent gain is the change in the ex-ante utility times $(\rho+\delta)$. From equations (7) and (9), the consumption-equivalent utility of cohort τ is

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

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

The welfare of each cohort has two components. The first is the average wage across skill types net of the skill lottery cost. This term depends on the relative wage ω_t through $\log(Q_\tau(i))$ because a fraction of workers is employed in high-tech production. The second term is the consumption price index that equally affects all skill types. Importantly, due to the cost of good H, the price index in (3) is decreasing in high-tech productivity A but increasing in ω_t.

To obtain an average welfare measure across generations $\bar{U}$, we take an utilitarian approach by considering a weighted average of the ex-ante utility of different generations, where generation-τ's weight is $re^{-r\tau}$ as in Calvo and Obstfeld (1988). To obtain a measure of average lifetime welfare inequality $\bar{\Omega}$, we first convert lifetime inequality for generation τ into a consumption-equivalent measure $(\rho+\delta)\log q_\tau$ and then simply aggregate all generations using the welfare weights. Then, average welfare and lifetime welfare inequality are

$$\bar{U} = r\int_0^\infty e^{-r\tau}U_\tau d\tau \qquad \text{and} \qquad \bar{\Omega} \equiv (\rho+\delta)r\int_0^\infty e^{-r\tau}\log(q_\tau)d\tau.$$

Appendix A.5 characterizes the first order changes in average welfare $\Delta\bar{U} \equiv \bar{U} - U_{0^-}$ and lifetime welfare inequality $\Delta\bar{\Omega} \equiv \bar{\Omega} - \log(q_{0^-})$ caused by $\Delta\log(A)$. Everything else equal, average lifetime welfare inequality trivially increases when the long-run change of q is larger, but also when transitional dynamics are more important since q remains high for longer. Moreover, while average welfare increases following the innovation, this increase is partially offset if inequality is higher (for a given magnitude of the innovation) whenever the H income share is larger than its employment share (i.e., the average wage of H workers is higher than L workers). This is because the increase in the average wage of workers in H is offset by the higher cost of H goods – and thus the price index – for all workers.

Importantly, the proposition below shows that the welfare consequences of technological innovations depend crucially on how slow the adjustment is.

Proposition 2 (Changes in average welfare and lifetime welfare inequality) *Assume that the H income share is larger than its employment share and that $r = \delta$. Then, for a given long-run elasticity $\frac{\Delta log(q_\infty)}{\Delta log(A)}$,*

$$\frac{\partial \left(\frac{\Delta \bar{U}}{\Delta \log(A)} \right)}{\partial DCIR(q)} < 0 \quad \textit{and} \quad \frac{\partial \left(\frac{\Delta \bar{\Omega}}{\Delta \log(A)} \right)}{\partial DCIR(q)} > 0.$$

Proof. See Appendix A.5.

Intuitively, a higher $DCIR(q)$ arises when the relative supply of H goods increases more slowly over many generations. This then implies that, during their lifetimes, generations born before the shock expect to see a smaller fraction of the long-run increase in relative output and decline in the price index. Yet, they expect to experience even higher inequality compared to those generations born in the long-run. As a result, everything else equal, the average welfare increase will be smaller and the average lifetime welfare inequality increase will be larger.

3.3 LeChatelier Principle and the risks of extrapolation

This section connects the predictions of our theory to those of a reduced-form demand-supply framework. Appendix A.6 shows that our theory yields a reduced-form supply elasticity φ_t such that relative output and wage solve

$$\Delta \log(x_t) = (\theta - 1)\Delta \log(A) - \theta \Delta \log(\omega_t), \tag{21}$$

$$\Delta \log(x_t) = \varphi_t \Delta \log(\omega_t). \tag{22}$$

The main feature of our theory is a time-varying elastic supply of skills across technologies, which implies that φ_t is positive and increasing over time.[11] This arises from two sources. First, even if skills are exogenous ($\psi = 0$), the relative supply elasticity is positive because a fraction of the heterogeneous workers in the economy decides to reallocate across technologies in response to changes in the

[11] The canonical model of Katz and Murphy (1992) is a special case of our theory with $\varphi_t = 0$.

relative wage. Second, the change in the skill investment decision of cohorts born after the shock introduces an additional adjustment margin for relative supply. This margin becomes stronger over time as younger cohorts replace older cohorts, driving φ_t upwards along the transition when $\psi > 0$. Thus, our theory implies a form of Samuelson's LeChatelier principle: the relative supply of high-tech output is more elastic over longer horizons due to changes in the skill distribution.

By microfounding the dynamics of the relative supply elasticity, our results point to two types of risks associated with using reduced-form estimates of φ_t. The first arises when extrapolating from observed responses in the economy over any given horizon. 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. Suppose this researcher then uses her estimates to analyze the consequences of skill-biased innovations. The time-varying nature of 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 researcher will also obtain biased estimates of the welfare consequences of the shock as she will wrongly conclude that the change in lifetime welfare inequality is $(\rho + \delta)\Delta \log q_T = \Delta \log \omega_T$, which may be higher or lower than $\Delta\bar{\Omega}$ depending on the estimation horizon T.

The second type of risk arises when extrapolating from past technological transitions in different contexts: a type of threat to external validity. Consider a researcher that obtains estimates of the path of φ_t from a particular historical episode. Suppose this researcher uses such estimates to make predictions about the dynamic consequences of a new technology in a different economy or historical context. Appendix A.6 shows that if either technology-skill specificity (η) or the cost of skill investment (ψ) are different, then the path of φ_t will be different as well. Thus, the researcher will obtain biased predictions about the economy's adjustment at all horizons whenever the nature of technology and skills or the underlying flexibility of skill investment are significantly different across episodes.

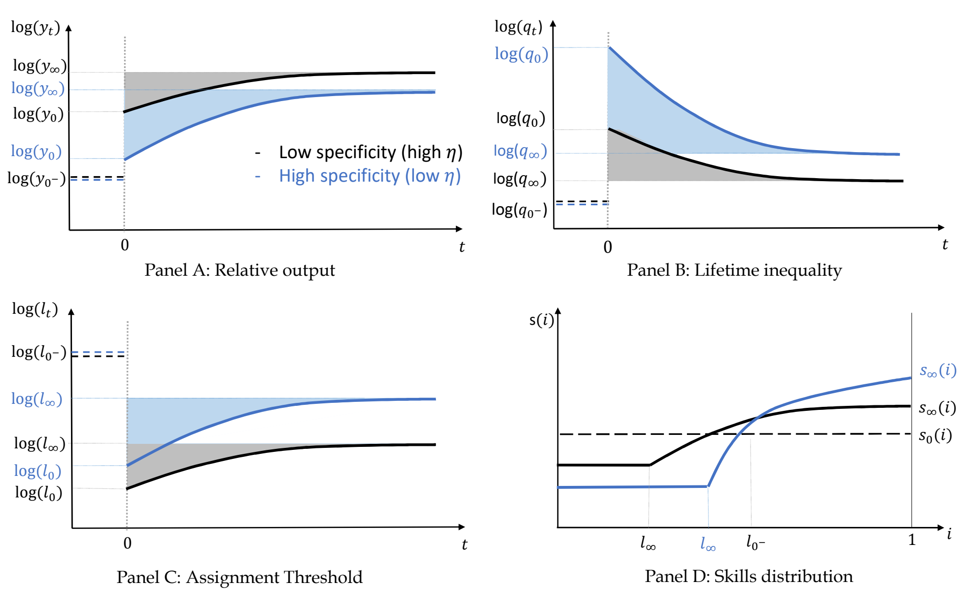
4 Determinants of Skill-Biased Transitions

This section analyzes how parameters governing technology-skill specificity and the cost of skill investment affect the economy's adjustment to a skill-biased innovation. The comparative static exercises speak to when is it that technological transitions are more unequal and slower, with the adjustment mainly driven by slow changes in the skill distribution across generations as opposed to fast reallocation of workers within a generation. As such, they help interpret differences between historical episodes or future transitions where the nature of technological innovations and associated skills are different.

Comparative statics with respect to technology-skill specificity. Consider first how the economy's dynamic responses change with the degree of technology-skill specificity. As reminder, we say that technology-skill specificity is higher when η is lower because the productivity of skills associated with higher i types decreases more when deployed to the L-technology rather than to the H-technology. This exercise speaks to differences in the dynamics across episodes in which skills of incumbent workers were more or less transferable for use in the new technology.

Figure 2 shows the responses of two economies that differ in their technology-skill specificity.[12] The black lines show the responses of an economy with a high value of η (i.e., low technology-skill specificity). The blue lines show the responses of an economy with a low value of η (i.e., high technology-skill specificity). In Appendix A.7, we support the graphical representation in Figure 2 with Proposition A.1 establishing how η affects the short- and long-run responses, the cumulative impulse response, and the rate of convergence.

Figure 2: Comparative statics with respect to η



[12]The figure shows the case where $\theta > 1$ and the threshold's cumulative response increases with η.

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 (i.e., $\frac{\partial|\Delta\log(l_0)|}{\partial\eta} > 0$). As a result, the short-run increase in relative wages and lifetime inequality q are larger (i.e., $\frac{\partial|\Delta\log(q_0)|}{\partial\eta} < 0$) and the increase in relative output is smaller (i.e., $\frac{\partial|\Delta\log(y_0)|}{\partial\eta} > 0$). The larger increase in q then implies that younger entering generations have stronger incentives to invest in those skills that are more complementary to the H technology. As a consequence, there are larger differences in skill heterogeneity across generations.[13] Then, the overall magnitude of the adjustment of y_t and q_t that happens along the transition is larger because larger changes in the skill distribution take place as younger generations replace older generations. Formally, we measure this as the cumulative impulse response function being larger (e.g. $\frac{\partial\int_0^\infty \hat{q}_t dt}{\partial\eta} < 0$) – graphically, the blue shaded areas being larger than the black shaded areas.

Moreover, while the larger changes in the skill distribution could have implied a larger (smaller) overall long-run adjustment in relative output (lifetime inequality), it turns out that the smaller (larger) short-run response dominates. Thus, the long-run adjustment in relative output (lifetime inequality) is smaller (larger) in the economy with higher technology-skill specificity.

Finally, we can come back to the DCIR to summarize how technology-skill specificity affects the importance of transitional dynamics.

Theorem 2.1 (DCIR comparative statics with respect to η) *Following a skill-biased innovation $\Delta\log(A)$, lifetime inequality (q) and relative output (y) adjust slower in economies with a higher degree of technology-skill specificity (lower η). Formally,*

$$\frac{\partial DCIR(q)}{\partial\eta} < 0, \quad \frac{\partial DCIR(y)}{\partial\eta} < 0.$$

Proof. We have that $DCIR(q) = \frac{\lambda\delta}{\lambda+\delta}\left|\frac{\int_0^\infty \hat{q}_t dt}{\Delta\log(A)}\right|$. From Proposition A.1 in Appendix A.7, we know that when η is higher then λ and $\left|\int_0^\infty \hat{q}_t dt\right|$ are both smaller. The proof for y_t is analogous. ■

The theorem shows that the DCIR is larger and transitional dynamics are more important in economies with a higher degree of technology-skill specificity. That is, the adjustment is slower and more back-loaded, with generations alive before

[13] This follows directly from the fact that $\tilde{s}_\tau(i)$ is proportional to $\hat{q}_\tau$ in (19) and $\hat{q}_\tau$ is larger for all τ when technology-skill specificity is higher.

the shock expecting to see less of the long-run changes during their lifetime. Intuitively, this is because the muted reallocation of workers at shorter horizons *causes* larger endogenous changes in the skill distribution along the transition due to the larger increases in lifetime inequality.[14]

To summarize, these results show that when technology-skill specificity is higher, technological transitions will be driven more by changes in the skill distribution across generations than the reallocation of workers within a generation. Therefore, they will be more unequal and play out slower over many generations.

It is also worth noting that the slower adjustment in economies with higher technology-skill specificity does not mechanically follow from the fact that reallocation is smaller in the short-run, or from the fact that old generations are replaced slowly at rate δ. Instead, it follows from the skill distribution responding more to the stronger relative wage change. To make this point clear, Proposition 3 shows that technology-skill specificity has no effect on the DCIR of q and y when either skill distribution or inequality changes are muted ($\psi \to 0$ or $\theta \to \infty$).

Proposition 3 (Interaction of technology-skill specificity with ψ and θ) *When the cost of skill investment is large ($\psi \to 0$) or when H and L are highly substitutable ($\theta \to \infty$), then technology-skill specificity has no effect on how slow q and y adjust.*

$$\left.\frac{\partial DCIR(y)}{\partial \eta}\right|_{\psi \to 0} = \left.\frac{\partial DCIR(q)}{\partial \eta}\right|_{\psi \to 0} = 0, \quad \left.\frac{\partial DCIR(y)}{\partial \eta}\right|_{\theta \to \infty} = \left.\frac{\partial DCIR(q)}{\partial \eta}\right|_{\theta \to \infty} = 0$$

Proof. See Appendix A.8. ■

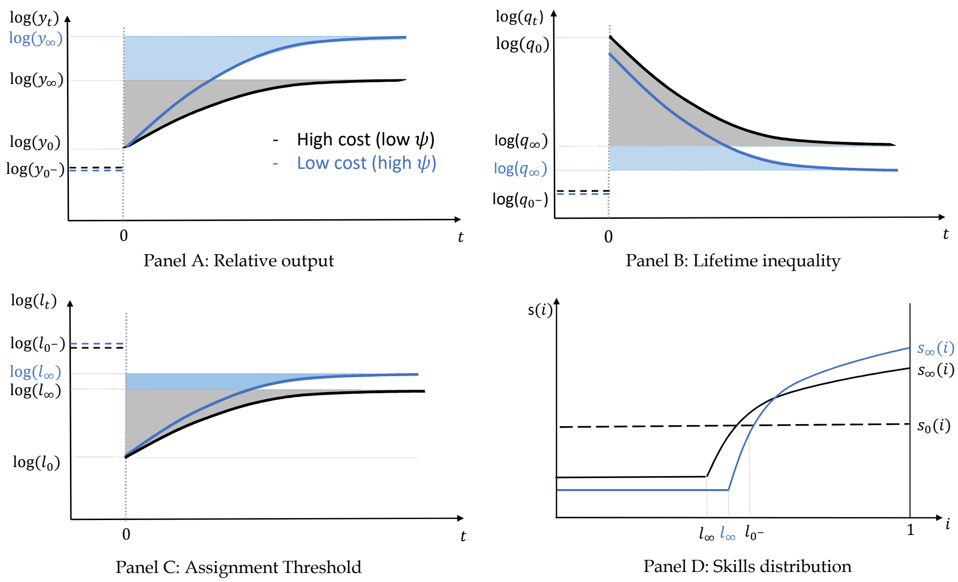
Comparative statics with respect to the cost of skill investment. We now consider how the parameter ψ affects the economy's adjustment to a skill-biased innovation. This comparative static exercise speaks to differences across historical episodes in the gap between younger and older generations' ability to invest in skills. Specifically, it captures situations in which younger generations may have found it easier to invest in skills in high demand than older generations due to, for example, better educational systems, the availability of vocational training for young workers, or the absence of re-training programs for older generations.

Figure 3 illustrates the responses of two economies that differ with respect to the skill investment cost of young generations. The blue lines depict the adjustment of an economy with a low investment cost (i.e., high value of ψ), and the

[14] More generally, after a shock, economies with a less mobile *stock* of a factor experience stronger changes in the *flow* of entrants because of larger changes in relative prices – e.g., if old vintages of physical capital are less adaptable to a new sector, then the flow of firm entrants with newer capital vintages will be larger.

black lines represent the responses of an economy with a high investment cost (i.e., low value of ψ). Proposition A.2 in Appendix A.7 supports this figure.

Figure 3: Comparative statics with respect to ψ



In the short-run, both economies exhibit identical responses in relative output and worker reallocation. 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 implied by the more responsive skill lottery in (19). The larger change in the skill distribution of the economy with a lower investment cost (i.e., higher ψ) has two important implications for its dynamic adjustment to the shock. First, in the long run, it implies that relative output (lifetime inequality) increases more (less). Second, as the following theorem shows, it implies a slower, more back-loaded adjustment in relative output and inequality. Intuitively, when the cost of skill investment for younger workers is lower, transitional dynamics become more important since there are larger changes in the skill distribution.

Theorem 2.2 (DCIR comparative statics with respect to ψ) *Following a skill-biased innovation $\Delta \log(A)$, lifetime inequality (q) and relative output (y) adjust slower in economies with a lower cost of skill investment for younger workers (higher ψ). Formally,*

$$\frac{\partial DCIR(q)}{\partial \psi} > 0, \quad \frac{\partial DCIR(y)}{\partial \psi} > 0$$

Proof. The proof is analogous to the one for Theorem 2.1 but using Proposition A.2 in Appendix A.7 instead. ■

Back to LeChatelier Principle and the risks of extrapolation. To better understand the previous comparative statics, it is useful to return to the reduced-form supply elasticity φ_t introduced in Section 3.3. The different dynamic implications of changing η or ψ arise because the two parameters shape different horizons of this elasticity. Figure 6 in Appendix A.6 illustrates these implications. Both higher values of η and ψ increase the elasticity in the long-run, but the timing differs. Specifically, increasing η flattens the path of φ_t but increasing ψ steepens it. Intuitively, a higher η front-loads more the response in the relative supply of H by making it easier for skill types to reallocate across technologies in response to the shock. This in turn reduces the relative wage changes and, as a result, attenuates changes in the skill distribution across generations and φ_t over time. In contrast, a higher ψ implies that it is easier for new generations to invest in skills, amplifying changes in the skill distribution across generations and φ_t across horizons.

This discussion indicates when researchers should be more cautious about extrapolating from observed changes at short horizons: economies where technology-skill specificity is higher and/or the cost of skill investment for young generations is lower. In such economies, the adjustment is slower and more back-loaded, implying larger changes in all outcomes across generations.

Additional determinants of skill distribution dynamics. The theory so far has ignored several determinants of the dynamics of the skill distribution along the transition. In Appendix B, we present three extensions that relax some of the assumptions of our baseline model. For all extensions, our comparative static results above with respect to η and ψ remain valid.

Our first extension considers a "learning-from-others" externality. This relaxes the assumption that the reference distribution $\bar{s}_t(i)$ in the skill investment problem is exogenous and fixed over time. Instead, we assume that certain skills may be easier to acquire than others because workers learn from others when such skills are already abundant in the economy. This extended model yields dynamic responses that are qualitatively similar to those of our baseline economy when ψ is higher and δ is lower, thus making the adjustment slower. Our second extension relaxes the assumptions that workers can only invest in new skills upon birth by allowing a fraction of older generations to re-train after the shock. This yields responses that are qualitatively similar to our baseline when η is higher, making the adjustment faster. Our third extension allows for population growth. This increases the rate of convergence λ, making the adjustment faster.

5 Observable Implications at Short Horizons

The results above show that technology-skill specificity and skill investment cost determine how unequal and slow technological transitions are. We now establish observable implications of our theory that are informative about η and ψ. We focus on implications at short horizons because: (i) the technologies of interest may be recent and the transition ongoing, or (ii) the effects of new technologies are typically harder to separate from other confounding shocks at longer horizons.

In deriving such observable predictions, we also take into account that measuring several of our theoretical outcomes would require strong assumptions and hence lack robustness. For example, "selection forces" imply that the relative wage in efficiency units, ω_t, is different than the relative average labor income. Furthermore, q_t is a forward-looking variable, so measuring it measure would require observing ω_t along the entire transition. Finally, the measurement of the skill distribution $s_t(i)$ and the technology-skill assignment l_t requires taking an explicit stance on observable attributes determining a worker's skill vector (e.g., college graduation), leading to misspecification if the chosen attributes do not fully determine technology-specific skills.

Given these challenges, we focus on changes in payroll and employment across sectors/occupations which are widely available across countries and periods. Our novel insight is that relative employment changes for different worker generations are connected to the technology-skill specificity and the cost of skill investment. As a result, even if we only observe these changes at short horizons, they are informative about how economies adjust at longer horizons.[15]

Specifically, we consider the short-run responses in (i) the outcomes of the "old" generation born before $t = 0$ (i.e., *within-generation* change) and (ii) the difference between outcomes of the "young" generations born at $t = 0$ and the "old" generations born before $t = 0$ (i.e., *between-generation* change). We formally define these elasticities for relative employment in high-tech production as

[15] The estimation of generation-specific responses are common in analysis of the impact of different types of shocks – e.g., Kim and Topel (1995) and Autor and Dorn (2009). We use our theory to connect such responses to structural parameters controlling the economy's dynamic adjustment to new technologies.

$$\varepsilon_0^{within} \equiv \frac{\log\left(\frac{\int_{l_0}^1 s_0(i)di / \int_0^{l_0} s_0(i)di}{\int_{l_{0-}}^1 s_0(i)di / \int_0^{l_{0-}} s_0(i)di}\right)}{\Delta \log(A)} \quad \text{and} \quad \varepsilon_0^{between} \equiv \frac{\log\left(\frac{\int_{l_0}^1 \tilde{s}_0(i)di / \int_0^{l_0} \tilde{s}_0(i)di}{\int_{l_0}^1 s_0(i)di / \int_0^{l_0} s_0(i)di}\right)}{\Delta \log(A)}.$$

The following theorem shows how η and ψ affect these elasticities.

Theorem 3 (Observable implications in the short-run)

$$\frac{\partial|\varepsilon_0^{within}|}{\partial\eta} > 0, \quad \frac{\partial|\varepsilon_0^{within}|}{\partial\psi} = 0, \qquad \textit{and} \qquad \frac{\partial|\varepsilon_0^{between}|}{\partial\eta} < 0, \quad \frac{\partial|\varepsilon_0^{between}|}{\partial\psi} > 0$$

Proof. See Appendix A.9. ■

In terms of relative employment in the short run, economies with higher technology-skill specificity (i.e., lower η) experience weaker within-generation responses for older workers, but stronger between-generation response differences for younger and older workers. Intuitively, the lower η weakens the reallocation of skill types across technologies, which reduces the reallocation of older workers, amplifies the increase in relative wages, and, consequently, increases skill differences across generations. In contrast, the lower investment cost (i.e., higher ψ) does not affect the within-generation response because older workers chose their skills before the shock. The lower cost however leads to stronger changes in the skill investment decisions of young workers, giving rise to stronger between-generation response differences.

Theorem 3 is informative about when a technological transition will be slower and more back-loaded given responses observed in the short-run. For example, suppose the within-generation elasticity is small, but the between-generation elasticity is large. Our results indicate that this is consistent with high technology-skill specificity and/or small costs of skill investment for young workers. As a consequence, through the lens of our theory, this technological transition should be more unequal and unfold slowly over many generations.

Finally, we note that the within- and between-generation employment elasticities are defined in terms of the shock, $\Delta \log(A)$. Yet, in some applications (like the one in the following sections), the shock is not directly observed. To circumvent this empirical challenge, we can use the observed response of relative payroll in Proposition 1 since it depends on the parameters of technology-skill specificity and skill investment costs, as well as the size of the shock.

6 Application: Cognitive-Biased Transitions

Our theoretical results established that technology-skill specificity and the cost of skill investment for entering generations are connected to within- and between-generation changes in relative employment following skill-biased technological innovations. In this section, we provide three pieces of evidence studying how these two margins affected the adjustment of developed economies to recent cognitive-biased innovations.

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. In line with our theory's predicted changes in skill investment, we use the unique features of the large-scale German training system to document higher growth in the number of trainees in more cognitive intensive occupations. Finally, we explore quasi-random cross-regional variation in adoption timing to estimate 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. Taken together, the evidence suggest that, for recent cognitive-biased innovations, technology-skill specificity is high and the cost of skill investment is smaller for younger generations. Therefore, cognitive-biased transitions are slow and unequal because they are mostly driven by skill changes across generations rather than worker reallocation within a generation.

6.1 Cognitive-intensive employment in developed economies

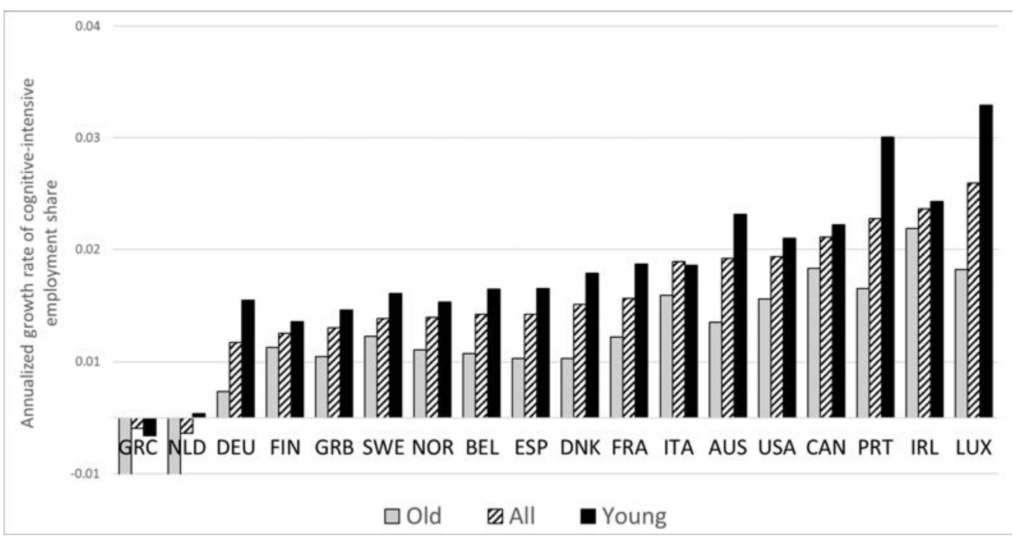
We define cognitive-intensive occupations as being those disproportionately augmented by recent innovations, like the computer and the internet, corresponding to the H technology in the theory. We follow an extensive literature documenting that recent innovations had different effects on jobs with different task content —e.g. Autor et al. (2003), Spitz-Oener (2006), Autor and Dorn (2013), Akerman et al. (2015). Specifically, this literature has documented that new technologies have a more positive impact on cognitive-intensive jobs whose daily activities require problem-solving, creativity, or complex interpersonal interactions. In fact,

Appendix C.1 shows that internet and computer usage is strongly correlated with performing cognitive tasks across occupations in Germany. We also document that there are no systematic differences in internet and computer usage across worker cohorts employed in the same occupation.

For 18 developed countries, we analyze changes in the occupation composition of males in two age groups: "Young" workers aged 15-39 yrs and "Old" workers aged 40-64 yrs. We consider employment in 9 aggregate occupation groups (2-digit ISCO occupations).[16] Using the German BERUFNET dataset, we rank occupations by time spent on tasks that intensively require analytical non-routine and interactive skills.[17] The cognitive-intensive occupations are the top 3 in this ranking: Managers, Professionals, Technicians and Associate Professionals.

Figure 4 displays the recent trends of employment in cognitive-intensive occupations across 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 et al. (2009) and Autor and Dorn (2013).

Figure 4: 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 US, 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 are the top 3 occupation in terms of time spent on cognitive tasks among the 9 2-digit ISCO occupations (Managers, Professionals, Technicians and Associate Professionals). For each country, annualized growth rate is the log-change of the cognitive-intensive employment share in the period divided by the number of years.

[16] Our data comes from Eurostat for European countries and IPUMS International for Canada and US.

[17] The BERUFNET dataset is based on the knowledge of experts about the skills required to perform tasks in each occupation. The occupation's cognitive intensity is the simple average of the time spent on analytical non-routine and interactive tasks in 2011-2013.

Figure 4 also shows that, while older workers increased cognitive-intensive employment in most countries, this increase was substantially stronger for younger generations. Across all countries, the annualized growth in cognitive-intensive employment of younger workers was 73% higher than that of older workers. The young-old gap was higher whenever overall reallocation was higher: there is a correlation of 0.43 between the young-old gap in cognitive-intensive employment growth and that of all workers.[18]

As discussed in Section 5, the different employment responses for young and old workers is consistent with an elastic supply of cognitive skills in the long-run driven by younger generations tilting their investment towards skills used in cognitive-intensive occupations. However, the trends in Figure 4 are subject to concerns about confounding shocks causing cognitive-intensive employment growth. They also do not provide any direct evidence about the skill investment mechanism in our theory. 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 turn to a more detailed investigation of the impact of cognitive-biased technologies on the German labor market.

6.2 Cognitive-intensive employment in Germany

We now study how the Germany adjusted to recent cognitive-biased innovations. We first describe the data used in our analysis. We then investigate the relative performance of occupations with a higher cognitive intensity in terms of employment, payroll, and numbers of trainees. Finally, we exploit quasi-random cross-regional variation in adoption timing of broadband internet to estimate the dynamic impact of this new technology on cognitive-intensive occupations.

6.2.1 Data

Our main data source is the LIAB Longitudinal Model in 1995-2014. We follow Card et al. (2013) to construct a sample of employed males aged 15-64. We first construct a dataset with yearly outcomes for 120 occupations in West Germany. We then construct a second dataset with annual outcomes for each occupation in 323 districts in West Germany.[19] We again use the BERUFNET data to measure

[18] Our new stylized fact complements the finding in Autor and Dorn (2009) that the average age of workers employed in contracting middle-wage occupations increased in the US in 1980-2005.

[19] Section 1.1 of the Online Additional Material describes the sample construction procedure. Our definition of a district as a regional labor market is the same used in Huber (2018).

each occupation's cognitive intensity using the share of time spent on analytical non-routine and interactive tasks.

We construct outcomes for two worker generations. The "Young" generation comprises workers born after 1960, and the "Old" generation includes all other workers. The young generation was at most 35 years old in 1995 when it represented 58% of the German labor force. Its overall employment share then increased to 89% in 2014 when workers were at most 54 years old. We also define a trainee sub-sample with workers whose employment status was a trainee, student trainee, or intern. In this sub-sample, 98% of all workers are below 30 years old and the mean age is 21.

6.2.2 Cognitive intensity and labor market outcomes across occupations

We now study the relationship between employment growth and cognitive intensity in our sample of 120 occupations in West Germany. We therefore move from the sharp predictions of the two-technology theory to look at employment trends across multiple occupations. This yields more variation to empirically investigate the main mechanisms in our theory. Specifically, for each worker generation g and year t, we estimate the following linear regression:

$$\log Y^g_{o,t} - \log Y^g_{o,1995} = \alpha^g_t + \beta^g_t \bar{C}_o + \epsilon^g_{o,t} \tag{23}$$

where $Y^g_{o,t}$ is an outcome in occupation o at year t for workers of generation g, and $\bar{C}_o$ is the cognitive intensity of occupation o.

Table 1 reports the estimation of equation (23) in the periods of 1995-2000 (Panel A) and 1995-2014 (Panel B). We report the estimated impact of the occupation's cognitive intensity on the log-change of employment in columns (1)–(3), payroll in columns (4)-(6), and number of trainees in column (7).

Column (1) indicates that occupations with higher cognitive intensity experienced stronger employment growth. This differential employment response is larger over longer horizons. Compared to the least cognitive-intensive occupation, the employment growth in the most cognitive-intensive occupation was 143% higher in 1995-2014. Importantly, columns (2)–(3) indicate that cognitive-intensive employment growth was weaker for older generations than for younger generations. In fact, estimated coefficients for the old generation are at most one-half those of the young generations. These results show that the German trends in Figure 4 also hold across more disaggregated occupations with different levels of cognitive intensity.

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

Dependent variable:	Employment Growth			Real Payroll Growth			Trainee Growth
	All	Young	Old	All	Young	Old	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Change in 1995-2000							
Cognitive intensity	0.388***	0.650***	0.113***	0.340***	0.616***	0.157***	0.379*
	(0.076)	(0.098)	(0.043)	(0.048)	(0.070)	(0.037)	(0.209)
Panel B: Change in 1995-2014							
Cognitive intensity	1.488***	1.894***	0.871***	1.535***	2.029***	1.044***	2.121***
	(0.225)	(0.234)	(0.229)	(0.227)	(0.238)	(0.223)	(0.385)

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.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Columns (4)-(6) show that the relative payroll responses are slightly stronger than the relative employment responses in 1995-2014. This suggests that there were only small relative changes in the average earnings of those employed in cognitive-intensive occupations. As discussed before, in our theory, these relative payroll responses combine changes in both the relative marginal value of labor and the worker "selection" in more cognitive intensive occupations. 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 strong selection forces created by entry of marginal workers with lower occupation-specific productivity than infra-marginal workers.[20]

Column (7) shows that occupations with a higher cognitive intensity experienced stronger growth in the number of trainees. Trainee programs are an important part of the formal training of young individuals in Germany – especially for non-degree occupations (Eckardt, 2019). As such, changes in the occupation allocation among trainees are a proxy for the changes in the skill investment decision of incoming generations in our theory. So, the estimated coefficient suggests that new generations tilted their investments towards cognitive-intensive skills in 1995-2014.

Taken together, this evidence again speaks qualitatively to the main mechanisms in our model. The small responses in employment for old workers suggest

[20] An extreme version of this pattern arises in assignment models with a Frechet distribution of technology-specific skills where average earnings are identical in all occupations (Hsieh et al., 2013; Burstein et al., 2016).

that skills are very specific to occupations with the same cognitive content. The large differences in employment responses between generations suggest that the cost of skill investment is smaller for younger workers than for older workers. In fact, young workers increase their investment on cognitive skills by becoming trainees in occupations with a higher cognitive intensity. The larger overall responses at longer horizons are consistent with LeChatelier's principle.

One concern with this interpretation of our estimates is that they may not be a consequence of a single technological innovation. Instead, they may be driven by different innovations introduced sequentially in the period – e.g. computers, industrial robots, or the internet. Thus, while our interpretation remains qualitatively valid, it is hard to quantitatively connect the estimates above to the mechanism in our theory because the empirical estimates are not impulse response functions to one-time permanent shocks. That is, the estimated dynamics may potentially confound both the endogenous skill distribution dynamics and the exogenous sequence of innovations. We address this concern in the next section.

Robustness. Section 1.3 in the Online Additional Material investigates the robustness of our findings. Table 1.2 shows that results are qualitatively similar over different horizons. Table 1.3 shows that the positive relative employment growth in cognitive occupations is driven by the top third of occupations by cognitive intensity. Table 1.4 shows that results are similar when changing the definition of the young generation, restricting the sample to native-born Germans, or controlling for exposure to trade and immigration shocks.

6.2.3 Dynamic adjustment to broadband internet adoption

In this section, we analyze the dynamic response to one cognitive-biased innovation: the introduction of broadband internet in the early 2000s.[21] 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 impulse response functions of labor market outcomes for different worker generations. We do so by following Falck et al. (2014) to iso-

[21] As shown by Akerman et al. (2015), broadband internet expanded the relative demand for more educated workers in non-routine jobs inside firms. In Appendix C.1, we show that this technology is disproportionately used in more cognitive-intensive occupations.

late 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

$$Y^g_{io,t} - Y^g_{io,1999} = \sum_{c \in \{\text{young, old}\}} (\alpha^c_t + \beta^c_t \bar{C}_o)\, 1_{[g=c]} DSL_i + \delta_{o,t} + \zeta_{g,t} + X^g_{io,t}\gamma^g_t + \epsilon^g_{io,t}, \tag{24}$$

where o is an occupation, i is a district, and g is a generation. In this specification, $Y^g_{io,t}$ is an outcome, DSL_i is the broadband internet penetration in district i in 2005 (normalized to have standard deviation of one), and $\bar{C}_o$ is the time-invariant measure of the cognitive intensity of occupation o. The specification includes generation-year fixed effects that capture nationwide trends for different cohorts, as well as occupation-year fixed effects that absorb confounding shocks affecting an occupation equally in all regions. We also include a control vector $X^g_{io,t}$ to absorb confouding shocks associated with the pretrend growth in 1995-1999 and the initial demographics of the district.[22]

We are mainly interested on the impact of broadband internet adoption on the relative outcomes of more cognitive intensive occupations for each generation: β^g_t in equation (24). 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. At year t, β^g_t is the difference between regions A and B in the relative outcome of more cognitive intensive occupation among workers of generation g.

The unbiased estimation of β^g_t requires an exogenous source of variation on the adoption of broadband internet across districts. However, 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.

[22]We follow Dix-Carneiro and Kovak (2017) and Freyaldenhoven et al. (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. In the Online Additional Material, we report estimates based on alternative control sets.

To circumvent this issue, we follow Falck et al. (2014) by exploiting variation in the location of pre-existing main distribution frames (MDFs) of the telephone network. The initial roll-out of DSL internet in Germany used the pre-existing telephone network. The transmission technology could not supply high-speed internet to areas more than 4200m away from an existing MDF. Thus, regions located close to MDFs were more likely to adopt broadband internet early. The use of MDF location as an instrument then 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. As argued by Falck et al. (2014), the location of MDFs in the 1960s did not take into account how attractive the nearby region was for broadband internet suppliers in the 2000s. This is reasonable because one of the main determinants of the MDF location was the availability of large empty building sites.

We construct two instrumental variables that measure the district'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 variables in Falck et al. (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-weighted centroid was further than the cut-off threshold of 4200m to the municipality's MDF. 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. We then estimate (24) using the exogenous variation induced by these two measures. Specifically, since the observation unit in equation (24) is an occupation-generation-district triple, our instrument vector includes the two measures interacted with generation dummies and the cognitive intensity of each occupation o, $\bar{C}_o$.[23]

Results. Panel A of Figure 5 reports the estimates of β_t^{old} and β_t^{young} implied by equation (24) for each year between 1996 and 2014. Prior to 2003, regions with early DSL expansion did not experience differential growth in the relative out-

[23] Appendix C.2 investigates how broadband internet penetration in 2005 responded to our two measures of the cost of expanding broadband access across districts. Figure 1.2 in the Online Additional Material presents the pattern of cross-district variation in the cost measures. Appendix Table 3 shows that regions with higher values of these cost measures had a lower share of households with broadband access in 2005. To test for weak instruments for the multiple endogenous variables in (24), Appendix Table 4 shows that we obtain high values for the Sanderson-Windmeijer F-statistics (Sanderson and Windmeijer, 2016).

comes of more cognitive intensive occupations for old and young workers. After 2005, there is a significant impact on the relative employment of young cohorts in 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 0.5 log-points more young workers employed in the most cognitive-intensive occupation than in the least cognitive-intensive occupation. However, we do not find such an effect for old cohorts – if anything, the effect is negative.[24]

In our theory, the small relative employment response of old generations suggests that technology-skill specificity is high (i.e., $\eta \approx 0$). In this case, most workers from old generations do not switch occupations as their skills would have a lower value in the more cognitive-intensive occupations augmented by the innovation. The positive between-generation difference in the relative employment response indicates that incoming cohorts tilt their investments towards skills more suitable for cognitive-intensive jobs (i.e., $\psi > 0$).[25] This creates changes in the skill distribution across generations.

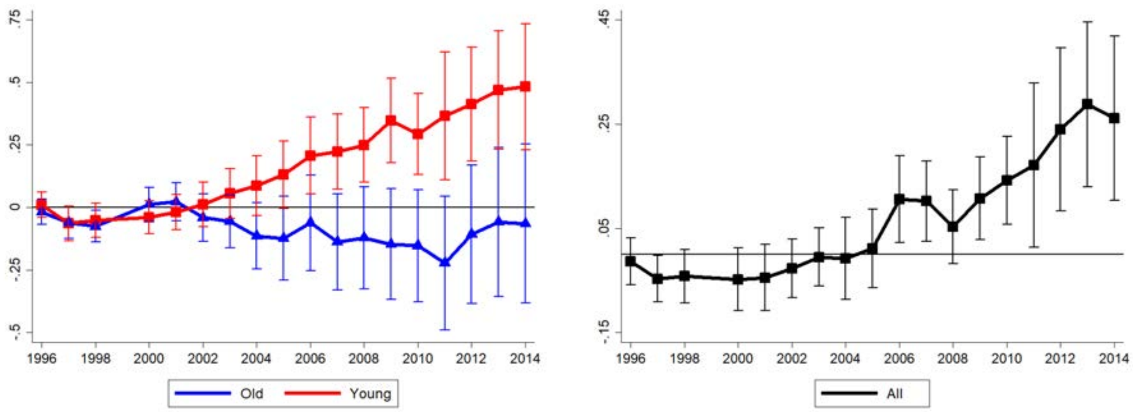
In Panel B of Figure 5, we investigate how early broadband expansion affected the relative payroll of more cognitive intensive occupations for *all* worker generations. Specifically, we estimate equation (24) with a single generation c containing workers of all ages in the district. This is the empirical analog of the impulse response function for relative output y_t presented in Section 3. Again, we find no evidence of responses in the pre-shock period of 1996-2005. Starting in 2006, there is a slow and steady increase in the relative payroll of more cognitive-intensive occupations. In our theory, this is consistent with broadband internet augmenting the relative productivity of cognitive intensive occupations when cognitive and non-cognitive intensive occupations are substitutes in production (i.e., $\theta > 1$).

Additional results and robustness. In the Online Additional Material, we analyze the robustness of our estimates. Table 1.5 shows that results are qualitatively similar when we drop the pretrend control, but estimated coefficients are less precise and slightly smaller in magnitude. Results are also similar when controlling for district-generation-year fixed-effects, so that identification comes purely from the differential effect of the shock on occupations with a higher cognitive intensity. Table 1.6 further shows that our estimates are similar when we use alternative

[24] Appendix Figure 9 shows that the between-generation difference is statistically significant for every year after 2006.

[25] In line with this mechanism, Table 1.7 of the Online Additional Material shows that early adopting regions experienced stronger growth in the number of trainees in more cognitive intensive occupations.

Figure 5: Impact of early DSL adoption on more cognitive-intensive occupations



(a) Relative employment response for each generation

(b) Relative payroll response for all generations

Note. Left panel: estimation of equation (24) for log-employment as dependent variable in the sample of 2 generations, 120 occupations and 323 districts. Right panel: estimation of equation (24) for log-payroll as dependent variable in the sample of 120 occupations, 323 districts, and a single generation with all working-age employed individuals. For each year, the dot is the point estimate of β_t^g. All regressions are weighted by the district population size in 1999 and include occupation-time and generation-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. Bars are the 90% confidence interval implied by the standard error clustered at the district level.

definitions of the young generation. Finally, Table 1.6 reports estimates using a sample excluding workers employed in large establishments that might have received broadband internet access early through private networks. Consistent with this intuition, this sample yields quantitatively stronger estimates.

6.3 Are cognitive-biased transitions different?

The evidence above suggests that recent cognitive-biased innovations triggered a transition that is particularly slow and unequal because of both high cognitive-technology specificity and large changes in skill investment of younger generations (compared to old generations). However, our theoretical results indicate that not all technological transitions are the same. The adjustment may be less unequal and faster if the transition entails more similar changes in relative employment for young and old workers. We build on this insight to investigate whether past changes in employment composition featured weaker between-generation differences and, consequently, may have been part of a transition with lower technology-skill specificity. Here, we only comment on the main results and leave the details for Appendix C.3.

Our analysis compares changes in the occupation composition of young and old workers in Germany and United States before and after 1990. To obtain a time-consistent measure of top expanding occupations, we compute generation-specific employment growth in the top tercile of occupations in terms of employment expansion among young workers.

We find that, in both countries, the two periods exhibit similar employment growth in the top expanding occupations among old workers. As in Figure 4, there is a large between-generation difference in recent years when most expanding occupations were cognitive intensive. However, such a between-generation difference was much smaller before 1990. In this earlier period, the set of expanding occupations was less cognitive intensive, with services and retail occupations at the top of the list in both countries. In fact, Germany did not have any cognitive intensive occupation among the fastest growing occupations before 1990. This evidence is consistent with lower technology-skill specificity for the expanding occupations before 1990, leading to a faster and less unequal transition.[26]

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 economy's dynamic adjustment following technological shocks. In addition, by presenting the full non-linear equilibrium dynamics, the numerical exercise also demonstrates that our theoretical insights are not 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, η) to match the estimates in Figure 5. The decline in the share of the old workers in total employment from 1997 to 2014 implies $\delta = 0.057$, i.e, an ex-

[26] This is just one of many possible interpretations of this evidence. For example, the generation-specific shocks in college graduation rates documented in Card and Lemieux (2001) may help explain why young and old generations have similar changes in employment composition in the 1970s and 1980s.

pected 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 $\psi = 0.35$. Section 2 of the Online Additional Material presents the matching procedure, 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 here on the impact of the shock on average welfare ($\Delta\bar{U}$) and lifetime welfare inequality ($\Delta\bar{\Omega}$), as well as the importance of transitional dynamics as measured by $DCIR(q)$.[28] Section 2 of the Online Additional Material shows the dynamic responses to the shock.

Table 2: Changes in Average Welfare and Lifetime Welfare Inequality

	Baseline ($\eta \approx 0, \psi = 0.35$)		Low specificity ($\eta = 0.75, \psi = 0.35$)	
	$\Delta\bar{U}$	$\Delta\bar{\Omega}$	$\Delta\bar{U}$	$\Delta\bar{\Omega}$
True	46%	39%	44%	29%
Short-run	31%	76%	40%	45%
Long-run	55%	30%	47%	24%
$DCIR(q)$	0.9		0.4	

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.

Table 2 shows that the increase in the average consumption-equivalent welfare 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 almost one-third of the economy's labor force.

Following up on the discussions in Section 3.3, the remaining rows of Table 2 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

[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 4 (Portugal and Netherlands, respectively). Thus, our results can be seen as analyzing the transitional dynamics of a 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.

the changes observed in the long-run were permanent and happened at impact.

We can see that these two calculations lead to substantial biases in welfare analysis. The 'Short-run' calculation severely understates the average welfare gains and overstates the inequality increases. The opposite is true for the 'Long-run' calculation. The biases arise because of the slow adjustment in the economy's skill distribution. For instance, the $DCIR(q)$ of 0.9 reported in the last row implies that a worker born right before the shock expects to experience in her lifetime a relative wage that is 90 percent larger than the long-run relative wage. Thus, the 'Short-run' approach misses the future accumulation of skills that increases relative output– thus reducing the price index– 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 skills necessary to achieve the long-run levels of relative output and wages.

In the remaining columns of Table 2, we analyze the same shock in an economy with a lower degree of technology-skill specificity (i.e., higher η). As discussed in Section 5, in this case, the between-generation difference in the relative employment response is smaller due to the smaller change in the skill distribution across generations. As such, we interpret the comparison between our baseline and this alternative calibration as a numerical illustration of the welfare consequences of the same shock if a lower technology-skill specificity resulted in more similar occupation composition changes for old and young workers (for example, as those discussed in Section 6.3 for the US and Germany before 1990). The second panel of Table 2 shows that the higher η implies a faster transition with $DCIR(q)$ falling from 0.9 to 0.4. This results from the stronger reallocation of old workers at impact which then leads to weaker increases in inequality and changes in the skill distribution. The larger short-run reallocation also leads to a stronger decline in the price index, which translates into a higher average welfare gain in the short-run. Importantly, the faster transition implies smaller biases from the short- and long-run welfare calculations.

8 Conclusions

We develop a theory where overlapping generations of workers are heterogeneous over a continuum of technology-specific skills. Technological transitions are driven both by the reallocation of workers within a generation and changes in the skill distribution across generations. 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 skill-biased innovation, as well as derive observable predictions for changes in labor market outcomes within and between generations. We use these insights to study the adjustment of developed economies to recent cognitive-biased technological innovations. Several pieces of evidence show strong responses of cognitive-intensive employment for young but not old generations.

Taken together, we derive two broad takeaways from this piece. First, the evidence suggests that cognitive-biased transitions may be particularly unequal and slow to play out because of the high specificity of cognitive skills. Most of the adjustment happens through slow changes in the skill distribution across generations as opposed to the fast reallocation of workers within a generation. These features are not universal though. They may be different in past or future technological transitions where a broader set of skills can be transferred to the occupations or sectors improved by the technological innovation. Second, caution should be exercised when interpreting technological transitions based on evidence spanning much less than a generation. This may lead to overly pessimistic views of the consequences of new technologies for inequality and average welfare. Yet, observed changes for different generations, even at short horizons, are useful when combined with a theory of technological transitions. Looking at the decisions of younger workers allows us to "see the future" and thus appropriately derive the full implications of technological innovations.

We think the ideas developed here can also help tackle other problems with a similar structure. It is straightforward to extend the theory to multiple sectors to study how economies adjust to more nuanced labor demand shocks (e.g., trade liberalizations, routine-biased innovations), to include different worker groups (e.g, gender, race) to analyze changes in discrimination, or to reinterpret sectors as regions to study migration. Moreover, future work can address normative questions related to the optimal speed of adjustment to technological (and trade) shocks or the role of workers with more transferable skills in providing aggregate insurance against them. Finally, the notion that within- and between-cohort changes at short horizons are informative about structural parameters governing elasticities at longer horizons can be used both to improve on empirical projections about future labor market conditions as well as to discipline other dynamic models with, for instance, incumbents and entrant firms.

References

Abramitzky, R., Lavy, V., and Segev, M. (2019). The effect of changes in the skill premium on college degree attainment and the choice of major. Technical report, National Bureau of Economic Research.

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier.

Acemoglu, D. and Restrepo, P. (2017). Robots and jobs: Evidence from us labor markets. Working Paper 23285, National Bureau of Economic Research.

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

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

Akerman, A., Gaarder, I., and Mogstad, M. (2015). The Skill Complementarity of Broadband Internet. *The Quarterly Journal of Economics*, 130(4):1781–1824.

Atkeson, A., Burstein, A., and Chatzikonstantinou, M. (2018). Transitional dynamics in aggregate models of innovative investment. Technical report, National Bureau of Economic Research.

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

Atkin, D. (2016). Endogenous skill acquisition and export manufacturing in mexico. *American Economic Review*, 106(8):2046–85.

Autor, D. and Dorn, D. (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, D. H. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American Economic Review*, 103(5):1553–97.

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

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

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

Buera, F. J., Kaboski, J. P., and Shin, Y. (2011). Finance and development: A tale of two sectors. *American economic review*, 101(5):1964–2002.

Burstein, A., Morales, E., and Vogel, J. (2016). Changes in between-group inequality: computers, occupations, and international trade. Technical report, mimeo.

Calvo, G. A. and Obstfeld, M. (1988). Optimal time-consistent fiscal policy with finite lifetimes. *Econometrica: Journal of the Econometric Society*, pages 411–432.

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

Card, D. and Lemieux, T. (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, P., Heckman, J. J., and Vytlacil, E. J. (2011). Estimating marginal returns to education. *American Economic Review*, 101(6):2754–81.

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

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

Charles, K. K., Hurst, E., and Notowidigdo, M. J. (2018). Housing booms and busts, labor market opportunities, and college attendance. *American Economic Review*, 108(10):2947–94.

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

Dix-Carneiro, R. and Kovak, B. K. (2017). Trade liberalization and regional dynamics. *American Economic Review*, 107(10):2908–46.

Dvorkin, M. and Monge-Naranjo, A. (2019). Occupation mobility, human capital and the aggregate consequences of task-biased innovations. Technical report.

Eckardt, D. (2019). Are chemists good bankers? returns to the match between training and occupation. Technical report, LSE.

Falck, O., Gold, R., and Heblich, S. (2014). E-lections: Voting behavior and the internet. *American Economic Review*, 104(7):2238–65.

Freyaldenhoven, S., Hansen, C., and Shapiro, J. M. (2018). Pre-event trends in the panel event-study design. Working Paper 24565, National Bureau of Economic Research.

Gabaix, X., Lasry, J.-M., Lions, P.-L., and Moll, B. (2016). The dynamics of inequality. *Econometrica*, 84(6):2071–2111.

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

Ghose, D. (2019). Trade, internal migration, and human capital: Who gains from india's it boom?

Goldin, C. D. and Katz, L. F. (2009). The Future of Inequality: The Other Reason Education Matters So Much. Scholarly articles, Harvard University Department of Economics.

Goos, M., Manning, A., and Salomons, A. (2009). Job polarization in europe. *American economic review*, 99(2):58–63.

Hansen, L. P. and Sargent, T. J. (2008). *Robustness*. Princeton university press.

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

Hobijn, B., Schoellman, T., and Vindas, A. (2019). Wages during structural transformation: The importance of cohort labor supply decisions.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2013). The allocation of talent and us economic growth. Technical report, National Bureau of Economic Research.

Huber, K. (2018). Disentangling the effects of a banking crisis: Evidence from german firms and counties. *American Economic Review*, 108(3):868–98.

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

Kim, D.-I. and Topel, R. H. (1995). Labor markets and economic growth: lessons from korea's industrialization, 1970-1990. In *Differences and changes in wage structures*, pages 227–264. University of Chicago Press.

Lagakos, D. and Waugh, M. E. (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review*, 103(2):948–80.

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

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98.

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

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

Moll, B., Rachel, L., and Restrepo, P. (2019). Uneven growth: Automation's impact on income and wealth inequality.

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

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

Porzio, T. and Santangelo, G. (2019). Does schooling cause structural transformation?

Rogerson, R. (2008). Structural transformation and the deterioration of european labor market outcomes. *Journal of political Economy*, 116(2):235–259.

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

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

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

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

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

Online Appendix

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). We can re-write it as

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

The right-hand side is strictly decreasing in l_t, converges to zero as $l_t \to 1$, and converges to infinity as $l_t \to 0$. Then, existence and uniqueness of a solution follows from applying Bolzano's theorem.

A.2 Proof of 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

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

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 Theorem 1

Part 1. We start by taking 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

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

We now guess and verify that l_t converges monotonically along the equilibrium path. We establish this starting from $\hat{l}_0 < 0$. We omit the analogous proof for $\hat{l}_0 > 0$. 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 $s > t + \tau(i)$. Thus, given the definition of $Q_t(i)$, we get

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

This implies the following expression for the optimal lottery:

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

The log-linearization of (A.6) implies

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

By replacing (A.7) into the expression inside the parenthesis in (A.4), we obtain

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

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:

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

We then obtain (17) by replacing this expression back in (A.4).

To show (18), we differentiate the definition of $log(q_t)$ with respect to time:

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

Notice that indifference condition (A.4) immediately implies that $\hat{\omega}_t = -(1/\eta)\hat{l}_t$. Then, by log-linearizing the expression above and replacing, we obtain (18)

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

Part 2. We now derive the policy functions, show that the equilibrium is saddle-path stable, and verify that l_t increases monotonically along the equilibrium path.

We start by guessing that the policy functions are given by $\frac{\partial \hat{l}_t}{\partial t} = -\lambda \hat{l}_t$ and $\hat{q}_t = \zeta \hat{l}_t$. By replacing this guess into (17)–(18), we obtain the following system:

$$\begin{aligned} -\lambda =& -\delta + \frac{\eta}{\kappa\eta + \theta}\delta\psi\zeta \\ -\zeta\lambda =& \frac{1}{\eta} + (\rho + \delta)\zeta. \end{aligned}$$

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

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

which yields the following solutions

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

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

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 theorem.

Part 3. Notice that $\int s(i)\hat{\hat{s}}_t(i)di = \int(\tilde{s}_t(i) - s(i))di = 0$. Using (A.7) , we have that

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

We can use use the same arguments as in Appendix A.3 to show that the last term is of second order. Thus,

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

and, therefore,

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

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

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

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 follow from equations A.1 and A.2:

$$
\begin{aligned}
\hat{s}_t(i) &= \hat{s}_0(i)e^{-\delta t} + \int_0^t e^{\delta(\tau - t)}\hat{\hat{s}}_\tau(i)d\tau, \\
\hat{y}_t &= (\theta - 1)\frac{1}{\eta}\hat{l}_t
\end{aligned}
$$

A.4 Proof of Proposition 1

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

$$\begin{aligned}\Delta log(y_t) &= (\theta-1)\left(\Delta log(A) - \Delta log(\omega) - \hat{\omega}_t\right) \\ &= (\theta-1)\left(\Delta log(A) - \left(\Delta log(\omega) + \hat{\omega}_0 e^{-\lambda t}\right)\right) \qquad \text{(A.8)} \\ \Delta log(q_t) &= \Delta log(q) + \hat{q}_t \\ &= \frac{1}{\rho+\delta}\Delta log(\omega) + \frac{1}{\rho+\delta+\lambda}\hat{\omega}_0 e^{-\lambda t} \qquad \text{(A.9)}\end{aligned}$$

Furthermore,

$$\Delta log(l_t) = -\eta \Delta log(\omega_t) = -\eta\left(\Delta log(\omega) + \hat{\omega}_0 e^{-\lambda t}\right) \tag{A.10}$$

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

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

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,

$$\Delta log(\omega) = \frac{1}{\left(\theta+\frac{\psi}{\rho+\delta}\right)+\eta\kappa}(\theta-1)\,\Delta log(A) \tag{A.11}$$

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,

$$\hat{\omega}_0 = \frac{1}{\theta + \kappa\eta} \frac{\psi}{\rho + \delta} \Delta log(\omega) \tag{A.12}$$

Dynamic responses We now use the derivations above to show that

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

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

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

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

A.5 Proof of Proposition 2

The following proposition first characterizes the induced changes in average welfare $\Delta \bar{U} \equiv \bar{U}_0 - U_{0^-}$ and lifetime inequality $\Delta \bar{\Omega} \equiv \bar{\Omega}_0 - log(q_{0^-})$. It is easy to then see that Proposition 2 follows from the expressions for $\Delta \bar{U}$ and $\Delta \Omega$ because, when $r = \delta$, $\frac{r\lambda}{r+\lambda} \int_0^\infty \frac{\hat{q}_\tau}{\Delta log(A)} d\tau = DCIR(q)$.

Proposition 4 (Average welfare and lifetime welfare inequality) *The changes in average welfare $\Delta \bar{U}$ and lifetime inequality $\Delta \bar{\Omega}$ are approximately:*

$$\Delta \bar{U} = \frac{y_\infty}{1 + y_\infty} \Delta \log(A) - \left(\frac{y_\infty}{1 + y_\infty} - \frac{e_\infty}{1 + e_\infty} \right) \Delta \bar{\Omega}$$

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

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 high-tech employment in the long-run.

Proof.

We have that, because of the envelope theorem, for any $\tau \geq 0^-$

$$U_\tau = \int \tilde{s}_\tau(i) V_\tau(i) di - \frac{1}{\psi} \int \tilde{s}_\tau(i) log \left(\frac{\tilde{s}_\tau(i)}{\bar{s}(i)} \right) di$$

$$\approx \int s(i)(V_\tau(i) - V(i)) di + U_\infty$$

Then, for $\tau \geq 0$

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

Also, for $\tau = 0^-$

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

Then,

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

Finally, using Proposition 1,

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

■

A.6 Proof of Demand-Supply representation in (21)–(22)

The demand equation in (2) immediately implies that

$$\Delta \log x_t = (\theta - 1)\Delta \log(A) - \theta \Delta \log \omega_t.$$

We guess and verify the responses in Proposition 1 can be derived from a relative supply equation with the following form:

$$\Delta \log x_t = \varphi_t \log \omega_t.$$

By combining the supply and demand equations, the change in relative wage is given by

$$\Delta \log \omega_t = \frac{1}{\varphi_t + \theta}(\theta - 1)\Delta \log(A)$$

We now derive the expression for $\Delta \log \omega_t$ implied by Proposition 1. The demand equations in (2) implies that

$$\Delta \log \omega_t = \Delta \log(A) + \frac{1}{1-\theta}\Delta \log y_t,$$

which combined with Proposition 1 yields

$$\Delta \log \omega_t = \left[\left(\frac{1}{\theta + \kappa\eta}\right) - \frac{\psi}{\chi}(1 - e^{-\lambda t})\frac{1}{\theta + \kappa\eta}\right](\theta - 1)\Delta \log(A).$$

Equalizing the two expressions above for $\Delta \log \omega_t$, we obtain

$$\varphi_t + \theta = \frac{\theta + \kappa\eta}{1 - \frac{\psi}{\chi}(1 - e^{-\lambda t})},$$

which implies that

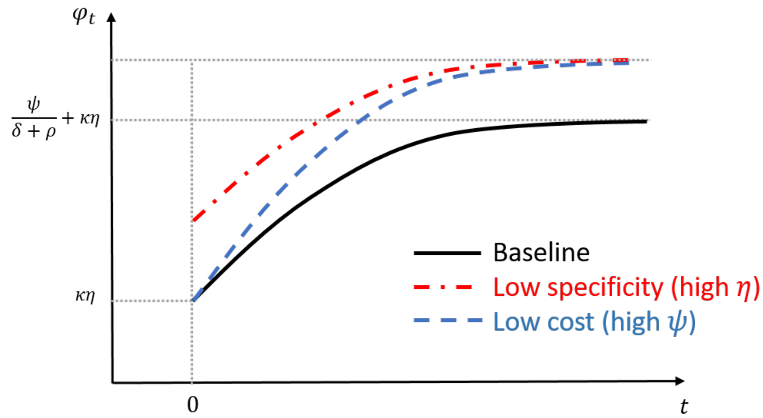
$$\varphi_t = \frac{\kappa\eta\chi + \theta\psi(1 - e^{-\lambda t})}{(\theta + \kappa\eta)(\delta + \rho) + \psi e^{-\lambda t}}.$$

This establishes the representation in (21)–(22) that yields the same path for $\Delta \log \omega_t$ and $\Delta \log y_t$ implied by Proposition 1. Since $e^{-\lambda t} \leq 1$ for all $t \geq 0$, this expression implies that $\varphi_t > 0$ for all t. In addition, we can verify that φ_t is increasing over time because

$$\frac{\partial \varphi_t}{\partial t} = \frac{\theta(\theta + \kappa\eta)(\delta + \rho) + \kappa\eta\chi + \theta\psi}{((\theta + \kappa\eta)(\delta + \rho) + \psi e^{-\lambda t})^2}\psi\lambda e^{-\lambda t} > 0.$$

Finally, Figure 6 illustrates how these elasticities change with η and ψ.

Figure 6: Effect of η and ψ on the elasticity of relative output supply (φ_t)



A.7 Comparative Statics with respect to η and ψ

Proposition A.1 (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*

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

3. *Rate of convergence*

$$\frac{\partial \lambda}{\partial \eta} < 0$$

4. *Cumulative impulse response*

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

Proposition A.2 (Comparative statics with respect to ψ) *Assume that $\theta > 1$. Then,*

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. *Rate of convergence*

$$\frac{\partial \lambda}{\partial \psi} > 0$$

4. *Cumulative impulse response*

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

Next, we prove each of the items of the two propositions above.

1. Short-run adjustment

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

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.

2. 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.

3. Rate of convergence

From the expression for λ in Theorem 1 it is straightforward to see that is decreasing in η and increasing in ψ.

4. Cumulative impulse response

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

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:

$$\frac{\partial log(\frac{1}{\lambda}\frac{\frac{\psi}{\rho+\delta}}{\theta+\kappa\eta+\frac{\psi}{\rho+\delta}}\frac{\theta-1}{\theta+\kappa\eta})}{\partial\eta} = \frac{1}{\lambda}\frac{1}{\rho+2\lambda}\frac{\psi\delta\kappa}{(\theta+\kappa\eta)^2} - \frac{\kappa}{\theta+\kappa\eta+\frac{\psi}{\rho+\delta}} - \frac{\kappa}{\theta+\kappa\eta}$$

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

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.8 Proof of Proposition 3

From the proof of Proposition A.1 in Appendix A.7, we have that

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

$$DCIR(y) = \frac{\delta\lambda}{\lambda+\delta}\frac{\int_0^\infty |\hat{y}_t|dt}{\Delta log(A)} = \frac{\delta}{\lambda+\delta}\frac{\frac{\psi}{\rho+\delta}}{\theta+\kappa\eta+\frac{\psi}{\rho+\delta}}\frac{(\theta-1)^2}{\theta+\kappa\eta} = \frac{\delta}{\lambda+\delta}\frac{(\rho+\delta+\lambda)}{\lambda(\rho+\lambda)}(\lambda-\delta)\frac{(\theta-1)^2}{(\theta+\kappa\eta)^2}.$$

The definition of λ in Theorem 1 implies that $\lambda|_{\psi\to 0} = \lambda|_{\theta\to\infty} = \delta$ and $\frac{\partial\lambda}{\partial\eta}|_{\psi\to 0} = \frac{\partial\lambda}{\partial\eta}|_{\theta\to\infty} = 0$. Taken together, they immediately imply that $\frac{\partial DCIR(q)}{\partial\eta}|_{\psi\to 0} = \frac{\partial DCIR(q)}{\partial\eta}|_{\theta\to\infty} = 0$ and $\frac{\partial DCIR(y)}{\partial\eta}|_{\psi\to 0} = \frac{\partial DCIR(y)}{\partial\eta}|_{\theta\to\infty} = 0$

A.9 Proof of Theorem 3

We start by deriving the elasticity of relative employment of old generations with respect to $\Delta\log(A)$. 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:

$$\varepsilon_0^{within} \approx \frac{1}{\Delta log(A)}\frac{1}{e_\infty(1-e_\infty)}\left(\int_{l_0}^1 s_0(i)di - \int_{l_{0^-}}^1 s_0(i)di\right)$$

Taking a first-order approximation around l,

$$
\begin{aligned}
\int_{l_0}^{1} s_0(i)di - \int_{l_{0-}}^{1} s_0(i)di &\approx -s_0(l)l\left(\hat{l}_t + \Delta\log(l_\infty)\right) \\
&\approx s_0(l)l\eta\Delta\log(\omega_t) \\
&\approx \frac{s_0(l)l}{e_\infty(1-e_\infty)}\eta\left(-\frac{1}{\theta-1}\Delta\log y_0 + \Delta\log A\right).
\end{aligned}
$$

where the last line uses the demand expression in (2).

Then, using Proposition 1,

$$
\varepsilon_0^{within} \approx \frac{s_0(l)l}{e_\infty(1-e_\infty)}\frac{\eta}{\theta+\kappa\eta}(\theta-1),
$$

Thus,

$$
\frac{\partial|\varepsilon_0^{within}|}{\partial\eta} = \frac{s_0(l)l}{e_\infty(1-e_\infty)}\frac{\theta}{(\theta+\kappa\eta)^2}|\theta-1| > 0 \quad \text{and} \quad \frac{\partial|\varepsilon_0^{within}|}{\partial\psi} = 0.
$$

We first use a first-order approximation to write the relative high-tech employment in terms of changes in the high-tech employment share:

$$
\begin{aligned}
\varepsilon_0^{between} \approx& \frac{1}{e_\infty(1-e_\infty)}\frac{1}{\Delta\log A}\left(\int_{l_0}^{1}(\tilde{s}_0(i) - s_0(i))di\right) \\
\approx& \frac{1}{e_\infty(1-e_\infty)}\frac{1}{\Delta\log A}\left(\int_{l}^{1}s(i)(\hat{\tilde{s}}_0(i) - \hat{s}_0(i))di\right)
\end{aligned}
$$

To write this expression in term of fundamentals, we derive the changes in the skill distribution between stationary equilibria. Using the expression for the stationary skill distribution in (16),

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

Recall also that the third part of Theorem 1 yields

$$
\hat{\tilde{s}}_0(i) = \left(\mathbb{I}_{i>l} - \int_l^1 s(i)di\right)\psi\hat{q}_0 + o_0(i).
$$

Combining the expressions above,

$$\begin{aligned}
\varepsilon_0^{between} \approx& \frac{1}{\Delta \log A}\left(\psi \hat{q}_0 + \frac{\psi}{\rho+\delta}\Delta \log(\omega)\right) \\
\approx& \frac{1}{\Delta \log A}\psi\left(\hat{q}_0 + \Delta \log(q_\infty)\right) \\
\approx& \frac{1}{\Delta \log A}\psi \Delta \log(q_0)
\end{aligned}$$

Using the expression for $\Delta \log(q_0)$ in Proposition 1,

$$\varepsilon_0^{between} \approx \frac{\psi}{(\rho+\lambda)(\theta+\kappa\eta)}(\theta-1).$$

Using the expressions derived in Appendix A.8 and defining $\varrho \equiv \left(\frac{\rho}{2}\right)^2 + \delta\left((\rho+\delta)+\frac{\psi}{\theta+\kappa\eta}\right)$, we obtain

$$\begin{aligned}
\frac{\partial|\varepsilon_0^{between}|}{\partial\psi} =& \frac{1-e_\infty}{(\theta+\kappa\eta)(\rho+\lambda)^2}\left(\rho+\lambda-\psi\frac{\partial\lambda}{\partial\psi}\right)|\theta-1| \\
=& \frac{1}{(\theta+\kappa\eta)(\rho+\lambda)^2}\left(\frac{\rho}{2}+\varrho^{1/2}-\frac{1}{2}\varrho^{-1/2}\frac{\delta\psi}{\theta+\kappa\eta}\right)|\theta-1| \\
=& \frac{1}{(\theta+\kappa\eta)(\rho+\lambda)^2}\varrho^{-1/2}\left(\frac{\rho}{2}\varrho^{-1/2}+\varrho-\frac{1}{2}\frac{\delta\psi}{\theta+\kappa\eta}\right)|\theta-1| \\
=& \frac{1}{(\theta+\kappa\eta)(\rho+\lambda)^2}\varrho^{-1/2}\left(\frac{\rho}{2}\varrho^{-1/2}+\left(\frac{\rho}{2}\right)^2+\delta\left(\rho+\delta\right)+\frac{1}{2}\frac{\delta\psi}{\theta+\kappa\eta}\right)|\theta-1|,
\end{aligned}$$

which implies that $\frac{\partial|\varepsilon_0^{between}|}{\partial\psi} > 0$.

Using the expressions derived in Appendix A.8,

$$\begin{aligned}
\frac{\partial|\varepsilon_0^{between}|}{\partial\eta} =& -(1-e_\infty)\frac{\psi}{\left[(\rho+\lambda)(\theta+\kappa\eta)\right]^2}\left(\kappa(\rho+\lambda)+(\theta+\kappa\eta)\frac{\partial\lambda}{\partial\eta}\right)|\theta-1| \\
=& -\frac{\psi}{\left[(\rho+\lambda)(\theta+\kappa\eta)\right]^2}\kappa\varrho^{-1/2}\left(\frac{\rho}{2}\varrho^{-1/2}+\varrho-\frac{1}{2}\frac{\delta\psi}{(\theta+\kappa\eta)}\right)|\theta-1| \\
=& -\frac{\psi}{\left[(\rho+\lambda)(\theta+\kappa\eta)\right]^2}\kappa\varrho^{-1/2}\left(\frac{\rho}{2}\varrho^{-1/2}+\left(\frac{\rho}{2}\right)^2+\delta(\rho+\delta)+\frac{1}{2}\frac{\delta\psi}{(\theta+\kappa\eta)}\right)|\theta-1|,
\end{aligned}$$

which implies that $\frac{\partial|\varepsilon_0^{between}|}{\partial\eta} < 0$.

Appendix B Extensions

This section discusses the extensions described in Section 4.

B.1 Learning-from-others

We relax the assumption that the reference distribution $\bar{s}_\tau(i)$ in the skill investment problem is exogenous and fixed over time. Instead, we assume that certain skills may be easier to acquire than others because workers can "learn from others" when such skills are already abundant in the economy. Formally, we assume that the baseline distribution $\bar{s}_\tau(i)$ is a geometric average of a fixed distribution $\bar{\epsilon}(i)$ and the current skill distribution in the economy $s_\tau(i)$ at the time where generation τ is born,

$$\bar{s}_\tau(i) = s_\tau(i)^\gamma \bar{\epsilon}(i)^{1-\gamma}, \qquad \gamma \in [0,1).$$

Note that as γ increases it becomes easier for workers to choose skill lotteries that put more weight in those skill types that are already abundant in the economy. As opposed to our benchmark case ($\gamma = 0$), this extension with $\gamma > 0$ introduces a backward-looking element to the skill investment problem and complementarities in skill investment decisions across generations.

In what follows, we reproduce the key steps that change in the proofs in Appendix A.3. First, we log-linearize the extended version of (A.6). We begin by noting that the stationary distribution exist and is

$$s(i) = \frac{s(i)^\gamma \epsilon(i)^{1-\gamma} w(i)^{\frac{\psi}{\rho+\delta}}}{\int_0^1 s(j)^\gamma \epsilon(j)^{1-\gamma} w(j)^{\frac{\psi}{\rho+\delta}} dj} \Longrightarrow s(i) = \frac{\epsilon(i) w(i)^{\frac{1}{1-\gamma}\frac{\psi}{\rho+\delta}}}{\int_0^1 \epsilon(i) w(i)^{\frac{1}{1-\gamma}\frac{\psi}{\rho+\delta}} di}.$$

Then, we obtain that

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

Second, replacing the above in the expression inside the parenthesis in (A.4),

$$\left(\int_l^1 \hat{\bar{s}}_t(i) \frac{\alpha(i)\sigma(i)s(i)}{\int_l^1 \alpha(i)\sigma(i)s(i)di} di - \int_0^l \hat{\bar{s}}_t(i) \frac{\alpha(i)s(i)}{\int_0^l \alpha(i)s(i)di} di\right) =$$

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

$$\gamma \frac{\kappa\eta + \theta}{\eta} \hat{l}_t + \psi \hat{q}_t - \psi \int_{l_t}^l \left(\hat{q}_t - \hat{q}_{t+\tau(i)}\right) \frac{\alpha(i)s(i)}{\int_0^l \alpha(x)s(x)dx} di$$

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

Third, as in the proof in Appendix A.3, we can show that the last term inside the integral is of second order. Thus, replacing the above expression back in (A.4), we obtain the Kolmogorov-Forward equation for $\hat{l}_t$ in the economy with learning-from-others,

$$\frac{\partial \hat{l}_t}{\partial t} = -\delta(1-\gamma)\hat{l}_t + \frac{\eta}{\kappa\eta + \theta}\delta\psi\hat{q}_t.$$

Fourth, since the law of motion for $\hat{q}_t$ is the same as in the benchmark model, this implies that the equilibrium is saddle-path stable where the new λ in the economy with learning-from-others is the positive solution to

$$(\delta(1-\gamma)-\lambda)(\rho+\delta+\lambda) + \frac{\psi\delta}{\kappa\eta+\theta} = 0.$$

Finally, the optimal lottery in the economy with learning-from-others is

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

Next, we reproduce the key steps that change in Appendices A.4 and A.7. First, from the expression for the stationary distribution above, note that the long-run skill supply elasticity in the learning-from-others economy is $\frac{1}{1-\gamma}\psi$ as opposed to simply ψ.

This implies that the dynamic responses are

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

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

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

where the last line follows from the equation for the new λ.

Second, note that the short-run responses for l_t and y_t are identical than in the benchmark model. The long-run responses are larger (smaller) in magnitude for y_t (for l_t) in the economy with learning-from-others since the long-run skill supply elasticity is larger and thus $\frac{1}{\theta+\kappa\eta+\frac{1}{1-\gamma}\frac{\psi}{\rho+\delta}}\frac{1}{1-\gamma}\frac{\psi}{\rho+\delta}$ is larger. As for the DCIR, note that λ is smaller in the learning-from-others economy. Together with the fact that $\frac{1}{\theta+\kappa\eta+\frac{1}{1-\gamma}\frac{\psi}{\rho+\delta}}\frac{1}{1-\gamma}\frac{\psi}{\rho+\delta}$ is larger, they imply that the DCIR of both y_t and l_t is higher in the learning-from-others economy.

Third, for q_t we have that

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

$$\Delta log(q_0) = \frac{1}{\theta+\kappa\eta}\frac{1}{(\rho+\delta+\frac{\psi\delta}{\theta+\kappa\eta}\frac{1}{\rho+\delta+\lambda})}(\theta-1)\Delta log(A)$$

$$\int_0^\infty \hat{q}_t dt = \frac{\frac{1}{1-\gamma}\frac{\psi}{\rho+\delta}}{\theta+\kappa\eta+\frac{1}{1-\gamma}\frac{\psi}{\rho+\delta}}\frac{1}{\rho+\delta+\lambda}\frac{1}{\lambda}\frac{1}{\theta+\kappa\eta}(\theta-1)\Delta log(A).$$

Then, since λ is smaller, the short- and long-run responses are smaller in magnitude and the DCIR is larger in the economy with learning-from-others.

Finally, we note that the proofs for the comparative statics in Appendix A.7 with respect to η and ψ are unchanged. To see this, it suffices to show that the dynamics for q_t, l_t, y_t in the economy with learning-from-others are equivalent to those from a re-parameterized benchmark economy where $\delta' = \delta(1-\gamma)$, $\psi' = \frac{1}{1-\gamma}\psi$ and $\rho' = \rho + \delta\gamma$.

B.2 Old generations skill investment

We now let a fraction of workers that were present before the shock re-optimize their skill investment "as if" they were a young generation entering at time $t = 0$. Formally, the skill distribution on impact now becomes

$$s_0(i) = (1-\beta)s_{0^-}(i) + \beta\tilde{s}_0(i),$$

where β is the fraction of workers in the generation present before the shock that can re-optimize.

The first thing to note is that this does not change any of the transitional dynamics given the new initial skill distribution on impact. As such Theorem 1 is unchanged. However, the initial conditions and the dynamic responses do change. Next, we reproduce the key steps that change in Appendix A.4.

The deviation from the skill distribution on impact from the new stationary distribution is now

$$\begin{aligned}\hat{s}_0(i) &= \hat{s}_{0^-}(i) + \beta\left(\hat{\tilde{s}}_0(i) - \hat{s}_{0^-}(i)\right)\\ &= (1-\beta)\left(\hat{s}_0(l) - \mathbb{I}_{i>l}\frac{\psi}{\rho+\delta}\Delta log(\omega)\right) + \beta\left(\mathbb{I}_{i>l} - \int_l^1 s(i)di\right)\psi\hat{q}_0 + \beta o(i)\end{aligned}$$

where the long-run change $\Delta log(\omega)$ is the same as in the benchmark model. Following the same steps as in the benchmark proof, this then implies that

$$\begin{aligned}\left(\frac{\theta}{\eta}+\kappa\right)\hat{l}_0 &= \int_l^1 \frac{\sigma(i)\alpha(i)s(i)}{\int_l^1 \sigma(i)\alpha(i)s(i)}\hat{s}_0(i)di - \int_0^l \frac{\alpha(i)s(i)}{\int_0^l \alpha(i)s(i)}\hat{s}_0(i)di \\ &= -(1-\beta)\frac{\psi}{\rho+\delta}\Delta log(\omega)+\beta\psi\hat{q}_0.\end{aligned}$$

Thus,

$$\begin{aligned}\hat{\omega}_0 &= -\frac{1}{\eta}\hat{l}_0 \\ &= \frac{1}{\theta+\kappa\eta}\left(\frac{\psi}{\rho+\delta}\Delta log(\omega)-\beta\left(\frac{\psi}{\rho+\delta}\Delta log(\omega)+\psi\hat{q}_0\right)\right) \\ &= \frac{1}{\theta+\kappa\eta}\left(\frac{\psi}{\rho+\delta}\Delta log(\omega)-\beta\left(\frac{\psi}{\rho+\delta}\Delta log(\omega)+\frac{\psi}{\rho+\delta+\lambda}\hat{\omega}_0\right)\right) \\ &= \frac{1-\beta}{1+\beta\frac{\psi}{\rho+\delta+\lambda}\frac{1}{\theta+\kappa\eta}}\frac{1}{\theta+\kappa\eta}\frac{\psi}{\rho+\delta}\Delta log(\omega).\end{aligned}$$

Finally, using the above together with the expression for $\Delta log(\omega)$ in equations (A.8)-(A.10), we obtain:

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

Then, mathematically, the dynamic responses in the economy where old generations can re-optimize their skills are similar to those in the benchmark economy except that the function $e^{-\lambda t}$ is now multiplied by $\frac{1-\beta}{1+\beta\frac{\lambda-\delta}{\delta}} < 1$. This immediately implies that: the long-run responses are the same in both economies, the short-run responses of y and l (of q) are now larger (smaller) in magnitude, and the DCIR of all variables is now smaller. Hence, in many ways, this new economy behaves qualitatively similar to an economy with a lower degree of skill specificity (higher η), with the exception that long-run responses are unchanged.

B.3 Population growth

We now assume that the size of entering generations is μ as opposed to δ. This implies that the population growth rate is $\mu - \delta$. The Kolmogorov-Forward equation describing the evolution of the skill distribution becomes

$$\frac{\partial e^{(\mu-\delta)t} s_t(i)}{\partial t} = -\delta e^{(\mu-\delta)t} s_t(i) + \mu e^{(\mu-\delta)t} \tilde{s}_t(i).$$

Then, we have that

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

The remaining elements in the model remain the same. Hence, the economy with population growth is identical to our benchmark economy except that the convergence rate λ is higher iff $\mu > \delta$ since it is now the positive solution to:

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

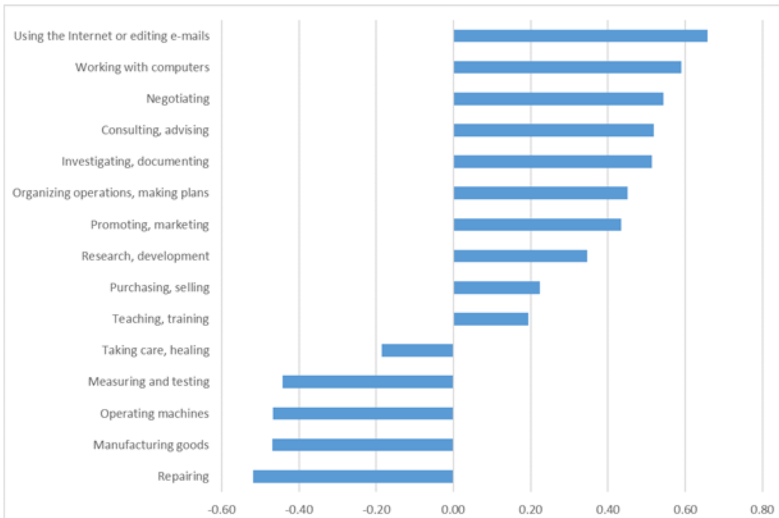
Then, if $\mu > \delta$, the short- and long-run dynamic responses for y_t, l_t remain unchanged, the short-run response of q is smaller in magnitude, and the DCIR of all variables is lower. The opposite holds when $\mu < \delta$.

Appendix C Empirical Analysis

C.1 Cognitive intensity and use of new technologies across occupations in Germany

This section analyzes the types of tasks required by cognitive-intensive occupations. Figure 7 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 7 are consistent with the evidence establishing the heterogeneous impact of new technologies on different tasks performed by workers – e.g., Autor et al. (2003), Spitz-Oener (2006), Autor and Dorn (2013), and Akerman et al. (2015).

Figure 7: 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).

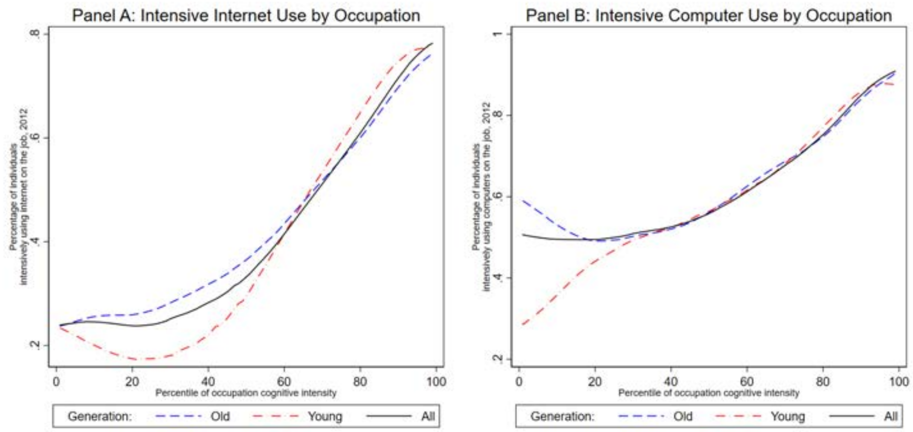
We then investigate whether these new technologies affected worker generations differently conditional on their occupation. We consider two generations: a young generation aged less than 40 years and an old generation aged more than 40 years.[29] Figure 8 shows that, while internet and computer usage 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 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.

C.2 Dynamic adjustment to broadband internet adoption across regions and occupations

We start by examining the first-stage regression that relates the initial telephone network to DSL access. Although the unit of observation in equation (24) is a district-occupation-generation triple, the exogenous variation in the instrument vector comes only from cross-district variation. Therefore, to provide a clear picture of the exogenous variation underlying the first-stage regression, we first examine the impact of the instrument vector Z_i on the district's share of population

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

Figure 8: 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.

with broadband internet access in 2005, DSL_i. That is, we begin by estimating the following linear regression:

$$DSI_i = Z_i\rho + X_i\gamma + \epsilon_i$$

where Z_i is the vector of instruments described in Section 6.2 and X_i is the vector of district-level controls used in the estimation of (24). We refer to the first instrument as "MDF Density Measure" (number of municipalities in the district without a MDF in a 4200m radius) and to the second instrument as "Alternative MDF Availability" (number of municipalities in the district without a MDF in a 4200m radius and without access to an alternative MDF in a neighboring district in a 4200m radius).

Table 3 shows that districts with adverse initial conditions for internet adoption had a lower share of households with high-speed internet in 2005. Columns (1) reports 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.

As discussed in Section 6, equation (24) has multiple endogenous variables since they include DSL access interacted with occupation cognitive intensity and worker generation dummies. To test for weak instruments in this setting, we provide the Sanderson-Windmeijer F-statistics (Sanderson and Windmeijer, 2016) for the first stage of each specification in Table 4. This test statistic checks for whether any of our endogenous variables are weakly instrumented, as well as whether there are sufficiently many strong instruments to instrument the multiple endogenous variables. As shown in the table, we obtain uniformly high first-stage SW F-statistics in all specifications, indicating that our instrument vector has enough power to estimate responses for different worker cohorts.

Table 3: First-stage regressions – Share of households with DSL access in 2005

	(1)	(2)
MDF density measure	-0.020***	-0.018***
	(0.005)	(0.005)
Alternative MDF availability	0.002	-0.001
	(0.001)	(0.002)
Baseline controls	Yes	No
F statistic	26.49	43.06

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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: First-stage SW F-statistics for estimation of equation (24) reported in Panel A of Figure 5

Instrumented Variable	1997	2007	2014
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

Note. Sample of 2 cohorts, 120 occupations and 323 districts. Table reports the Sanderson-Windmeijer F-statistic for each endogenous regressor when estimating equation (24).

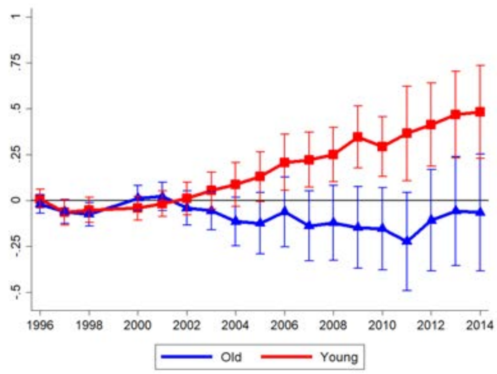
Finally, Figure 9 shows the between-generation employment differences

C.3 Occupations composition changes for young and old workers in different periods

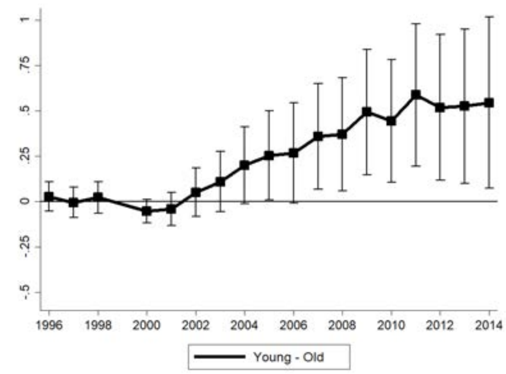
This section analyzes occupation composition trends for young and old workers in the fastest growing occupations in different periods. We focus on Germany and United States before and after 1990. We again use the nine aggregate occupations in the 2-digit ISCO classification used in Section 6.1. Due to data availability, our early period is 1970-1987 for Germany and 1960-1990 for the United States. For both countries, the recent period is the same as the one used to compute the trends in Figure 4.[30] To obtain a measure of the expanding occupations that is

[30]Our primary data source for the early period is the individual-level Census data downloaded from IPUMS international, which contains information on the 2-digit ISCO occupation of males aged 16-64 in each Census year. For the recent period, we compute all outcomes using the same underlying data of Figure 4. We select a sample of employed males in each country-year and split them into two age groups: "Young" workers aged 15-39yrs and "Old" workers aged 40-64yrs.

Figure 9: Impact of early DSL adoption on employment in more cognitive-intensive occupations: Old and Young generations



(a) Relative employment response for each generation



(b) Between-generation response difference

Note. Estimation of equation (24) in the sample of 2 cohorts, 120 occupations and 323 districts. Dependent variable: log employment. The left panel reports β_t^g for olf and young generations, and the right panel reports $\beta_t^{young} - \beta_t^{old}$. 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. Bars are the associated 90% confidence interval implied by the standard error clustered at the district level.

consistent over time, we no longer rely on the set of cognitive intensive occupations since past shocks may have augmented a different set of skills. Instead, for each country and period, we define the expanding occupations as the three occupations with the highest change in log employment share among young workers. Through the lens of our theory, since young workers adjust their skills to work on occupations that became more attractive, their employment decisions provide a revealed-preference way of recovering the occupations experiencing positive demand shocks under the assumption of no shocks to the cost of investing on different skills.

Table 5 reports the employment growth trends in the three occupations with the highest growth among young workers. Columns (1) and (3) report substantial growth in these occupations for both periods and countries. Interestingly, columns (2) and (4) show that the two periods differ in the relative magnitude of the between- and within-generation components of employment changes. As in Figure 4, there is a large between-generation difference in recent years when most expanding occupations were cognitive intensive. However, such a between-generation difference was much smaller before 1990 when changes in the occupation composition was more similar for young and old generations. In this earlier period, the set of expanding occupations was less cognitive intensive with services and retail occupations at the top of the list in both countries. In fact, Germany did not have any cognitive intensive occupation among the fastest growing occupations in 1970-1987.

To further investigate this trend reversal, we focus on the United States because it has individual-level data from the U.S. Census containing 2-digit ISCO occupation information for 1960, 1970, 1980, 1990, 2000, 2010, and 2015. We use this data to select a sample of males aged 16-64 years old in each year. For each occupation o, we use this sample to compute the change in the average age of its workers between years t and t_0 ($\Delta \bar{A}_{o,t} \equiv \bar{A}_{o,t} - \bar{A}_{o,t_0}$) and the change in the employment share in the same period ($\Delta e_{o,t} \equiv e_{o,t} - e_{o,t_0}$). We then compute the correlation between $\Delta \bar{A}_{o,t}$ and $\Delta e_{o,t}$ across the nine occupations weighted by their employment share in 1960.

Table 6 shows that, in line with Figure 4, the expanding occupations in recent periods attracted young individuals, leading to reductions in the average age of its workers. However, this was not the case in previous periods. Between 1960 and 1990, the correlation between changes in average age and employment share were much weaker. In fact, this correlation was positive in 1960-1980.

Table 5: Changes in between-generation employment differences and employment shares across occupations in different periods

	Early period		Recent period	
	$\frac{1}{T}\Delta \log e_t^{all}$	$\frac{\Delta \log(e_t^{young}/e_t^{old})}{\Delta \log e_t^{old}}$	$\frac{1}{T}\Delta \log e_t^{all}$	$\frac{\Delta \log(e_t^{young}/e_t^{old})}{\Delta \log e_t^{old}}$
	(1)	(2)	(3)	(4)
Germany	1.94%	0.173	1.59%	0.456
United States	0.91%	0.206	1.45%	0.476

Note. Columns (1) and (3) report the annualized growth rate in the three 2-digit ISCO occupations with the highest change in log employment share in the period among young workers in the country (where T is the number of years in the period). For the top 3 occupations by log-employment growth for young workers, columns (2) and (4) report the ratio between the log-change in the between-generation employment share and the log-change in employment share for old workers. Early period: 1970-1987 for West Germany and 1960-1990 for the United States. Recent period: 1997-2017 for Germany and 2000-2015 for the United States. Sample of males in two age groups: "Young" workers aged 15-39yrs and "Old" workers aged 40-64yrs.

Table 6: Changes in mean age and employment share across occupations, United States

Period	$Corr(\Delta \bar{A}_{o,t}, \Delta e_{o,t})$
2000-2015	-0.53
1990-2010	-0.63
1980-2000	-0.60
1970-1990	-0.03
1960-1980	0.35

Note. For each period, the table reports the correlation between $\Delta \bar{A}_{o,t}$ and $\Delta e_{o,t}$ across the nine 2-digit ISCO occupations (weighted by their employment share in 1960). For each occupation and period, $\Delta \bar{A}_{o,t}$ is the change in the mean age and $\Delta e_{o,t}$ is the change in employment share. Sample of males 16-64 years old in the United States.