Online Appendix (not for publication)

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 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 in (19), 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 $\int s(i)\hat{\tilde{s}}_t(i)di = \int(\tilde{s}_t(i) - s(i))di = 0$. Using (A.7) , we have that

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

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

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

and, therefore,

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

To prove the result, we use 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 y_t were already derived in equations A.1 and A.2.

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,

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

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

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

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

A.5 Proof of Proposition 2

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

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

Then, 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 (25)–(27)

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 (25)–(27) 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

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

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*

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

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

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

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

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

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

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

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

$$\int_0^\infty |\hat{l}_t| dt = \frac{\eta}{\theta - 1}\int_0^\infty |\hat{y}_t| dt$$

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

Part 1. 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,

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

Part 2. 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),

$$s_0(i) = \frac{\bar{s}(i)\alpha(i)^{\frac{\psi}{\rho+\delta}} \left(\omega_{0^-}\sigma(i)\right)^{\frac{\psi}{\rho+\delta}\mathbb{I}_{i>l_{0^-}}}}{\int_0^{l_{0^-}} \bar{s}(j)\alpha(j)^{\frac{\psi}{\rho+\delta}} dj + \int_{l_{0^-}}^1 \bar{s}(j)\alpha(j)^{\frac{\psi}{\rho+\delta}} \left(\omega_{0^-}\sigma(j)\right)^{\frac{\psi}{\rho+\delta}} dj}$$

$$\Longrightarrow$$

$$\hat{s}_0(i) \approx -\left(\mathbb{I}_{i>l} - \int_l^1 s(j)dj\right) \frac{\psi}{\rho+\delta} \Delta \log(\omega)$$

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

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

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

Using the expressions derived in Appendix A.8,

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

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

Appendix B Extensions

This section discusses the extensions described in Section 4.4. In Section B.1, we introduce "learning-from-others" by allowing the innate ability distribution $\bar{s}_t(i)$ to depend on the economy's skill distribution $s_t(i)$. In Section B.2, we introduce re-training by allowing a fraction of the older generations to make skill investment following the shock. Section B.3 introduces population growth by allowing the birth rate to be higher than the death rate.

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 $\bar{e}(i)$ and the current skill distribution in the economy $s_\tau(i)$ at the time where generation τ is born,

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

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

$$\hat{\hat{s}}_t(i) = \gamma\left(\hat{s}_t(i) - \hat{s}_t(l)\right) + \hat{\hat{s}}_t(l) - \psi\hat{q}_t\mathbb{I}_{i<l_t} - \psi\hat{q}_{t+\tau(i)}\mathbb{I}_{i\in(l_t,l)}. \tag{B.2}$$

Second, we replace the above in the expression inside the parenthesis in (A.4), we obtain

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

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,

$$\frac{\partial\hat{l}_t}{\partial t} = -\delta(1-\gamma)\hat{l}_t + \frac{\eta}{\kappa\eta+\theta}\delta\psi\hat{q}_t. \tag{B.3}$$

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:

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

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

This appendix presents additional empirical results that complement those presented in Section 6. Section C.1 discusses details about the data construction procedure. Section C.2 reports the summary statistics of our baseline sample. Section C.3 presents evidence on the types of tasks required by cognitive-intensive occupations. Section C.4 presents additional empirical results that attest the robustness of the estimates presented in Section 6.2. Section C.5 presents the correlation between changes in the mean age of an occupation and its employment share in the United States over different time periods.

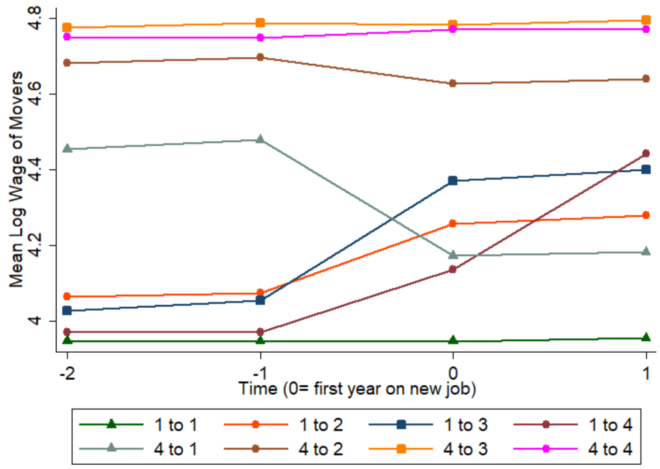
C.1 Data construction

The raw data in the LIAB comes in the form of entire job histories of workers in the sample. Individual entries therefore contain worker information, as well as information on the start and end date of a job spell for that individual, the location (establishment), and characteristics of the job spell. We transform this data into an annual panel dataset following the steps in Card, Heining, and Kline (2013), with minor modifications. Specifically, we sequentially restrict our sample by selecting (i) males in West Germany, (ii) those aged 15-64 years at the time of the job spell, and (iii) the job-spell within a calendar year with maximum earnings. We then adjust wages by (i) deflating earnings using German CPI information from FRED (Series id: DEUCPIALLMINMEI) and (ii) replacing daily wages with Upper Earnings Limits in Card, Heining, and Kline (2013) for daily wages above censor limit. Finally, we impute the district of employment using the district of the establishment if the district of employment is missing (before 1999).

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

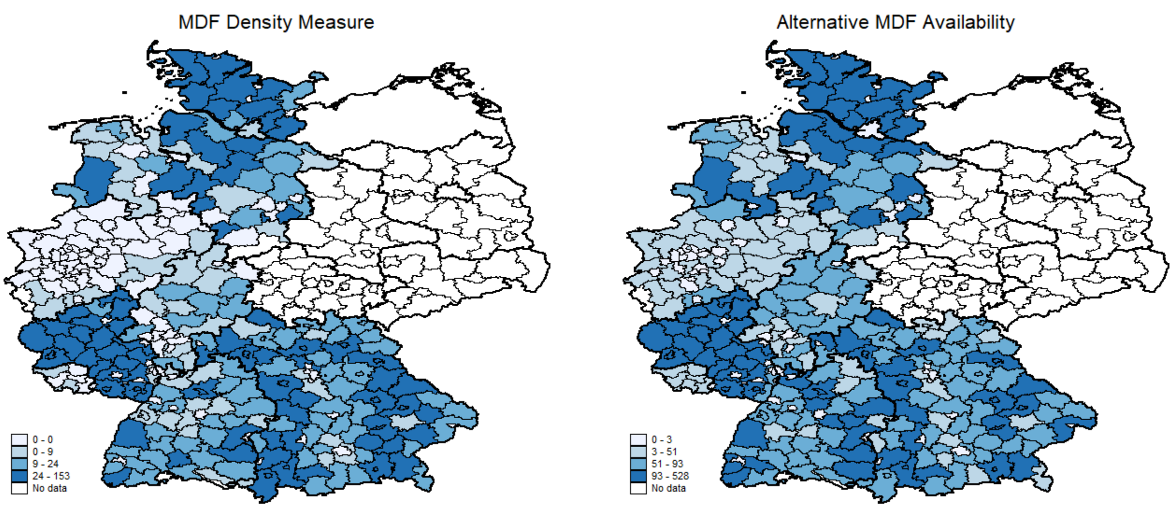
We link our LIAB-based worker panel to the DSL access data from Falck, Gold, and Heblich (2014) using the district identifiers in both datasets. We then construct the instrumental variables discussed in Section 6.2.1. Figure B2 illustrates the spatial variation of the instruments used to estimate our baseline results.

Figure B1: Replication of Card, Heining, and Kline (2013)



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

Figure B2: Spatial Variation in the Instrumental Variables



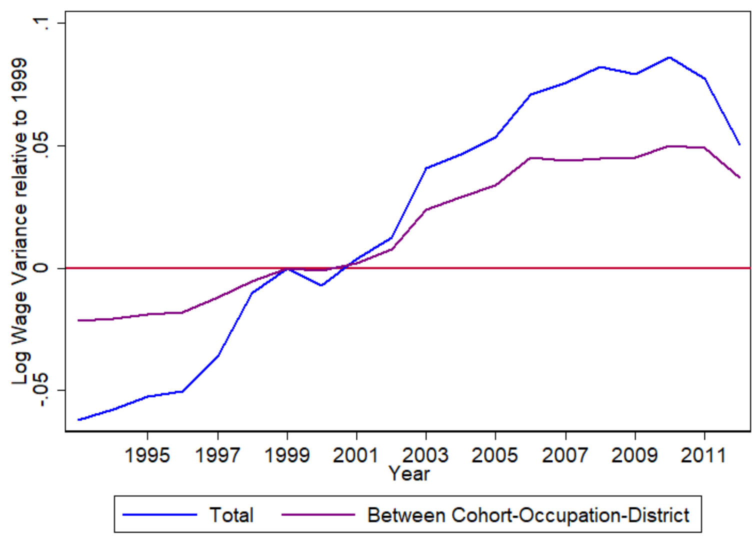
Note. Panel A illustrates the number of municipalities across districts in Germany that did not have access to an MDF within the 4200m radius ("MDF Density Measure"), as described in Section 6.2.1. Panel B illustrates the number of municipalities across districts that did not have their own MDF and did not have access to an alternative MDF in a neighboring district "Alternative MDF Availability".

C.2 Sample statistics

This section reports the summary statistics of our baseline sample. We begin by illustrating the increase in inequality, measured by the standard deviation of log wages, in our sample. Figure B3 compares the overall change in inequality together with the between district-generation-occupation component, measured using the residual log-wage dispersion from a mincer regression including dummies for the district-generation-occupation estimated on the sample in each year. Between 1997-2012, overall inequality in our sample increased by about 8.5 log points. As the figure illustrates, the between district-generation-occupation component explains about half of the increase in inequality during this period. In results available on request, we attest that each of these these characteristics alone does not account for the inequality rise.[39]

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

Figure B3: Aggregate Trends in Log Wage Variance



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

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

Table B1: Summary Statistics: German Microdata

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

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

C.3 Cognitive intensity and use of new technologies across occupations in Germany

This section analyzes the types of tasks required by cognitive-intensive occupations. Figure B4 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 B4 are consistent with the evidence establishing the heterogeneous impact of new technologies on different tasks performed by workers – e.g., Autor, Levy, and Murnane (2003), Spitz-Oener (2006), Autor and Dorn (2013), and Akerman, Gaarder, and Mogstad (2015).

We then investigate whether these new technologies affected worker generations differently conditional on their occupation. We consider two generations: a young generation aged less than 40 years and an old generation aged more than 40 years.[40] Figure B5 shows that, while internet and computer usage are biased towards cognitive-intensive occupations, there were only small differences in the usage of these new technologies across worker cohorts employed in the same occupation in 2012. These results are complement the finding in Spitz-Oener (2006) that there were small between-cohort differences in the change of the task content of German occupations in the 1990s.

C.4 Impact of new technologies on cognitive-intense occupations in Germany

This section investigates the robustness of the results presented in Section 6.2.

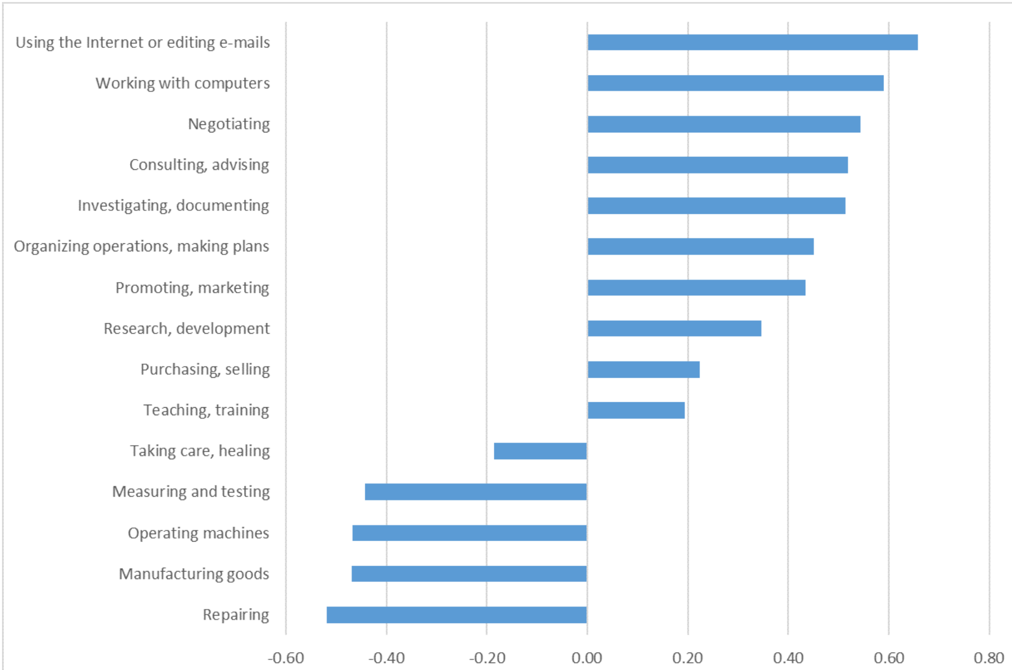
Cognitive intensity and labor market outcomes across occupations. We first report the impact of cognitive-intensity on occupation employment growth for different time horizons. The estimates in Table B2 show that results are qualitatively similar for 1995-2000 and 1995-2010.

We then investigate the impact of cognitive-intensity on occupation employment growth with a more flexible specification that allows for different coefficients for different levels of cognitive-intensity. As is clear from Table B3, the results in Table 1 are driven largely by an increase in employment for all generations in the most cognitive intensive occupations (above the 60th percentile of cognitive intensity). This increase is substantially stronger for the young generation. Some evidence of polarization is also evident for the young generation, as they also disproportionately enter the least cognitive intensive occupations.

Appendix Table B4 investigates the robustness of the estimates of equation (28) reported in Table 1. Panel A of Table B4 reports similar results when we include occupation-level controls for import and export exposure and the growth in the fraction of migrants in the occupation. Panel B shows that results are also robust to restricting the sample to native-born German males only. Panel C presents results where the "Young" generation is defined alternatively as those born after 1965 or 1955. As expected, when the definition of the young generation is further restricted to include only more recent cohorts, the coefficient on "Young" is stronger.

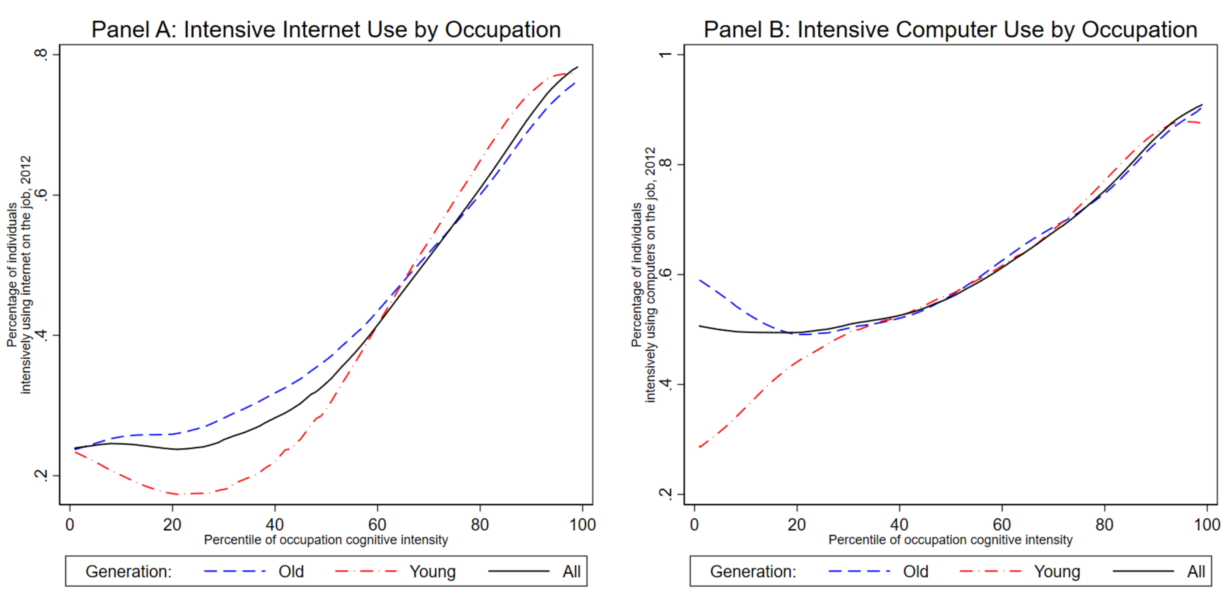
[40] Results are similar if we define young generations to include workers who are less than 30, 35 or 45 years old.

Figure B4: 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).

Figure B5: 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.

The opposite happens if we relax the young definition to include older cohorts. Panel C also shows that results are similar if the "Young" generation is defined as those aged below 40 in each year (as in Figure 5).

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

$$DSI_i = Z_i\rho + X_i\gamma + \epsilon_i \tag{C.1}$$

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

Table B5 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 (29) 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 B6. This test statistic checks for whether any of our endogenous variables are weakly instrumented, as well as whether there are sufficiently many strong instruments to instrument the multiple endogenous variables. As shown in the table, we obtain uniformly high first-stage SW F-statistics in all specifications, indicating that our instrument vector has enough power to estimate responses for different worker cohorts.

We now turn to a more careful investigation of the robustness of the results in Figures 6. Table B7 investigates how our baseline set of controls affects estimates. The three panels of Table B7 present estimates for the entire post-shock period of the sample (1999-2014, Panel A), the period during which DSL was rolled out across German regions (1999-2007, Panel B), and the period before the shock (1996-1999, Panel C). Each panel includes the results of our baseline specification, as well as alternative specifications in which (i) we drop the pre-trend control, and (ii) we augment baseline controls with district-generation-year fixed effects.

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

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

Table B8 investigates the robustness of the baseline estimates in Figures 6 to the sample specification. The two panels present estimates for the entire post-shock period of the sample (1999-2014, Panel A), and the period during which DSL was rolled out across German regions (1999-2007, Panel B). All specifications include the baseline set of controls.

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

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

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

Finally, Table B9 investigates the impact of early DSL adoption on investment in cognitive skills by young workers in the district. Specifically, it reports the estimated impact, on the growth in the number of trainees in an occupation-district, of the interaction between the occupation's cognitive intensity and the district's DSL access in equation (29). The estimates are for a single generation of working-age individuals whose employment status is a trainee or intern in each year. Our estimates suggest that regions where DSL expansion happened faster also experienced stronger growth in the number of trainees in more cognitive intensive occupations. This evidence is consistent with our model's prediction that, after the arrival of a cognitive-intensive innovation, incoming cohorts increase their investment in the cognitive skills used in more cognitive intensive occupations.

Table B2: 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 1995-2005							
Cognitive intensity	0.778***	1.150***	0.290***	0.741***	1.158***	0.404***	0.545
	(0.111)	(0.130)	(0.079)	(0.086)	(0.114)	(0.063)	(0.427)
Panel C: Change in 1995-2010							
Cognitive intensity	1.110***	1.523***	0.454***	1.036***	1.525***	0.539***	0.768*
	(0.137)	(0.149)	(0.125)	(0.111)	(0.133)	(0.091)	(0.430)
Panel D: 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

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

Dependent variable:	Employment Growth			Real Payroll Growth			Trainee Growth
	All	Young	Old	All	Young	Old	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Low: below P30	-0.012	0.286**	-0.946***	0.220*	0.590***	-0.875***	-0.790***
	(0.128)	(0.137)	(0.129)	(0.132)	(0.140)	(0.135)	(0.178)
Medium: P30-P60	-0.054	-0.046	0.031	-0.086	-0.036	-0.014	-0.112
	(0.194)	(0.205)	(0.208)	(0.195)	(0.208)	(0.202)	(0.274)
High: above P60	0.812***	1.038***	0.531***	0 .816***	1.099***	0 .592***	1.052***
	(0.156)	(0.166)	(0.016)	(0.158)	(0.169)	(0.157)	(0.237)

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

Table B4: Cognitive intensity and labor market outcomes across occupations in Germany: Robustness

Dependent variable:	Employment Growth				Trainee Growth
	All	Young	Old	Between	
	(1)	(2)	(3)	(4)	(5)
Panel A: Alternative control set, 1995-2014					
Controls for immigration and trade	1.426***	1.807***	0.841***	0.966***	2.029***
	(0.261)	(0.279)	(0.252)	(0.376)	(0.457)
Panel B: Alternative sample definition, 1995-2014					
Native-born Males Only	1.396***	1.807***	0.778***	1.029***	2.194***
	(0.226)	(0.235)	(0.231)	(0.340)	(0.385)
Panel C: Alternative generation definition, 1995-2014					
Young: Born after 1965	1.488***	2.137***	0.857***	1.280***	2.121***
	(0.225)	(0.299)	(0.246)	(0.387)	(0.385)
Young: Born after 1955	1.488***	1.639***	0 .967***	0.671*	2.121***
	(0.225)	(0.268)	(0.290)	(0.395)	(0.385)
Young: Aged Below 40 in each year	1.488***	1.748***	0.773***	0.975**	2.121***
	(0.225)	(0.294)	(0.246)	(0.383)	(0.385)

Note. Sample of 120 occupations, sample periods as defined in the table. Columns (1)–(3) report the estimated coefficient on the occupation's cognitive intensity in equation (28). Column (4) reports the difference between the coefficients in columns (3) and (2). Each row defines a separate robustness exercise. The row "Controls for immigration and trade" includes a set of baseline controls: growth in occupational exposure to exports during the sample period, growth in occupational exposure to imports during the sample period, and growth in the fraction of immigrants in the occupation during the sample period. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B5: 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 B6: First-stage SW F-statistics for estimation of equation (29) reported in Panel A of Figure 6

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 (29).

Table B7: Impact of early DSL adoption on more cognitive-intensive occupations: Alternative control sets

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

Note. Sample of 2 cohorts, 120 occupations and 323 districts. Sample periods as defined in the table. Column (1) reports the estimated coefficient on interaction between the occupation cognitive intensity and district DSL access in equation (29) for a single generation of working-age employed individuals. Columns (2)-(3) report the estimated coefficients on interaction between the occupation cognitive intensity, generation dummies and district DSL access in equation (29) for the old and young generations. Column (4) reports the difference between the coefficients in columns (3) and (2). Generations are the baseline generations with young workers those born after 1960. All regressions are weighted by the district population size in 1999 and include a set of baseline district-level controls as well as occupation-year and generation-year fixed effects. Each row defines a separate robustness exercise. "District-Year Effects" are estimated as district-year fixed effects in column (1) and as district-year-generation fixed effects in columns (2)-(4). Standard errors clustered at the district-level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B8: Impact of early DSL adoption on more cognitive-intensive occupations: Sample selection

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

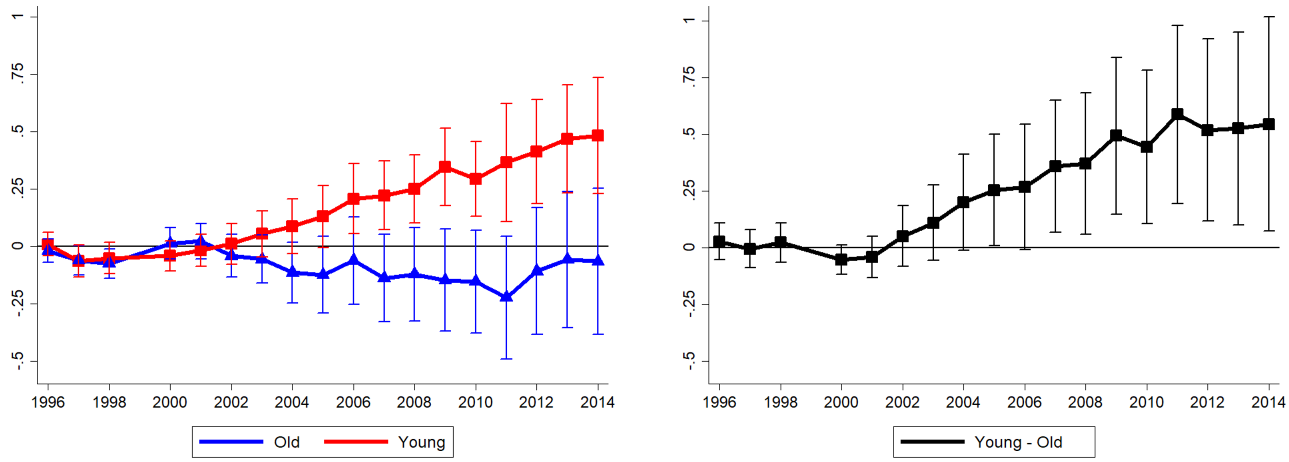
Note. Sample of 2 cohorts, 120 occupations and 323 districts. Sample periods as defined in the table. Column (1) reports the estimated coefficient on interaction between the occupation cognitive intensity and district DSL access in equation (29) for a single generation of working-age employed individuals. Columns (2)-(3) report the estimated coefficients on interaction between the occupation cognitive intensity, generation dummies and district DSL access in equation (29) for the old and young generations. Column (4) reports the difference between the coefficients in columns (3) and (2). All regressions are weighted by the district population size in 1999 and include a set of baseline district-level controls, pretrend controls, occupation-year and generation-year fixed effects. Each row defines a separate sample selection exercise: (i) baseline sample restricted to only Germans ("Native-born"), (ii) different definitions of young workers based on year of birth or age cutoff in each year, and (iii) baseline sample restricted to workers employed in establishments below the 75th percentile of all establishment sizes ("Small Establishments Only"). Standard errors clustered at the district-level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B9: Impact of early DSL adoption on the number of trainees in more cognitive-intensive occupations

Dependent variable:	Training Growth
	(1)
Panel A: 1999-2014	
Baseline	0.415*
	(0.237)
Panel B: 1999-2007	
Baseline	0.305
	(0.247)
Panel C: 1996-1999	
Baseline	-0.093
	(0.044)

Note. Sample of 120 occupations and 323 districts. Sample periods as defined in the table. Table reports the estimated coefficient on interaction between the occupation cognitive intensity and district DSL access in equation (29) for a single generation of working-age individuals whose employment status is a trainee or intern in each year. All regressions are weighted by the district population size in 1999 and include a set of baseline district-level controls as well as occupation-year fixed effects. Standard errors clustered at the district-level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

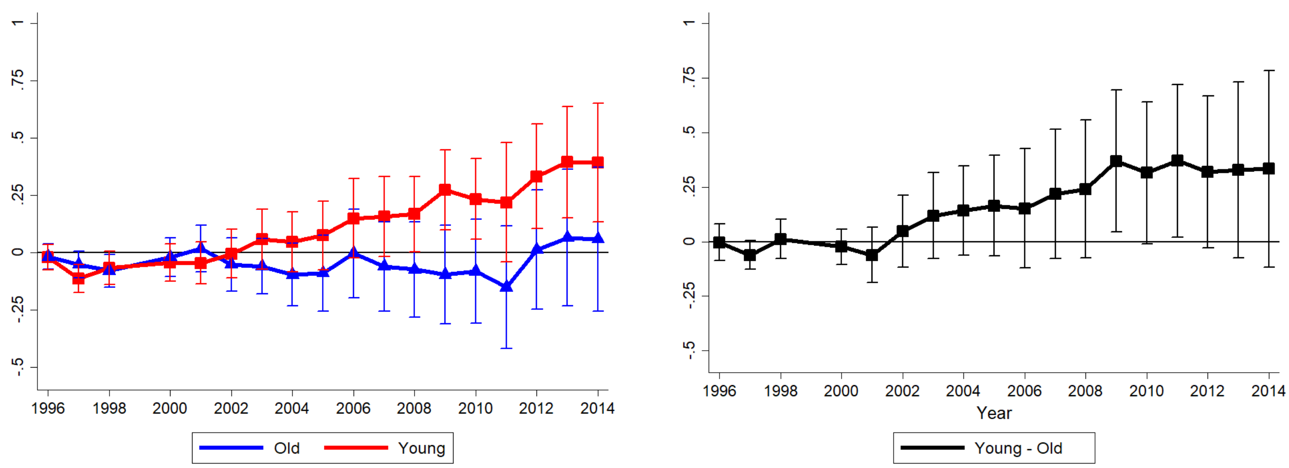
Figure B6: Impact of early DSL adoption on employment in more cognitive-intensive occupations: Old and Young generations



(a) Relative employment response for each generation (b) Between-generation response difference

Note. Estimation of equation (29) 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.

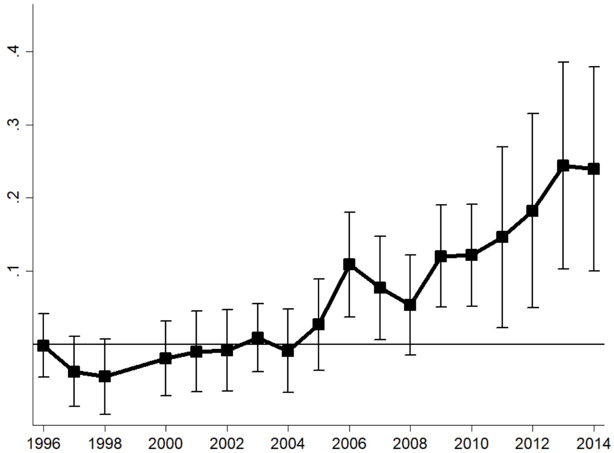
Figure B7: Impact of early DSL adoption on payroll in more cognitive-intensive occupations: Old and Young generations



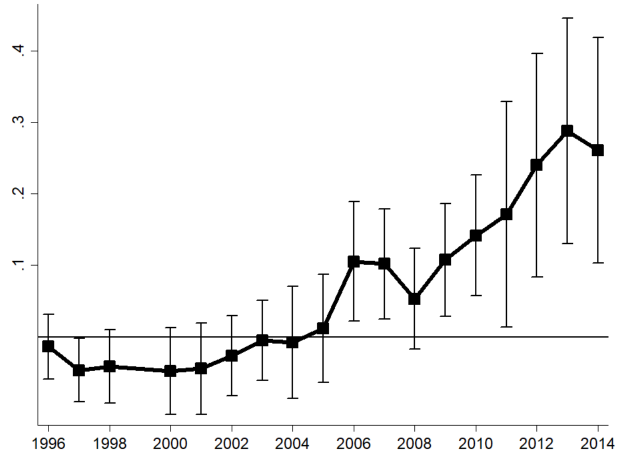
(a) Relative payroll response for each generation (b) Between-generation response difference

Note. Estimation of equation (29) 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.

Figure B8: Impact of early DSL adoption on more cognitive-intensive occupations: All generations



(a) Relative employment response for all generations



(b) Relative payroll response for all generations

Note. Estimation of equation (29) in the sample of 120 occupations and 323 districts for a single generation of working-age employed individuals. Dependent variable: log employment (left) and log payroll (right). All regressions are weighted by the district population size in 1999 and include occupation-time fixed-effects. Baseline controls include the following district variables in 1999: college graduate population share, manufacturing employment share, immigrant employment share, district age composition, and the dependent variable pre-shock growth in 1995-1999. For each year, the dot is the point estimate of β_t^{all}, and the bar is the associated 90% confidence interval implied by the standard error clustered at the district level.

C.5 Changes in mean age and employment share across occupations in the United States

This section investigates the correlation between changes in the mean age of an occupation and its employment share in the United States over different time periods. We use the U.S. Census data downloaded from IPUMS international for 1960, 1970, 1980, 1990, 2000, 2010, and 2015. It contains individual-level occupation information for males aged 16-64 years old in the nine 2-digit ISCO occupation used to construct the aggregate trends reported in Section 6.1. For each occupation o, we 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 B10 shows that, in line with Figure 5, 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, our results show that this correlation was positive in 1960-1980.

Table B10: 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.

Appendix D Numerical Analysis

This appendix discusses in detail the parameterization of the model. Section D.1 presents the theoretical impulse response functions for the relative employment of different worker generations. Section D.2 describes the procedure to select the parameters that match the theoretical and empirical impulse response functions. In Section D.3, we use the parametrized model to quantitatively evaluate the dynamic adjustment to cognitive-biased technological innovations.

D.1 Impulse response functions of relative employment by generation

As a first step to parametrize our theory using the empirical impulse response functions in Section 6, we derive the theoretical responses of generation-specific relative employment. To this end, we consider the same one-time permanent change in A at $t = 0$. We define older generations as those born before period $t = -x$ and younger generations as those born at period $t = -x$. In period $t \geq 0$, the relative high-tech employment of these worker generations are given by

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

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

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

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

Relative employment of old generation. We show below that the change in the relative employment of old generations is

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

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

Among old generations, the increase in the relative productivity of high-tech production induces the reallocation of workers towards high-tech production whenever $\theta > 1$. The expression indicates that this positive effect on relative high-tech employment becomes weaker over time. This follows from the expansion of high-i skills among younger generations, which displaces old workers with marginal skills from high-tech production – i.e., those with skills $i \in (l_0, l_\infty)$. Importantly, expression (D.1) shows that the magnitude of the increase in relative employment of older generations is decreasing in the degree of technology-skill specificity (i.e., increasing in η).

Relative employment of young generation. Turning to the employment response among young generations, we show below that

$$\Delta \log e_t^{young} \approx \Delta \log e_t^{old} + \frac{\psi}{\chi} \frac{1 - e^{-\lambda t}}{1 - (1 - \tilde{x}_0)e^{-\delta t}} (\theta - 1)\Delta \log A. \tag{D.2}$$

This expression indicates that the evolution of the allocation of young workers has two components. The first term captures the change in technology-skill assignment and, since it is the only determinant of the relative employment of old generations, it can be approximated by $\Delta \log e_t^{old}$. The second term captures the change in the skill investment decision of incoming cohorts. At each point in time, this term is positive as young workers distort skill investment towards high-i skills that became more valuable in high-tech production. We can also show that the between-generation difference grows shortly after the shock. Importantly, expression (D.1) indicates that the between-generation difference in the response of relative employment is decreasing in the skill investment cost (i.e., it is increasing in ψ).

D.1.1 Proof of equations (D.1)–(D.2)

Proof of equation (D.1). We first use a first-order approximation to write the log-change in relative high-tech employment in terms of changes in the high-tech employment share:

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

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

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

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

We have that

$$e^{old}_{H,t}-e^{old}_{H,0^-}=\int_{l_t}^{1}s_0(i)di-\int_{l_{0^-}}^{1}s_0(i)di$$

By approximating these expressions around l,

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

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

Combining the two expressions,

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

Using the demand expression in (2),

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

Using the expression for the evolution of y_t in Proposition 1,

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

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

which is identical to (D.1).

Proof of equation (D.2). We first use a first-order approximation to write the log-change in relative high-tech employment in terms of changes in the high-tech employment share:

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

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

Using the definition of employment shares for each generation,

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

Thus,

$$\log(\frac{e_t^{young}}{e_{0^-}^{young}}) - \log(\frac{e_t^{old}}{e_{0^-}^{old}}) \approx \frac{1}{(1-e_{H,\infty})e_{H,\infty}} \frac{1}{1-(1-\tilde{x}_0)e^{-\delta t}} \left(\delta \int_0^t e^{\delta(\tau - t)} \int_{l_t}^1 \left(\tilde{s}_\tau(i) - s_0(i) \right) di d\tau \right) \tag{D.3}$$

We now consider the following approximation:

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

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

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

Using the third part of Theorem 1,

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

We now apply this expression into (D.3):

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

Notice that Proposition 1 implies that

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

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

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

Thus,

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

which is equivalent to (D.2).

D.2 Parameterization by impulse response matching

We now describe how to parametrize our theory to match the empirical impulse response functions in Section 6. To this end, we map the H technology in our theory to the set of production activities performed by cognitive-intensive occupations. We calibrate our theory in two steps. In the first step, we exogenously specify a subset of parameters and functions in the theory. We set the discount rate to match an annual interest rate of 2%, $\rho = 0.02$. We calibrate the elasticity of substitution across cognitive and non-cognitive intensive occupations to $\theta = 3$. Finally, for all welfare calculations, we specify welfare-weights re^{-rt} with $r = \rho + \delta$ so that the social discounting of future generations is identical to the discounting of worker's future utility.

We also specify functional forms for the productivity of skill types in the two technologies. We abstract from differences in non-cognitive productivity across skills by normalizing $\alpha(i) \equiv 1$. This implies that, for any given worker generation, employment and payroll responses are both driven by the degree of technology-skill specificity in the economy.[41] In addition, we

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

assume that $\sigma(i)$ takes the form of a logistic function:

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

where l is the assignment threshold in the initial stationary equilibrium. This specification is a tractable manner of capturing technology-skill specificity in the economy. It implies that the equilibrium exists for any $\sigma > 0$ since the relative productivity is bounded. Also, by setting the midpoint of the function to l, the parameter σ controls the elasticity of $\sigma(i)$ for the marginal skill types in the initial equilibrium (i.e., i close to l). Thus, σ specifies the magnitude of technology-skill specificity, $1/\eta$.

In the second step, we use the estimated responses of Section 6 to calibrate (δ, σ, ψ). In doing so, we select the distribution of innate ability to normalize the initial skill distribution to be uniform: $s_0(i) \equiv 1$.[42] We formally present the parametrization procedure next, along with an analysis of the model fit. For all parameters, we assume that the shock starts with the roll-out of broadband internet in 2003. We then select parameters to match the estimates for the period of 2008 to 2014 in which we find statistically significant response in the relative payroll and relative employment of cognitive-intensive occupations.

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

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

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

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

Rate of convergence: λ. Proposition 1 implies that it is possible to write the impulse response function of relative output as

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

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

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

$$\hat{\lambda} = arg \min_{\lambda} \sum_{t=2008}^{2014} \left[\left(\hat{\beta}^y_t - \hat{\beta}^y_{2007} \right) - \alpha_1 e^{-\lambda(t-2007)} \right]^2 \tag{D.4}$$

where $\hat{\beta}^y_t$ are the estimated coefficient reported in Panel B of Figure 6.

The minimization problem in (D.4) yields $\hat{\lambda} = 0.135$. Figure C1 shows the fit of the calibrated model

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

Cost of skill investment: ψ. Theorem 1 implies that

$$\kappa\eta = \psi\hat{\alpha} - \theta \tag{D.5}$$

where

$$\alpha = \delta\left[\left(\frac{\rho}{2}+\hat{\lambda}\right)^2 - \left(\frac{\rho}{2}\right)^2 - \delta(\rho+\delta)\right]^{-1}. \tag{D.6}$$

Using expression (D.2), we have that

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

From Proposition 1,

$$(\theta-1)\,\Delta log(A) = \Delta \log(y_t)\left(\frac{1+\kappa\eta}{\theta+\kappa\eta} + \frac{\psi}{\chi}\frac{\theta-1}{\theta+\kappa\eta}(1-e^{-\lambda t})\right)^{-1} \tag{D.7}$$

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

Combining these two expressions, we get that

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

Using the expression for $\kappa\eta$ in (D.5),

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

We then define the function:

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

To calibrate ψ, we first our calibrated values of (λ,δ,ρ) to compute α using (D.6). Our baseline calibration implies that $\hat{\alpha} = 3.484$. We then select the parameter ψ to match the ratio of the between-generation employment response and the payroll response:

$$\hat{\psi} = arg\min_{\psi}\sum_{t=2008}^{2014}\left[\frac{\hat{\beta}_t^{young} - \hat{\beta}_t^{old}}{\hat{\beta}_t^{y}} - F^{\psi}(\psi,t)\right]^2 \tag{D.8}$$

where $\hat{\beta}_t^{y}$ are the estimated coefficients reported in Panel B of Figure 6, and $\hat{\beta}_t^{young} - \hat{\beta}_t^{old}$ is the between-generation employment response obtained with the estimated coefficients reported in Panel A of Figure 6.

The minimization problem in (D.8) yields $\hat{\psi} = 0.345$. Figure C2 shows the fit of the calibrated model.

Technology-skill specificity: η. The combination of (D.1) and (D.7) implies that

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

Using the expression for $\kappa\eta$ in (D.5),

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

We then define

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

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

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

$$\hat{\eta} = arg\min_{\eta} \sum_{t=2008}^{2014} \left[\frac{\hat{\beta}_t^{old}}{\hat{\beta}_t^{y}} - F^\eta(\eta, t) \right]^2 \tag{D.9}$$

$\hat{\beta}_t^y$ are the estimated coefficients reported in Panel B of Figure 6, and $\hat{\beta}_t^{old}$ are the estimated coefficients reported in Panel A of Figure 6.

The negative point estimates reported in Panel A of Figure 6 imply that the minimization problem in (D.9) yields $\hat{\eta} < 0$. Since the employment response of old generations is small and nonsignificant, we assume that they are identical to zero, which yields $\hat{\eta} = 0$. Hence, we calibrate $\eta = 0.01$ and evaluate the model predictions under alternative specifications of this parameter.

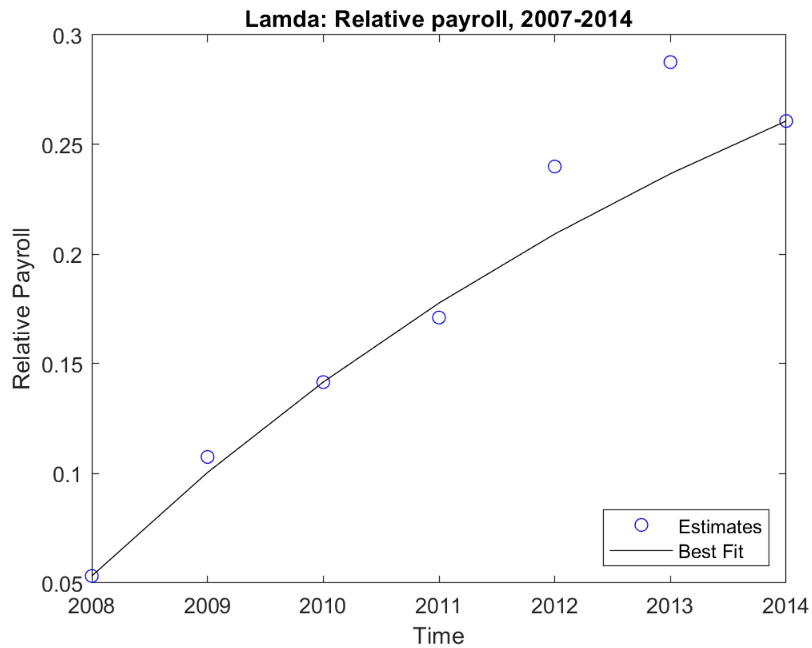


Figure C1: Calibration of λ

Note. Blue dots represent the point estimates reported in Panel B of Figure 6. Black solid curve represents the bet fit line with $\lambda = 0.135$ obtained from the solution of (D.4).

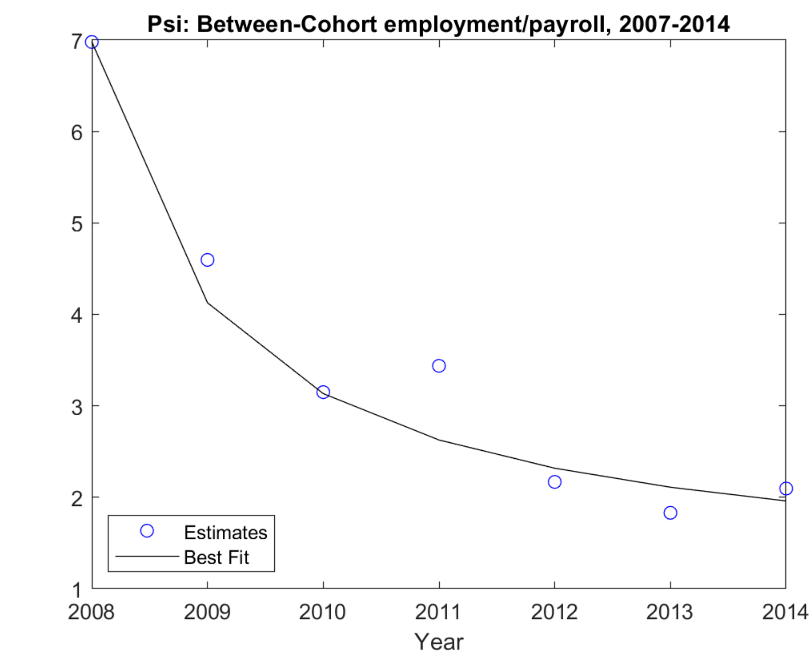


Figure C2: Calibration of ψ

Note. Blue dots represent the point estimates of $\frac{\hat{\beta}_t^{young} - \hat{\beta}_t^{old}}{\hat{\beta}_t^{y}}$ using the estimates reported in Figures 6. Black solid curve corresponds to $F^{\psi}(\hat{\psi}, t)$ with $\hat{\psi} = 0.354$ obtained from the solution of (D.8).

D.3 Dynamic Adjustment to cognitive-biased technological innovations

We now present the theoretical impulse response functions in our parametrized model. As in Section 7, we evaluate a shock to A that leads to an increase in the employment share in cognitive-intensive occupations from 20% to 50%. Figure C3 presents the predicted impulse response functions of labor market outcomes.

Consider first the response at $t = 0$. Given that our theory abstracts from several additional sources of dynamics, it would be wrong to interpret the impact adjustment as happening instantaneously in reality. We view this short-run response as capturing changes over the time window encompassing dynamic forces triggered by other variables that are likely to move faster than the distribution of skills (e.g., physical capital). In other words, we prefer to interpret the "length" of the impact adjustment as related to the time that it takes for such faster moving variables to converge to the new long-run equilibrium. Results show that there is a substantial increase in the relative cognitive-intensive output in the short-run. This large response is a consequence of the large magnitude of the shock. This becomes clear when we take into account that relative employment almost does not change at impact because of the high technology-skill specificity (i.e., $\eta \approx 0$). The combination of the large increase in relative output and the small increase in relative employment translates into large changes in lifetime inequality.

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

Appendix E Additional Results

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

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

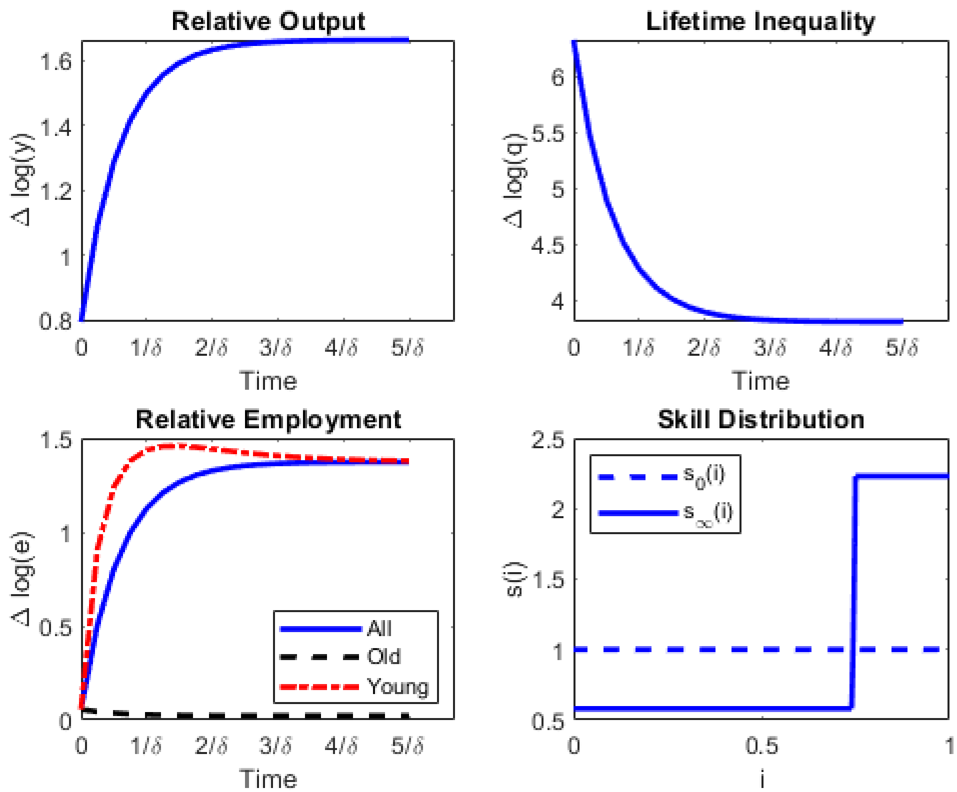
$$X_{kt} = \int_0^1 x_{kt}(i)s_{kt}(i)di,$$

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

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

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

Figure C3: Transitional dynamics to a cognitive-biased innovation at $t = 0$



Note. The figure reports the theoretical impulse response function with a shock calibrated to increase the employment share in cognitive-intensive occupations from 20% to 50% between stationary equilibria. Baseline calibration described in Appendix D.2.

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

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

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

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

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

E.2 Welfare Consequences of Adjustment Across Generations

This section investigates how calculations of the welfare consequences of technological shocks are affected by the speed of adjustment of labor market outcomes along the transition to the new equilibrium. In our theory, the transitional dynamics arise from the changes in the skill distribution. So, in order to evaluate its consequences, we consider a static version of our model in which we shut down any skill investment of young workers. However, we allow this static model to match labor market responses over one particular horizon. This exercise thus speaks directly to the risks of ignoring the adjustment across generations by focusing on estimates of the impact of new technologies on labor market outcomes over fixed time horizons.

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

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

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

Proposition 4 *Consider an economy in which η and ψ are positive. Assume that $\Delta \log A$ generates a change in lifetime inequality between $t = 0$ and $t = T$ of $\Delta \log q_T$. Consider predictions under two alternative static parametrizations of the model ($\psi^1 = \psi^2 = 0$).*

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

Proof. See Appendix E.2.1. ■

This proposition shows that there are multiple ways in which researchers can use a static version of our model to match observed inequality responses over a fixed horizon. All versions ignore the transitional dynamics of labor market outcomes generated by changes in the skill distribution across generations. This introduces biases in the evaluation of the welfare consequences of new technologies. If the researcher only matches inequality responses in short horizons (i.e. T is low), then she will think that inequality will remain high in the future. This makes her overpredict the present value of lifetime inequality, and underpredict the average welfare gain. Alternatively, a researcher using the first approach would reach the opposite conclusions if she matches inequality responses in long horizons (i.e. T is high).

Such biases will be larger when the adjustment is slower due to the larger changes in the skill distribution along the transition. As shown in Section 4, this is the case whenever the skill supply elasticity is low in the short-run (i.e, η is low) but large in the long-run (i.e., ψ is large).

E.2.1 Proof of Proposition 4

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

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

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

Notice also that the combination of Theorem 1 and Proposition 2 implies that

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

and, therefore,

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

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

$$\Delta\bar{\Omega}^P = (\rho + \delta)\Delta \log(q_T) \tag{E.3}$$

for $P = 1, 2$.

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

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

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

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

By taking the ratio between the expressions in (E.2) and (E.3),

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

The expression of $\Delta\bar{U}$ in Proposition 2 immediately implies that $\Delta\bar{\Omega}^1 > \Delta\bar{\Omega} \iff \Delta\bar{U}^1 < \Delta\bar{U}$ whenever $y_\infty > e_\infty$.

Part 2. In the second approach, we set $\eta^2 = \eta$. So, by equation (E.1), we must set .

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

Expressions in (E.2) and (E.3) also hold in this case, so the same steps used above guarantee that $\Delta\bar{\Omega}^2 > \Delta\bar{\Omega}$ if, and only if, $T < T^1$. To establish the result, it is sufficient to show that $\Delta \log A^2 \leq \Delta \log A$ because, by Proposition 2, $\Delta\bar{\Omega}^2 > \Delta\bar{\Omega}$ and $\Delta \log A^2 \leq \Delta \log A$ imply that $\Delta\bar{U}^2 < \Delta\bar{U}$.

We now show that $\Delta \log A^2 \leq \Delta \log A$. By combining Proposition 1 and equation (E.1), we have that

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

and, therefore,

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

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

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

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

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

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