Online Appendix (For Online Publication Only)

Table of Contents

A Proofs

A.1 Proof of Lemma 1: Optimal Labor Allocation

The planner's problem is

$$V^*\left(\{q_{i0}\}\right) \equiv \max_{\{\ell_{it},s_{it}\}} \int_0^\infty e^{-\rho t} \sum_{i=1}^{K} \beta_i \ln y_{it} \, \mathrm{d}t,$$

subject to constraints (3), (4), (5), and (6). Substituting using (3) and (4), the objective can be re-written as

$$V^*\left(\{q_{i0}\}\right) \equiv \max_{\{\ell_{it},s_{it}\}} \int_0^\infty e^{-\rho t} \sum_{i=1}^{K} \beta_i \ln q_{it}^{\psi} \ell_{it} \, \mathrm{d}t.$$

The FOC with regard to ℓ_{it} gives: $\frac{\beta_i}{\ell_{it}} = \frac{\beta_j}{\ell_{jt}}$. Therefore, for all t, $\ell_{it} = \beta_i \bar{\ell}$ for each sector i.

A.2 Proof of Proposition 1: Optimal R&D Allocation in the Baseline Model

The social planner's problem is

$$\begin{gathered}\max_{\{\gamma_t\} \text{ s.t. } \gamma_t' 1 = 1 \forall t} \int_0^\infty e^{-\rho t} \boldsymbol{\beta}' \ln \boldsymbol{q}_t \, \mathrm{d}t \\ \text{s.t. } \mathrm{d} \ln \boldsymbol{q}_t / \mathrm{d}t = \lambda \cdot \left(\ln \boldsymbol{\eta} + \ln \bar{s} + \ln \boldsymbol{\gamma}_t + (\boldsymbol{\Omega} - \boldsymbol{I}) \ln \boldsymbol{q}_t\right),\end{gathered}$$

The control variable is $\boldsymbol{\gamma}_t$ and the state variable is $\boldsymbol{q}_t$. Denote the co-state variables as $\boldsymbol{\mu}_t$. The current-value Hamiltonian writes

$$H(\gamma_t, \boldsymbol{q}_t, \boldsymbol{\mu}_t, \zeta) = \boldsymbol{\beta}' \ln \boldsymbol{q}_t + \lambda \boldsymbol{\mu}_t' \left(\ln \boldsymbol{\eta} + \ln \bar{s} + \ln \boldsymbol{\gamma}_t + (\boldsymbol{\Omega} - \boldsymbol{I}) \ln \boldsymbol{q}_t\right) + \zeta(1 - \boldsymbol{\gamma}_t' \mathbf{1}).$$

For notational simplicity we suppress dependence on time for the control, state, and co-state variables:

$$\begin{aligned} H(\{\gamma_i\}, \{q_i\}, \{\mu_i\}, \zeta, t) &= \sum_i \beta_i \ln q_i + \zeta(1 - \sum_i \gamma_i) \\ &\quad + \lambda \sum_i \mu_i \left(\ln \eta_i + \ln \bar{s} + \ln \gamma_i + \sum_j \omega_{ij} \ln q_j - \ln q_i \right) \end{aligned}$$

By the maximum principle

$$H_{\gamma_i} = 0 \Longleftrightarrow \frac{\lambda \mu_i}{\gamma_i} = \zeta \quad \forall i \tag{A1}$$

$$H_{\ln q_i} = \rho\mu_i - \dot{\mu}_i \Longleftrightarrow \beta_i - \lambda\mu_i + \lambda \sum_j \mu_j \omega_{ji} = \rho\mu_i - \dot{\mu}_i \tag{A2}$$

First, we show that the transversality condition $\lim_{t\to\infty} e^{-\rho t} H(\{\gamma_i\}, \{q_i\}, \{\mu_i\}, \zeta, t) = 0$ implies $\dot{\mu}_i = 0$ for all i. It is then immediate that the optimal R&D allocation $\boldsymbol{\gamma}$ is time invariant.

Note the matrix formula of equation (A2) is

$$\dot{\boldsymbol{\mu}}_t = \left[(\rho+\lambda)\boldsymbol{I} - \lambda\boldsymbol{\Omega}'\right] \boldsymbol{\mu}_t - \boldsymbol{\beta} \tag{A3}$$

Then

$$\begin{aligned} \boldsymbol{\mu}_t &= e^{[(\rho+\lambda)\boldsymbol{I}-\lambda\boldsymbol{\Omega}']t}\boldsymbol{\mu}_0 - \left(\int_0^t e^{[(\rho+\lambda)\boldsymbol{I}-\lambda\boldsymbol{\Omega}'](t-s)}\,\mathrm{d}s\right)\boldsymbol{\beta} \\ &= e^{[(\rho+\lambda)\boldsymbol{I}-\lambda\boldsymbol{\Omega}']t}\boldsymbol{\mu}_0 - \left(e^{[(\rho+\lambda)\boldsymbol{I}-\lambda\boldsymbol{\Omega}']t} - \boldsymbol{I}\right)\left[(\rho+\lambda)\boldsymbol{I} - \lambda\boldsymbol{\Omega}'\right]^{-1}\boldsymbol{\beta}. \end{aligned}$$

By transversality,

$$\begin{aligned} 0 &= \lim_{t\to\infty} e^{-\rho t}\boldsymbol{\mu}_t \\ &= \lim_{t\to\infty} e^{[\lambda(\boldsymbol{I}-\boldsymbol{\Omega}')]t}\left[\boldsymbol{\mu}_0 - \left[(\rho+\lambda)\boldsymbol{I} - \lambda\boldsymbol{\Omega}'\right]^{-1}\boldsymbol{\beta}\right]. \end{aligned}$$

Hence it must be the case that $\boldsymbol{\mu}_0 = \left[(\rho+\lambda)\boldsymbol{I} - \lambda\boldsymbol{\Omega}'\right]^{-1}\boldsymbol{\beta}$. Plugging it to the explicit solution of $\boldsymbol{\mu}_t$ and then back to (A3), we can get $\dot{\boldsymbol{\mu}}_t = 0$. Hence $\boldsymbol{\mu}_t$ and $\boldsymbol{\gamma}_t$ are time invariant.

We then can calculate $\boldsymbol{\gamma}$. First obtain $\boldsymbol{\mu}$ directly from FOC (A3):

$$(\rho+\lambda)\boldsymbol{\mu}_t'\left(\boldsymbol{I} - \frac{\boldsymbol{\Omega}}{1+\rho/\lambda}\right) = \boldsymbol{\beta}' \Longleftrightarrow \boldsymbol{\mu}_t' = \frac{1}{\rho+\lambda}\left(\boldsymbol{I} - \frac{\boldsymbol{\Omega}}{1+\rho/\lambda}\right)^{-1}.$$

According to Equation (A1), $\boldsymbol{\gamma}$ is proportional to $\boldsymbol{\mu}$ and subject to $\sum_i \gamma_i = 1$. We can then find $\boldsymbol{\gamma}$:

$$\boldsymbol{\gamma}' = \frac{\rho}{\rho+\lambda}\boldsymbol{\beta}'\left(\boldsymbol{I} - \frac{\boldsymbol{\Omega}}{1+\rho/\lambda}\right)^{-1},$$

since

$$\begin{aligned}
\frac{\rho}{\rho+\lambda}\boldsymbol{\beta}'\left(\boldsymbol{I}-\frac{\boldsymbol{\Omega}}{1+\rho/\lambda}\right)^{-1}\mathbf{1} &= \frac{\rho}{\rho+\lambda}\boldsymbol{\beta}'\left(\sum_{s=0}^{\infty}\left(\frac{\boldsymbol{\Omega}}{1+\rho/\lambda}\right)^{s}\mathbf{1}\right) \\
&= \frac{\rho}{\rho+\lambda}\sum_{s=0}^{\infty}\left(\frac{1}{1+\rho/\lambda}\right)^{s} \\
&= 1,
\end{aligned}$$

as desired.

A.3 Proof of Lemma 2: Economic Growth Rate Along a Balanced Growth Path

Consider a BGP in which R&D allocation shares follow the vector $\boldsymbol{b}$ and the growth rate of sectoral knowledge stock is time-invariant. The law of motion for stock vector is

$$\mathrm{d}\ln\boldsymbol{q}_t/\mathrm{d}t = \lambda\cdot(\ln\boldsymbol{\eta}+\ln\bar{s}+\ln\boldsymbol{b}+(\boldsymbol{\Omega}-\boldsymbol{I})\ln\boldsymbol{q}_t).$$

Taking derivative with respect to time,

$$\mathbf{0} = \lambda\left(\boldsymbol{\Omega}-\boldsymbol{I}\right)\frac{\mathrm{d}\ln\boldsymbol{q}_t}{\mathrm{d}t},$$

implying that the vector of sectoral growth rates $\frac{\mathrm{d}\ln\boldsymbol{q}_t}{\mathrm{d}t}$ is the right-Perron eigenvector of $\boldsymbol{\Omega}$. Because $\boldsymbol{\Omega}$ is a row-stochastic matrix, this implies that $\frac{\mathrm{d}\ln\boldsymbol{q}_t}{\mathrm{d}t}$ must be a constant vector, meaning the knowledge stock in every sector must grow at the same rate $g^q(\boldsymbol{b})$. Hence,

$$g^q(\boldsymbol{b})\mathbf{1} = \frac{\mathrm{d}\ln\boldsymbol{q}_t}{\mathrm{d}t} = \lambda\cdot\left(\ln\boldsymbol{\eta}+\ln\bar{s}+\ln\boldsymbol{b}+(\boldsymbol{\Omega}-\boldsymbol{I})\ln\boldsymbol{q}_t\right). \tag{A4}$$

Left-multiply by the centrality $\boldsymbol{a}'$ of $\boldsymbol{\Omega}$ on both sides:

$$\begin{aligned}
g^q(\boldsymbol{b}) &= \boldsymbol{a}'\cdot g(\boldsymbol{b})\,\mathbf{1} \\
&= \lambda\cdot(\boldsymbol{a}'\ln\boldsymbol{\eta}+\boldsymbol{a}'\cdot\mathbf{1}\ln\bar{s}+\boldsymbol{a}'\ln\boldsymbol{b}+\boldsymbol{a}'(\boldsymbol{\Omega}-\boldsymbol{I})\ln\boldsymbol{q}_t) \\
&= \lambda\cdot(\boldsymbol{a}'\ln\boldsymbol{\eta}+\ln\bar{s}+\boldsymbol{a}'\ln\boldsymbol{b}) \\
&= \text{const}+\lambda\cdot\boldsymbol{a}'\ln\boldsymbol{b}.
\end{aligned}$$

The third equation is based on the properties of the innovation centrality vector: $\boldsymbol{a}' = \boldsymbol{a}'\boldsymbol{\Omega}$ and $\sum_{i=1}^{K} a_i = 1$. That $g^y(\boldsymbol{b}) = \psi\cdot g^q(\boldsymbol{b})$ is immediate from the production function $y_i = q_i^{\psi}\ell_i$.

A.4 Proof of Proposition 2

Starting from $\boldsymbol{\gamma}' = \frac{\rho}{\rho+\lambda}\boldsymbol{\beta}'\left(\boldsymbol{I} - \frac{\boldsymbol{\Omega}}{1+\rho/\lambda}\right)^{-1}$, right-multiply both sides by $\frac{\rho+\lambda}{\lambda}\left(\boldsymbol{I} - \frac{\boldsymbol{\Omega}}{1+\rho/\lambda}\right)$ to get

$$\boldsymbol{\gamma}'\left(\frac{\rho+\lambda}{\lambda}\boldsymbol{I} - \boldsymbol{\Omega}\right) = \frac{\rho}{\lambda}\boldsymbol{\beta}' \quad \Longleftrightarrow \quad \boldsymbol{\gamma}'\left(\boldsymbol{I} - \boldsymbol{\Omega}\right) + \frac{\rho}{\lambda}\left(\gamma' - \boldsymbol{\beta}'\right) = \boldsymbol{0}'.$$

Taking the limit as $\rho/\lambda \to 0$, $\gamma'\left(\boldsymbol{I} - \boldsymbol{\Omega}\right) \to \boldsymbol{0}$ implies $\boldsymbol{\gamma} \to \boldsymbol{a}$; taking the limit as $\rho/\lambda \to \infty$, $\boldsymbol{\gamma} \to \boldsymbol{\beta}$, as desired.

A.5 Proof of Proposition 3: Welfare Impact of R&D Reallocation

The law of motion for knowledge stock $\ln \boldsymbol{q}$ under R&D allocation $\boldsymbol{b}$ is

$$\frac{\mathrm{d}\ln \boldsymbol{q}}{\mathrm{d}t} = \lambda\left(\ln \boldsymbol{\eta} + \ln \bar{s} \cdot \boldsymbol{1} + \ln \boldsymbol{b} + \left(\boldsymbol{\Omega} - \boldsymbol{I}\right)\ln \boldsymbol{q}\right)$$

Let $\boldsymbol{a}$ denote the left-eigenvector centrality of $\boldsymbol{\Omega}$ (normalized to sum to one). We separately analyze $\boldsymbol{a}'\ln \boldsymbol{q}_t$, i.e., the centrality-weighted average knowledge stock, and the deviation of knowledge stock from this average, $\left(\boldsymbol{I} - \boldsymbol{1}\boldsymbol{a}'\right)\ln \boldsymbol{q}_t$.[11] We first show the former always grows at a constant rate even away from a BGP, whereas the latter converges to a constant vector as the economy converges to a BGP.

From the law of motion, we know

$$\begin{aligned} \boldsymbol{a}'\frac{\mathrm{d}\ln \boldsymbol{q}}{\mathrm{d}t} &= \lambda\left(\boldsymbol{a}'\ln \boldsymbol{\eta} + \ln \bar{s}' \cdot \boldsymbol{a}'\boldsymbol{1} + \boldsymbol{a}'\ln \boldsymbol{b} + \boldsymbol{a}'\left(\boldsymbol{\Omega} - \boldsymbol{I}\right)\ln \boldsymbol{q}\right) \\ &= \lambda\boldsymbol{a}'\left(\ln \boldsymbol{\eta} + \ln \bar{s} \cdot \boldsymbol{1} + \ln \boldsymbol{b}\right) \end{aligned}$$

Hence, given time-invariant R&D allocation $\boldsymbol{b}$, $\boldsymbol{a}'\ln \boldsymbol{q}_t$ always grows at a constant rate (and it equals to the rate of growth along a BGP) and can be solved in closed-form:

$$\boldsymbol{a}'\ln \boldsymbol{q}_t = \boldsymbol{a}'\ln \boldsymbol{q}_0 + \lambda\boldsymbol{a}'\left(\ln \boldsymbol{\eta} + \ln \bar{s} \cdot \boldsymbol{1} + \ln \boldsymbol{b}\right)t$$

Note that $\boldsymbol{a}'\boldsymbol{1} = 1$; hence $\left(\boldsymbol{I} - \boldsymbol{1}\boldsymbol{a}'\right)\left(\ln \bar{s} \cdot \boldsymbol{1}\right) = \boldsymbol{0}$. Let $\boldsymbol{A} \equiv \boldsymbol{1}\boldsymbol{a}'$. Note that the row-stochastic matrix $\boldsymbol{\Omega}$ represents a Markov chain, for which $\boldsymbol{a}$ is the stationary distribution, and $\boldsymbol{A} \equiv \lim_{s\to\infty}\boldsymbol{\Omega}^s$. Also note that

$$\begin{aligned} \left(\boldsymbol{I} - \boldsymbol{A}\right)\left(\boldsymbol{\Omega} - \boldsymbol{I}\right) &= \left(\boldsymbol{\Omega} - \boldsymbol{I}\right)\left(\boldsymbol{I} - \boldsymbol{A}\right) \\ &= -\left(\boldsymbol{I} - \boldsymbol{\Omega} + \boldsymbol{A}\right)\left(\boldsymbol{I} - \boldsymbol{A}\right) \end{aligned}$$

[11] We separate these two objects because, the matrix $\left(\boldsymbol{I} - \boldsymbol{\Omega}\right)$ is not invertible, but $\left(\boldsymbol{I} - \boldsymbol{\Omega} + \boldsymbol{1}\boldsymbol{a}'\right)$ generically is. The proof shown below utilizes the invertibility of $\left(\boldsymbol{I} - \boldsymbol{\Omega} + \boldsymbol{1}\boldsymbol{a}'\right)$ to solve for $\left(\boldsymbol{I} - \boldsymbol{1}\boldsymbol{a}'\right)\ln \boldsymbol{q}_t$.

Left-multiply the law of motion by $(\boldsymbol{I}-\boldsymbol{A})$, substitute the above, and let $\widetilde{\ln \boldsymbol{q}_t} \equiv (\boldsymbol{I}-\boldsymbol{A}) \ln \boldsymbol{q}_t$, we get

$$\frac{\mathrm{d}\widetilde{\ln \boldsymbol{q}_t}}{\mathrm{d}t} = \lambda (\boldsymbol{I}-\boldsymbol{A}) (\ln \boldsymbol{\eta} + \ln \boldsymbol{b}) - \lambda (\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A}) \widetilde{\ln \boldsymbol{q}_t}$$

We can integrate the ODE system:

$$\begin{aligned}
\widetilde{\ln \boldsymbol{q}_t} &= e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t} \left[\widetilde{\ln \boldsymbol{q}_0} + \lambda \int_0^t e^{\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})s} (\boldsymbol{I}-\boldsymbol{A}) (\ln \boldsymbol{\eta} + \ln \boldsymbol{b})\ \mathrm{d}s \right] \\
&= e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t} \widetilde{\ln \boldsymbol{q}_0} + (\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})^{-1} \left(\boldsymbol{I} - e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t} \right) (\boldsymbol{I}-\boldsymbol{A}) (\ln \boldsymbol{\eta} + \ln \boldsymbol{b})
\end{aligned}$$

Which implies that there's a closed-form solution for the sectoral knowledge stock along the entire path of the economy:

$$\begin{aligned}
\ln \boldsymbol{q}_t &= \widetilde{\ln \boldsymbol{q}_t} + \boldsymbol{A} \ln \boldsymbol{q}_t \\
&= \boldsymbol{A} \ln \boldsymbol{q}_0 + \lambda \boldsymbol{A} (\ln \boldsymbol{\eta} + \ln \bar{s} \cdot \mathbf{1} + \ln \boldsymbol{b})\, t \\
&\quad + e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t} \widetilde{\ln \boldsymbol{q}_0} + (\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})^{-1} \left(\boldsymbol{I} - e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t} \right) (\boldsymbol{I}-\boldsymbol{A}) (\ln \boldsymbol{\eta} + \ln \boldsymbol{b})
\end{aligned}$$

Starting from the same initial knowledge stock $\boldsymbol{q}_0$ but with two different time-invariant R&D allocations $\widetilde{\boldsymbol{b}}$ and $\boldsymbol{b}$, we have the following difference in knowledge stock over time:

$$\begin{aligned}
\ln \boldsymbol{q}_t \left(\widetilde{\boldsymbol{b}} \right) - \ln \boldsymbol{q}_t (\boldsymbol{b}) &= \boldsymbol{A} \ln \boldsymbol{q}_t \left(\widetilde{\boldsymbol{b}} \right) - \boldsymbol{A} \ln \boldsymbol{q}_t (\boldsymbol{b}) \\
&\quad + \widetilde{\ln \boldsymbol{q}_t} \left(\widetilde{\boldsymbol{b}} \right) - \widetilde{\ln \boldsymbol{q}_t} (\boldsymbol{b}) \\
&= \left[\boldsymbol{A} \lambda t + (\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})^{-1} \left(\boldsymbol{I} - e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t} \right) (\boldsymbol{I}-\boldsymbol{A}) \right] \left(\ln \widetilde{\boldsymbol{b}} - \ln \boldsymbol{b} \right)
\end{aligned}$$

Note

$$\int_0^\infty e^{-\rho t} \lambda t \,\mathrm{d}t = -\frac{1}{\rho} e^{-\rho t} \lambda t \Big|_0^\infty + \int_0^\infty \frac{1}{\rho} e^{-\rho t} \lambda \,\mathrm{d}t = \frac{\lambda}{\rho^2}$$

The difference in consumer welfare under two time-invariant paths of R&D allocations is

$$\begin{aligned}
& V\left(\boldsymbol{q}_0; \{\boldsymbol{\ell}_t\}, \widetilde{\boldsymbol{b}}\right) - V\left(\boldsymbol{q}_0; \{\boldsymbol{\ell}_t\}, \boldsymbol{b}\right) \\
= \;& \psi \boldsymbol{\beta}' \int_0^\infty e^{-\rho t} \left[\ln \boldsymbol{q}_t\left(\widetilde{\boldsymbol{b}}\right) - \ln \boldsymbol{q}_t\left(\boldsymbol{b}\right)\right] \, \mathrm{d}t \\
= \;& \psi \boldsymbol{\beta}' \int_0^\infty e^{-\rho t} \left[\boldsymbol{A}\lambda t + \left(\boldsymbol{I} - \boldsymbol{\Omega} + \boldsymbol{A}\right)^{-1} \left(\boldsymbol{I} - e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t}\right)\left(\boldsymbol{I} - \boldsymbol{A}\right)\right] \, \mathrm{d}t \left(\ln \widetilde{\boldsymbol{b}} - \ln \boldsymbol{b}\right) \\
= \;& \frac{\psi\lambda}{\rho^2} \boldsymbol{\beta}' \boldsymbol{A} \left(\ln \widetilde{\boldsymbol{b}} - \ln \boldsymbol{b}\right) \\
& + \psi \boldsymbol{\beta}' \left(\boldsymbol{I} - \boldsymbol{\Omega} + \boldsymbol{A}\right)^{-1} \left[\frac{1}{\rho}\boldsymbol{I} - \int_0^\infty \left(e^{-((\rho+\lambda)\boldsymbol{I} - \lambda(\boldsymbol{\Omega}-\boldsymbol{A}))t}\right) \, \mathrm{d}t\right] \left(\boldsymbol{I} - \boldsymbol{A}\right) \left(\ln \widetilde{\boldsymbol{b}} - \ln \boldsymbol{b}\right) \\
= \;& \psi \boldsymbol{\beta}' \left[\frac{\lambda}{\rho^2}\boldsymbol{A} + \left(\boldsymbol{I} - \boldsymbol{\Omega} + \boldsymbol{A}\right)^{-1} \left[\frac{1}{\rho}\boldsymbol{I} - \frac{1}{\rho+\lambda}\left(\boldsymbol{I} - \frac{\lambda}{\rho+\lambda}\left(\boldsymbol{\Omega} - \boldsymbol{A}\right)\right)^{-1}\right] \left(\boldsymbol{I} - \boldsymbol{A}\right)\right] \left(\ln \widetilde{\boldsymbol{b}} - \ln \boldsymbol{b}\right) \\
= \;& \psi \boldsymbol{\beta}' \left[\frac{\lambda}{\rho^2}\boldsymbol{A} + \frac{1}{\rho}\left(\boldsymbol{I} - \boldsymbol{\Omega} + \boldsymbol{A}\right)^{-1} \left[\frac{\lambda}{\rho+\lambda}\left(\boldsymbol{I} - \left(\boldsymbol{\Omega} - \boldsymbol{A}\right)\right)\left(\boldsymbol{I} - \frac{\lambda}{\rho+\lambda}\left(\boldsymbol{\Omega} - \boldsymbol{A}\right)\right)^{-1}\right] \left(\boldsymbol{I} - \boldsymbol{A}\right)\right] \left(\ln \widetilde{\boldsymbol{b}} - \ln \boldsymbol{b}\right) \\
= \;& \frac{\psi}{\rho} \boldsymbol{\beta}' \left[\frac{\lambda}{\rho}\boldsymbol{A} + \frac{\lambda}{\rho+\lambda}\left(\boldsymbol{I} - \frac{\lambda}{\rho+\lambda}\left(\boldsymbol{\Omega} - \boldsymbol{A}\right)\right)^{-1} \left(\boldsymbol{I} - \boldsymbol{A}\right)\right] \left(\ln \widetilde{\boldsymbol{b}} - \ln \boldsymbol{b}\right) \\
= \;& \frac{\psi}{\rho} \boldsymbol{\beta}' \frac{\lambda}{\rho+\lambda}\left(\boldsymbol{I} - \frac{\lambda}{\rho+\lambda}\left(\boldsymbol{\Omega} - \boldsymbol{A}\right)\right)^{-1} \left[\left(\boldsymbol{I} - \frac{\lambda}{\rho+\lambda}\left(\boldsymbol{\Omega} - \boldsymbol{A}\right)\right)\frac{\rho+\lambda}{\rho}\boldsymbol{A} + \left(\boldsymbol{I} - \boldsymbol{A}\right)\right] \left(\ln \widetilde{\boldsymbol{b}} - \ln \boldsymbol{b}\right) \\
= \;& \frac{\psi}{\rho^2} \boldsymbol{\beta}' \frac{\lambda}{\rho+\lambda}\left(\boldsymbol{I} - \frac{\lambda}{\rho+\lambda}\left(\boldsymbol{\Omega} - \boldsymbol{A}\right)\right)^{-1} \left[\rho\boldsymbol{I} + \lambda\boldsymbol{A}\right] \left(\ln \widetilde{\boldsymbol{b}} - \ln \boldsymbol{b}\right)
\end{aligned}$$

Note

$$\left(\rho\boldsymbol{I} + \lambda\boldsymbol{A}\right)^{-1} = \frac{1}{\rho}\left(\boldsymbol{I} - \frac{1}{1+\rho/\lambda}\boldsymbol{A}\right)$$

To see this,

$$\begin{aligned}
& \left(\rho\boldsymbol{I} + \lambda\boldsymbol{A}\right)\frac{1}{\rho}\left(\boldsymbol{I} - \frac{1}{1+\rho/\lambda}\boldsymbol{A}\right) \\
= \;& \frac{1}{\rho}\left(\rho\boldsymbol{I} + \lambda\boldsymbol{A} - \left(\rho\boldsymbol{I} + \lambda\boldsymbol{A}\right)\frac{1}{1+\rho/\lambda}\boldsymbol{A}\right) \\
= \;& \boldsymbol{I} + \frac{1}{\rho}\left(\lambda\boldsymbol{A} - \lambda\boldsymbol{A}\right) \\
= \;& \boldsymbol{I}
\end{aligned}$$

Hence,

$$
\begin{aligned}
& V\left(\boldsymbol{q}_0;\{\boldsymbol{\ell}_t\},\widetilde{\boldsymbol{b}}\right)-V\left(\boldsymbol{q}_0;\{\boldsymbol{\ell}_t\},\boldsymbol{b}\right) \\
= & \frac{\psi}{\rho}\boldsymbol{\beta}'\frac{\lambda}{\rho+\lambda}\left(\left(\boldsymbol{I}-\frac{1}{1+\rho/\lambda}\boldsymbol{A}\right)\left[\boldsymbol{I}-\frac{\lambda}{\rho+\lambda}(\boldsymbol{\Omega}-\boldsymbol{A})\right]\right)^{-1}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right) \\
= & \frac{\psi}{\rho}\boldsymbol{\beta}'\frac{\lambda}{\rho+\lambda}\left(\boldsymbol{I}-\frac{\lambda}{\rho+\lambda}(\boldsymbol{\Omega}-\boldsymbol{A})-\frac{1}{1+\rho/\lambda}\boldsymbol{A}\right)^{-1}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right) \\
= & \frac{\psi}{\rho}\boldsymbol{\beta}'\frac{\lambda}{\rho+\lambda}\left(\boldsymbol{I}-\frac{1}{1+\rho/\lambda}\boldsymbol{\Omega}\right)^{-1}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right) \\
= & \frac{\psi\lambda}{\rho^2}\boldsymbol{\gamma}'\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right),
\end{aligned}
$$

as desired.

A.6 Proof of Proposition 4: Consumption-Equivalent Welfare Gains from Adopting the Optimal R&D

For a given consumption path $\{y_t\}$, the welfare gain under the alternative consumption path $\{\mathcal{L}\cdot y_t\}$ is $\int e^{-\rho t}\ln\mathcal{L}\,\mathrm{d}t=\frac{\ln\mathcal{L}}{\rho}$. The result thus immediately follows Proposition 3.

A.7 Proof of Proposition 5: General Functional Forms and Endogenous Innovation Network

Consider the economic environment outlined in Section 2.5, with preferences

$$
\int_0^\infty e^{-\rho t}\ln\mathcal{Y}\left(\left\{q_{it}^{\psi}\ell_{it}\right\}\right)\,\mathrm{d}t
$$

and knowledge stock law of motion

$$
\mathrm{d}\ln q_{it}/\,\mathrm{d}t=\lambda\cdot\left(\ln\left(b_{it}\bar{s}\right)+\ln\mathcal{X}_i\left(\{q_{jt}\}\right)\right)\quad\forall i,
$$

where ℓ_{it} is the measure of production workers allocated to each variety in sector i at time t.

Consider the economy initially at $t=0$ in a BGP with R&D allocation $\boldsymbol{b}$. Define

$$
\beta_i\equiv\left.\frac{\partial\ln\mathcal{Y}\left(\{y_{jt}\}\right)}{\partial\ln y_{it}}\right|_{t=0},\qquad \omega_{ij}\equiv\begin{cases}\left.\frac{\partial\ln\mathcal{X}_i(\{q_{kt}\})}{\partial\ln q_{jt}}\right|_{t=0} & \text{if } i\neq j\\ 1+\left.\frac{\partial\ln\mathcal{X}_i(\{q_{it}\})}{\partial\ln q_{it}}\right|_{t=0} & \text{otherwise.}\end{cases}
$$

$\boldsymbol{\beta}\equiv[\beta_i]$ and $\boldsymbol{\Omega}\equiv[\omega_{ij}]$ are the consumption and innovation spillover elasticities evaluated in the initial BGP. Note that (1) $\mathcal{X}_i(\cdot)$ being homogeneous-of-degree-zero with positive cross-sector spillovers and (2) $\left|\partial\ln\mathcal{X}_i(\cdot)/\partial\ln q_{jt}\right|\leq 1\forall i,j$ jointly imply that $\omega_{ij}\geq 0$ for all i,j. Define

$\boldsymbol{\gamma}' = \frac{\rho}{\rho+\lambda}\boldsymbol{\beta}'\left(\boldsymbol{I} - \frac{\boldsymbol{\Omega}}{1+\rho/\lambda}\right)^{-1}$.

We now derive the first-order welfare impact of perturbing R&D allocation. Let $V(\ln \boldsymbol{q}_0; \ln \boldsymbol{b})$ denote the welfare under log-R&D allocation $\ln \boldsymbol{b}$. Formally, we show that the Gateaux derivative of welfare with respect to log R&D allocation $\ln \boldsymbol{b}$ in the direction of $\boldsymbol{h}$ is

$$\lim_{\alpha\to 0} \frac{V(\ln \boldsymbol{q}_0; \ln \boldsymbol{b} + \alpha \boldsymbol{h}) - V(\ln \boldsymbol{q}_0; \boldsymbol{b})}{\alpha} = \frac{\psi\lambda}{\rho^2}\boldsymbol{\gamma}'\boldsymbol{h}.$$

Given log-R&D allocation $\ln \boldsymbol{b} + \alpha\boldsymbol{h}$, the law of motion for knowledge stock satisfies

$$\frac{\mathrm{d}\ln \boldsymbol{q}_t}{\mathrm{d}t} = \lambda\left(\ln \boldsymbol{b} + \alpha\boldsymbol{h} + \ln \boldsymbol{\chi}\left(\{\ln \boldsymbol{q}_t\}\right)\right)$$

$$\frac{\partial^2 \ln \boldsymbol{q}_t}{\partial\alpha\partial t} = \lambda\boldsymbol{h} + \lambda\left(\boldsymbol{\Omega} - \boldsymbol{I}\right)\frac{\partial \ln \boldsymbol{q}_t}{\partial\alpha}$$

$$\implies \frac{\partial \ln \boldsymbol{q}_t}{\partial \alpha} = \left(\boldsymbol{I} - \boldsymbol{\Omega}\right)^{-1}\left[\boldsymbol{I} - e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega})t}\right]\boldsymbol{h}$$

$$\begin{aligned}
& \lim_{\alpha\to 0} \frac{V(\ln \boldsymbol{q}_0; \ln \boldsymbol{b} + \alpha \boldsymbol{h}) - V(\ln \boldsymbol{q}_0; \boldsymbol{b})}{\alpha} \\
= & \int e^{-\rho t}\frac{\partial \ln Y(\psi \ln \boldsymbol{q}_t)}{\partial \ln \boldsymbol{q}_t}\frac{\partial \ln \boldsymbol{q}_t}{\partial \alpha}\,\mathrm{d}t \\
= & \psi\boldsymbol{\beta}'\left(\boldsymbol{I}-\boldsymbol{\Omega}\right)^{-1}\int e^{-\rho t}\left[\boldsymbol{I} - e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega})t}\right]\,\mathrm{d}t\boldsymbol{h} \\
= & \psi\boldsymbol{\beta}'\left(\boldsymbol{I}-\boldsymbol{\Omega}\right)^{-1}\left[\frac{1}{\rho}\boldsymbol{I} - \int e^{-((\rho+\lambda)\boldsymbol{I}-\lambda\boldsymbol{\Omega})t}\,\mathrm{d}t\right]\boldsymbol{h} \\
= & \psi\boldsymbol{\beta}'\left(\boldsymbol{I}-\boldsymbol{\Omega}\right)^{-1}\left[\frac{1}{\rho}\boldsymbol{I} - \frac{1}{\rho+\lambda}\left(\boldsymbol{I} - \frac{\lambda}{\rho+\lambda}\boldsymbol{\Omega}\right)^{-1}\right]\boldsymbol{h} \\
= & \psi\boldsymbol{\beta}'\left(\boldsymbol{I}-\boldsymbol{\Omega}\right)^{-1}\left[\frac{1}{\rho}\left(\boldsymbol{I} - \frac{\lambda}{\rho+\lambda}\boldsymbol{\Omega}\right) - \frac{1}{\rho+\lambda}\boldsymbol{I}\right]\left(\boldsymbol{I} - \frac{\lambda}{\rho+\lambda}\boldsymbol{\Omega}\right)^{-1}\boldsymbol{h} \\
= & \frac{\psi}{\rho}\boldsymbol{\beta}'\left(\boldsymbol{I}-\boldsymbol{\Omega}\right)^{-1}\left[\frac{\lambda}{\rho+\lambda}\boldsymbol{I} - \frac{\lambda}{\rho+\lambda}\boldsymbol{\Omega}\right]\left(\boldsymbol{I} - \frac{\lambda}{\rho+\lambda}\boldsymbol{\Omega}\right)^{-1}\boldsymbol{h} \\
= & \frac{\psi\lambda}{\rho^2}\frac{\rho}{\rho+\lambda}\boldsymbol{\beta}'\left(\boldsymbol{I} - \frac{\lambda}{\rho+\lambda}\boldsymbol{\Omega}\right)^{-1}\boldsymbol{h} \\
= & \frac{\psi\lambda}{\rho^2}\boldsymbol{\gamma}'\boldsymbol{h},
\end{aligned}$$

as desired.

A.8 Proof of Proposition 6: Optimal R&D in the Presence of Foreign Spillovers

First, note that given output y_t and the price of imports p_t^f, consumption, export, and import must solve

$$\bar{\mathcal{C}}^*\left(y_t, p_t^f\right) \equiv \max_{c_t^d, c_t^f} \mathcal{C}\left(c_t^d, c_t^f\right) \quad \text{s.t. } y_t - c_t^d = p_t^f c_t^f. \tag{A5}$$

Since $\mathcal{C}(\cdot)$ features constant-returns-to-scale, we can re-write the maximized consumption aggregator as $\bar{\mathcal{C}}^*\left(y_t, p_t^f\right) = y_t \mathcal{C}^*\left(p_t^f\right)$ for some function $\mathcal{C}^*$. Hence, for any $\boldsymbol{q}_t$, $\{\ell_{it}\}$ are chosen to maximize flow output; thus the optimal worker allocation features $\ell_{it}/\bar{\ell} = \beta_i$ as in the closed economy.

We next characterize the optimal R&D allocation. Let $\boldsymbol{\Theta} \equiv \boldsymbol{\Omega} \circ \boldsymbol{X}$. Given the law of motion for sectoral knowledge stock, we can solve for the evolution of knowledge stock in closed form as a function of R&D allocation $\boldsymbol{b}_t$:

$$\ln \boldsymbol{q}_t = e^{\lambda(\boldsymbol{\Theta} - \boldsymbol{I})t}\left[\ln \boldsymbol{q}_0 + \lambda \int_0^t e^{-\lambda(\boldsymbol{\Theta}-\boldsymbol{I})s}\left((\boldsymbol{\Omega} - \boldsymbol{\Theta}) \ln \boldsymbol{q}_s^f + \ln \boldsymbol{\eta} + \ln \bar{s} + \ln \boldsymbol{b}_s\right) \mathrm{d}s\right]. \tag{A6}$$

The optimal R&D allocation is

$$\begin{aligned}
\{\boldsymbol{\gamma}_t\} &= \arg\max_{\{\boldsymbol{b}_s\}} \int_0^\infty e^{-\rho t} \ln \bar{\mathcal{C}}^*\left(y_t\left(\{\boldsymbol{b}_s\}\right), p_t^f\right) \mathrm{d}t \\
&= \arg\max_{\{\boldsymbol{b}_s\}} \int_0^\infty e^{-\rho t} \ln y_t\left(\{\boldsymbol{b}_s\}\right) \mathrm{d}t \\
&= \arg\max_{\{\boldsymbol{b}_s\}} \int_0^\infty e^{-\rho t} \boldsymbol{\beta}' \ln \boldsymbol{q}_t\left(\{\boldsymbol{b}_s\}\right) \mathrm{d}t \\
&= \arg\max_{\{\boldsymbol{b}_s\}} \boldsymbol{\beta}' \int_0^\infty e^{-\rho t}\left[\lambda \int_0^t e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Theta})(t-s)} \ln \boldsymbol{b}_s \, \mathrm{d}s\right] \mathrm{d}t.
\end{aligned}$$

The optimal R&D allocation therefore coincides with the solution to the following problem:

$$\arg\max_{\{\boldsymbol{b}_s\}} \int_0^\infty e^{-\rho t} \boldsymbol{\beta}' \boldsymbol{m}_t \, \mathrm{d}t$$

$$\text{s.t. } \dot{\boldsymbol{m}}_t = \lambda\left(\boldsymbol{\Theta} - \boldsymbol{I}\right) \boldsymbol{m}_t + \lambda \ln \boldsymbol{b}_t, \quad \boldsymbol{m}_0 \text{ given},$$

which can be solved in closed form by forming the Hamiltonian, following a similar procedure as in the proof for Proposition 1. The solution features

$$\boldsymbol{\gamma}' = \xi^{-1} \frac{\rho}{\rho + \lambda} \boldsymbol{\beta}' \left(\boldsymbol{I} - \frac{\boldsymbol{\Omega} \circ \boldsymbol{X}}{1 + \rho/\lambda}\right)^{-1}, \quad \xi \equiv \frac{\rho}{\rho + \lambda} \boldsymbol{\beta}' \left(\boldsymbol{I} - \frac{\boldsymbol{\Omega} \circ \boldsymbol{X}}{1 + \rho/\lambda}\right)^{-1} \boldsymbol{1},$$

as desired.

A.9 Proof of Proposition 7: Welfare Impact of R&D in the Presence of Foreign Spillovers

Starting from an initial condition $\boldsymbol{q}_0$, a path of foreign knowledge and import prices $\left\{\boldsymbol{q}_t^f, p_t^f\right\}$, and a path of worker allocation $\{\boldsymbol{\ell}_t\}$, the welfare differences between an economy with optimal R&D allocation $\boldsymbol{\gamma}$ and an economy with time-invariant allocation $\boldsymbol{b}$ is

$$V(\boldsymbol{\gamma}) - V(\boldsymbol{b}) = \int_0^\infty e^{-\rho t}\left[\ln \bar{\mathcal{C}}^*\left(y_t(\boldsymbol{\gamma}), p_t^f\right) - \ln \bar{\mathcal{C}}^*\left(y_t(\boldsymbol{b}), p_t^f\right)\right] \mathrm{d}t,$$

where $\bar{\mathcal{C}}^*$ is defined in (A5). Following the proof to Proposition 6, $\bar{\mathcal{C}}^*\left(y_t, p_t^f\right) = y_t \mathcal{C}^*\left(p_t^f\right)$; hence the welfare differences can be re-written as

$$V(\boldsymbol{\gamma}) - V(\boldsymbol{b}) = \int_0^\infty e^{-\rho t}\left[\ln y_t(\boldsymbol{\gamma}) - \ln y_t(\boldsymbol{b})\right] \mathrm{d}t.$$

Since $\ln y_t$ is additive in $\psi\boldsymbol{\beta}'\ln\boldsymbol{q}_t$, we can re-write the welfare differences in terms of the discounted integral of $\boldsymbol{\beta}$-weighted differences in knowledge stock induced by the two different R&D allocation vectors. By (A6), we can re-write the welfare differences as

$$V(\boldsymbol{\gamma}) - V(\boldsymbol{b}) = \psi\boldsymbol{\beta}' \int_0^\infty e^{-\rho t}\left[\lambda \int_0^t e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Theta})(t-s)}\,\mathrm{d}s\right] \mathrm{d}t\,(\ln\boldsymbol{\gamma} - \ln\boldsymbol{b}),$$

where $\boldsymbol{\Theta} \equiv \boldsymbol{\Omega} \circ \boldsymbol{X}$. To simplify the integral we follow the proof to Proposition 3:[12]

$$\begin{aligned}
V(\boldsymbol{\gamma}) - V(\boldsymbol{b}) &= \psi\boldsymbol{\beta}' \int_0^\infty e^{-\rho t}\left[\lambda \int_0^t e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Theta})(t-s)}\,\mathrm{d}s\right] \mathrm{d}t\,(\ln\boldsymbol{\gamma} - \ln\boldsymbol{b}),\\
&= \psi\boldsymbol{\beta}'(\boldsymbol{I}-\boldsymbol{\Theta})^{-1}\left(\int_0^\infty e^{-\rho t}\left[\boldsymbol{I} - e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Theta})t}\right]\mathrm{d}t\right)(\ln\boldsymbol{\gamma} - \ln\boldsymbol{b})\\
&= \frac{\psi}{\rho}\boldsymbol{\beta}'(\boldsymbol{I}-\boldsymbol{\Theta})^{-1}\left(\boldsymbol{I} - \frac{\rho}{\rho+\lambda}\left(\boldsymbol{I} - \frac{1}{1+\rho/\lambda}\boldsymbol{\Theta}\right)^{-1}\right)(\ln\boldsymbol{\gamma} - \ln\boldsymbol{b})\\
&= \frac{\psi\lambda}{\rho^2}\frac{\rho}{\rho+\lambda}\boldsymbol{\beta}'\left(\boldsymbol{I} - \frac{1}{1+\rho/\lambda}\boldsymbol{\Theta}\right)^{-1}(\ln\boldsymbol{\gamma} - \ln\boldsymbol{b})\\
&= \frac{\psi\lambda}{\rho^2}\underbrace{\frac{\rho}{\rho+\lambda}\boldsymbol{\beta}'\left(\boldsymbol{I} - \frac{1}{1+\rho/\lambda}\boldsymbol{\Theta}\right)^{-1}\boldsymbol{1}}_{\equiv\xi}\underbrace{\frac{\boldsymbol{\beta}'\left(\boldsymbol{I} - \frac{1}{1+\rho/\lambda}\boldsymbol{\Theta}\right)^{-1}}{\boldsymbol{\beta}'\left(\boldsymbol{I} - \frac{1}{1+\rho/\lambda}\boldsymbol{\Theta}\right)^{-1}\boldsymbol{1}}}_{\equiv\boldsymbol{\gamma}'}(\ln\boldsymbol{\gamma} - \ln\boldsymbol{b})\\
&= \frac{\psi\lambda}{\rho^2}\xi\boldsymbol{\gamma}'(\ln\boldsymbol{\gamma} - \ln\boldsymbol{b}).
\end{aligned}$$

[12] Note that $\boldsymbol{I}-\boldsymbol{\Theta}$ is generically invertible—the economy with foreign spillovers exhibit aggregate decreasing-returns-to-scale in domestic R&D—so the proof here is simpler than in the baseline model.

For a given consumption path $\left\{\bar{\mathcal{C}}^*\left(y_t, p_t^f\right)\right\}$, the welfare gain under the alternative consumption path $\left\{\mathcal{L}\cdot\bar{\mathcal{C}}^*\left(y_t, p_t^f\right)\right\}$ is $\int e^{-\rho t}\ln\mathcal{L}\,\mathrm{d}t = \frac{\ln\mathcal{L}}{\rho}$. The consumption-equivalent welfare gains from adopting the optimal R&D allocation is thus

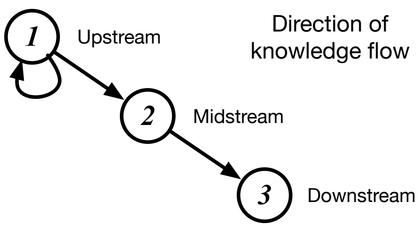
$$\mathcal{L}(\boldsymbol{b},\xi) = \exp\left(\frac{\psi\lambda}{\rho}\xi\boldsymbol{\gamma}'\left(\ln\boldsymbol{\gamma} - \ln\boldsymbol{b}\right)\right),$$

as desired.

B Theoretical Extensions

B.1 Three-Sector Example

To demonstrate Propositions 1 and 2, consider the following three-sector example, where knowledge flows from sector 1 to sector 2 and from sector 2 to sector 3. Sector 1 can thus be interpreted as the "upstream" sector of knowledge flows, and sector 3 is the knowledge "downstream." To ensure the knowledge aggregator χ_{it} has constant returns to scale in every sector, we specify that knowledge in sector 1 also benefits itself. For simplicity, we assume the consumer values goods from each sector equally, with consumption share $\beta_i = 1/3$ for all i.



$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}.$$

The socially optimal R&D allocations depend on the effective discount rate ρ/λ and should follow, according to Proposition 1,

$$\boldsymbol{\gamma}' = \frac{\rho}{\rho+\lambda}\boldsymbol{\beta}'\left(\boldsymbol{I} - \frac{\boldsymbol{\Omega}}{1+\rho/\lambda}\right)^{-1} = \begin{bmatrix} \frac{1+(1+\rho/\lambda)+(1+\rho/\lambda)^2}{3(1+\rho/\lambda)^2} & \frac{\rho/\lambda+\rho/\lambda(1+\rho/\lambda)}{3(1+\rho/\lambda)^2} & \frac{\rho/\lambda}{3(1+\rho/\lambda)} \end{bmatrix}.$$

When the effective discount rate ρ/λ is lower, more resources should be directed to upstream sector 1 and fewer to downstream sector 3. A myopic planner ($\rho/\lambda \to \infty$) chooses $\gamma_1 = \gamma_3$; as the when $\rho/\lambda = 1$, $\gamma_1/\gamma_3 \approx 3.5$; when $\rho/\lambda = 0.1$, $\gamma_1/\gamma_3 \approx 30.1$.

B.2 Embedding Input-Output Linkages into Production Functions

We now expand on Section 2.7.1 and introduce input-output linkages into the baseline model. As discussed in the main text, for the optimal R&D allocation $\boldsymbol{\gamma}' \propto \boldsymbol{\beta}'\left(\boldsymbol{I} - \frac{\boldsymbol{\Omega}}{1+\rho/\lambda}\right)^{-1}$, the presence of a production network requires a different construction for the $\boldsymbol{\beta}$ vector, but the innovation network $\boldsymbol{\Omega}$ term is unaffected. Formally, the $\boldsymbol{\beta}$ vector should capture the elasticity of aggregate consumption with respect to the knowledge stock in each sector; in the presence of a production network, it should reflect not only the consumer preferences but also the production network structure. With this adjustment, our main results continue to hold in this environment.

Specifically, suppose the production of good i requires other goods as intermediate inputs:

$$\ln y_{it} = \textstyle\sum_{j=1}^{K} \sigma_{ij} \ln m_{ijt} + \alpha_i \ln q_{it}^{\psi} \ell_{it} \, \mathrm{d}\nu, \quad \alpha_i + \textstyle\sum_{j=1}^{K} \sigma_{ij} = 1, \tag{A7}$$

where m_{ijt} is the quantity of good j used for the production of good i, α_i is sector i's output elasticity to value-added, and σ_{ij} is sector i's output elasticity to input j. The baseline model is a special case with $\sigma_{ij} = 0$ for all i, j. When an equal amount of labor ℓ_{it} is allocated to each variety within a sector, production function (A7) takes the standard form in the canonical production network model (Acemoglu et al., 2012):

$$y_{it} = \left(q_{it}^{\psi} \ell_{it} \right)^{\alpha_i} \textstyle\prod_{j=1}^{K} m_{ijt}^{\sigma_{ij}}. \tag{A8}$$

The market clearing condition for sectoral good follows

$$y_{jt} = \sum_i m_{ijt} + c_{jt}. \tag{A9}$$

The aggregate consumption bundle follows:

$$\ln y_t = \sum_{i=1}^{K} \beta_i \ln c_{it}. \tag{A10}$$

Consider the problem of choosing worker allocation to maximize flow consumption:

$$\ln y^* \left(\boldsymbol{q}_t \right) \equiv \max_{\{\ell_{it}\}} \sum_{i=1}^{K} \beta_i \ln c_{it}$$

subject to (A9) and (A8). Let $\boldsymbol{\Sigma} \equiv [\sigma_{ij}]$ denote the matrix of input-output elasticities. Standard results in the production networks literature (e.g., see Acemoglu et al., 2012 and Liu, 2019) imply

$$\ln y^* \left(\boldsymbol{q}_t \right) = \text{const} + \ln \bar{\ell} + \sum_i \hat{\beta}_i \ln q_{it},$$

where $\hat{\beta}_i \equiv \alpha_i \left[\boldsymbol{\beta}' \left(\boldsymbol{I} - \boldsymbol{\Sigma} \right)^{-1} \right]_i$ is the product between sectoral value-added elasticity α_i and the i-th entry of the influence vector $\boldsymbol{\beta}' \left(\boldsymbol{I} - \boldsymbol{\Sigma} \right)^{-1}$. $\hat{\beta}_i$ can be interpreted as the elasticity of aggregate output with respect to sectoral knowledge stock. Hence, results in the main text extend intuitively to this setting with input-output linkages: the optimal worker allocation follows the vector $\hat{\boldsymbol{\beta}}$, and the optimal R&D allocation $\gamma_{it} \equiv s_{it}/\bar{s}$ follows $\boldsymbol{\gamma}' \propto \hat{\boldsymbol{\beta}}' \left(\boldsymbol{I} - \frac{\boldsymbol{\Omega}}{1+\rho/\lambda}\right)^{-1}$.

B.3 Semi-Endogenous Growth

Our baseline model features endogenous growth: a positive growth rate of aggregate output along a balanced growth path in the absence of population growth. This is because the R&D technology features aggregate constant-returns-to-scale in sectoral knowledge stock. We now expand on Section 2.7.2 and embed our innovation network formulation into a semi-endogenous growth setting, with a constant growth rate in the total measure of scientists $\bar{s}_t = \bar{s}_0 e^{\bar{g}t}$. We show that the optimal R&D allocation follows $\boldsymbol{\gamma}' \propto \boldsymbol{\beta}' \left(\boldsymbol{I} - \frac{\boldsymbol{\Omega}}{1+\kappa+\rho/\lambda}\right)^{-1}$, and the consumption-equivalent welfare impact of adopting the optimal allocation is $\mathcal{L}(\boldsymbol{b}) = \exp\left(\frac{\lambda}{\rho+\kappa\lambda}\boldsymbol{\gamma}'(\ln\boldsymbol{\gamma} - \ln\boldsymbol{b})\right)$.

Specifically, replace the knowledge stock evolution equation (5) with

$$\dot{q}_{it}/q_{it} = \lambda \ln\left(n_{it}/q_{it}^{1+\kappa}\right),$$

where $\kappa \geq 0$ captures the rate at which proportional improvements in knowledge are getting harder to find (Bloom et al. 2020, Jones 2022). The knowledge law of motion (9) becomes

$$\mathrm{d}\ln\boldsymbol{q}_t / \mathrm{d}t = \lambda \cdot \left(\ln\boldsymbol{\eta} + \ln\boldsymbol{s}_t + \bar{g}t + \left(\boldsymbol{\Omega} - (1+\kappa)\boldsymbol{I}\right)\ln\boldsymbol{q}_t\right).$$

Integrating the ODE system over time, we get

$$\ln\boldsymbol{q}_t = e^{\lambda(\boldsymbol{\Omega}-(1+\kappa)\boldsymbol{I})t}\left[\ln\boldsymbol{q}_0 + \lambda\int_0^t e^{-\lambda(\boldsymbol{\Omega}-(1+\kappa)\boldsymbol{I})u}\left(\ln\boldsymbol{\eta} + \ln\boldsymbol{s}_u + \bar{g}u\right)\mathrm{d}u\right]$$

For given initial levels of knowledge stock and path of worker allocation, the difference in welfare under two R&D allocations $\widetilde{\boldsymbol{b}}$ and $\boldsymbol{b}$ is

$$\begin{aligned}
& V\left(\boldsymbol{q}_0;\{\boldsymbol{\ell}_t\},\widetilde{\boldsymbol{b}}\right)-V\left(\boldsymbol{q}_0;\{\boldsymbol{\ell}_t\},\boldsymbol{b}\right)\\
= \;& \psi\boldsymbol{\beta}'\int_0^\infty e^{-\rho t}\left[\ln\boldsymbol{q}_t\left(\widetilde{\boldsymbol{b}}\right)-\ln\boldsymbol{q}_t\left(\boldsymbol{b}\right)\right]\,\mathrm{d}t\\
= \;& \psi\lambda\boldsymbol{\beta}'\int_0^\infty e^{-\rho t}\left[\int_0^t e^{-\lambda((1+\kappa)\boldsymbol{I}-\boldsymbol{\Omega})(t-u)}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right)\,\mathrm{d}u\right]\,\mathrm{d}t\\
= \;& \psi\boldsymbol{\beta}'\left((1+\kappa)\boldsymbol{I}-\boldsymbol{\Omega}\right)^{-1}\left(\int_0^\infty e^{-\rho t}\left[\boldsymbol{I}-e^{-\lambda((1+\kappa)\boldsymbol{I}-\boldsymbol{\Omega})t}\right]\,\mathrm{d}t\right)\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right)\\
= \;& \frac{\psi}{\rho}\boldsymbol{\beta}'\frac{1}{1+\kappa}\left(\boldsymbol{I}-\frac{\boldsymbol{\Omega}}{1+\kappa}\right)^{-1}\left(\boldsymbol{I}-\frac{\rho}{\rho+\lambda+\lambda\kappa}\left(\boldsymbol{I}-\frac{\boldsymbol{\Omega}}{1+\kappa+\rho/\lambda}\right)^{-1}\right)\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right)\\
= \;& \frac{\psi\lambda}{\rho}\boldsymbol{\beta}'\frac{1}{\rho+\lambda+\kappa\lambda}\left(\boldsymbol{I}-\frac{\boldsymbol{\Omega}}{1+\kappa+\rho/\lambda}\right)^{-1}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right)\\
= \;& \frac{\psi\lambda}{\rho}\frac{1}{\kappa\lambda+\rho}\frac{\rho+\kappa\lambda}{\rho+\lambda+\kappa\lambda}\boldsymbol{\beta}'\left(\boldsymbol{I}-\frac{\boldsymbol{\Omega}}{1+\kappa+\rho/\lambda}\right)^{-1}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right)
\end{aligned}$$

It is easy to verify that $\boldsymbol{\gamma}'\equiv\frac{\rho+\kappa\lambda}{\rho+\lambda+\kappa\lambda}\boldsymbol{\beta}'\left(\boldsymbol{I}-\frac{\boldsymbol{\Omega}}{1+\kappa+\rho/\lambda}\right)^{-1}$ sums to one; hence we have

$$V\left(\boldsymbol{q}_0;\{\boldsymbol{\ell}_t\},\widetilde{\boldsymbol{b}}\right)-V\left(\boldsymbol{q}_0;\{\boldsymbol{\ell}_t\},\boldsymbol{b}\right)=\frac{\psi}{\rho}\frac{\lambda}{\rho+\kappa\lambda}\boldsymbol{\gamma}'\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right).$$

Clearly $\boldsymbol{\gamma}$ is the optimal allocation, and, analogous to the argument in Section A.6, the consumption-equivalent welfare impact of adopting the optimal allocation is $\mathcal{L}\left(\boldsymbol{b}\right)=\exp\left(\frac{\psi\lambda}{\rho+\kappa\lambda}\boldsymbol{\gamma}'\left(\ln\boldsymbol{\gamma}-\ln\boldsymbol{b}\right)\right)$.

B.4 An Illustrative Decentralized Equilibrium

In an innovation network, knowledge is a public good, as knowledge creation benefits subsequent R&D in other sectors and all future periods. To the extent that innovators do not fully internalize such future benefits,[13] a decentralized market does not implement the optimal R&D allocation. To demonstrate the potential inefficiency, in this section we construct a decentralized equilibrium in which innovators conduct R&D only in pursuit of profits, disregarding any beneficial spillovers their R&D activities may provide in the future. As we show, the decentralized allocation of R&D resources follows $\boldsymbol{\beta}$ along a BGP, which coincides with the planner's solution only if the society is completely myopic ($\rho/\lambda\to 0$).

It is important to note that our decentralized equilibrium lacks many real-world features of the market for innovation (e.g., multi-sector firms, mergers and acquisitions, and patent licensing). This is intentional: the goal of this section is not to capture quantitative realism but to

[13]While a patent gives its owner the legal right to exclude others from making or using an invention, it does not by design preclude others from benefiting from knowledge spillovers.

illustrate as clearly as possible the potential inefficiency of decentralized R&D decisions given knowledge spillovers. By comparing the R&D allocations in the data to the first-best, our notion of allocative efficiency—measured by the consumption-equivalent welfare impact of reallocating R&D optimally—does not require that we take a stance on firms' equilibrium conduct; instead, it has the advantage of directly calculating the welfare impact of reallocating R&D based on the economic environment

Consider a decentralized economy in which each intermediate variety is produced by a distinct monopolist. Different vintages of the same variety are perfect substitutes. Because the most recent vintage's quality is λ proportionally higher than the next best vintage, the monopolist conducts limit pricing and charges a markup $(1+\lambda)$. No vintages with dominated quality are produced in equilibrium.

In each sector, innovation is carried out by a continuum of potential entrants, who hire scientists to conduct R&D and generate new innovations according to equation (4). New innovation flows improve the quality of a random variety within sector i at Poisson rate $\ln(n_{it}/q_{it})$; the innovating firm overtakes as the monopolist of that variety until another successful innovation occurs in the future.

The representative consumer receives all workers' and scientists' income and firm profits. Given the initial state variables $\{q_{i0}\}_{i=1}^{K}$, a decentralized equilibrium is the time path of prices, quantities, and knowledge stocks such that production firms set prices to maximize profits, the consumer chooses bundles of goods to consume to maximize utility, and potential entrants hire scientists for R&D to maximize expected profits. A decentralized BGP is an equilibrium in which all sectors' knowledge stock grows at the same constant rate.

Proposition 8. *In the decentralized BGP, the allocations of R&D and production resources both follow the consumption shares:* $\ell_{it}(\nu) = \ell_{it} = \beta_i \bar{\ell}$ *and* $s_{it} = \beta_i \bar{s}$.

Proof. We normalize the consumer price index to one for all times t. The consumer at each time t spends a constant fraction β_i of their income on sectoral composite good i, with

$$p_{it} y_{it} = \beta_i y_t \quad \text{for all } i, t. \tag{A11}$$

The sectoral composite aggregator (3) further implies that the total revenue of each variety ν is also equal to $\beta_i y_t$, and, because each monopolist sets a markup $(1+\lambda)$, we derive the profits in each sector i as

$$\pi_{it}(\nu) = \frac{\lambda}{1+\lambda} \beta_i y_t \quad \text{for all } i, t, \nu. \tag{A12}$$

Because all varieties have identical markups, the worker allocation is identical across varieties within each sector. The total labor cost in each sector i is $\frac{1}{1+\lambda} \beta_i y_t$ and is thus also proportional to the consumption shares β_i:

$$\ell_{it}(\nu) = \ell_{it} = \beta_i \bar{\ell} \quad \text{for all } i, t, \nu. \tag{A13}$$

Along the BGP, a monopolist in each sector has the same Poisson rate to be replaced by an innovating entrant. Let δ denote that replacement rate; the value of a monopolistic firm is thus

$$v_{it} \equiv \int_t^\infty e^{-(r_s+\delta)(s-t)} \pi_{is} \, \mathrm{d}s, \tag{A14}$$

where r_s is the interest rate at time s and g_s is the growth rate of aggregate consumption y_s. Note we have suppressed the index for variety since all varieties have the same profits and thus the same value within each sector. Because sectoral profits are always proportional to the consumption shares at all times, we have

$$v_{it}/v_{jt} = \beta_i/\beta_j \quad \text{for all } i, j, t. \tag{A15}$$

Entrants hire scientists to conduct research in order to become future monopolists. The marginal value from an additional scientist ($v_{it} \times \partial \ln(n_{it}/q_{it}) / \partial s_{it}$) must be equalized across sectors. Substituting n_{it} using the innovation production function (4) and v_{it}/v_{jt} using equation (A15), we obtain that scientist allocation must also follow the consumption share, that is, $s_{it}/\bar{s} = \beta_i$ for all t, as desired. □

Intuitively, varieties in a sector with higher consumption share β_i have proportionally higher revenue, employment, and flow profits. Since the rate at which an innovating entrant replaces a producing monopolist is the same across all sectors along a BGP, a monopolistic firm's value is also proportional to the consumption share β_i of the sector. Because entrants conduct research to obtain that monopolistic value, the marginal value from an additional scientist must be equalized across sectors, and the innovation production function (4) thus implies that R&D allocation must follow $s_{it} = \beta_i \bar{s}$ along the BGP.

The decentralized R&D allocation $\boldsymbol{\beta}$ stands in contrast to the socially optimal allocation, $\boldsymbol{\gamma}' \propto \boldsymbol{\beta}' \left(\boldsymbol{I} + \frac{\boldsymbol{\Omega}}{1+\rho/\lambda} + \left(\frac{\boldsymbol{\Omega}}{1+\rho/\lambda} \right)^2 + \cdots \right)$. While the social planner takes into account both R&D's direct effect on product quality as well as the infinite rounds of indirect network spillover effects, the decentralized allocation is driven by firm profits and thus accounts only for the direct effect, as infinitesimal firms cannot monetize the future spillover effects of their own R&D.

B.5 Constrained Optimal R&D Allocations

In some settings, for instance under political or feasibility constraints, a planner may only be able to reallocate resources across a subset $\mathcal{K} \subset \{1, \ldots, K\}$ of sectors. We now generalize our results to such an environment. We show that our earlier results extend naturally: resources among sectors in $\mathcal{K}$ should be allocated proportionally to the unconstrained optimal allocation $\boldsymbol{\gamma}$. We generalize the welfare sufficient statistic to this setting as well.

For a generic allocation vector $\boldsymbol{b}$, we denote $\boldsymbol{b}^{\mathcal{K}}$ as the $|\mathcal{K}| \times 1$ allocation vector that sums to one with entries proportional to $\boldsymbol{b}$ for all sectors in $\mathcal{K}$ (i.e., $b_i^{\mathcal{K}} \equiv \frac{b_i}{\sum_{j \in \mathcal{K}} b_j}$ for $i \in \mathcal{K}$).

Proposition 9. *Suppose R&D allocations in sectors $k \notin \mathcal{K}$ are given exogenously and that the planner can only choose R&D allocations in sectors $k \in \mathcal{K}$ when solving the planning problem in (7). Along the entire equilibrium path, the constrained optimal R&D allocation is $s_i = \gamma_i^{\mathcal{K}} \left(\bar{s} - \sum_{k \notin \mathcal{K}} s_k\right)$ for $i \in \mathcal{K}$. The consumption-equivalent welfare gains from adopting the constrained-optimal R&D allocation (instead of allocation $\boldsymbol{b}$) is $\mathcal{L}^{\mathcal{K}}(\boldsymbol{b}) = \exp\left(\frac{\psi\lambda}{\rho}\left(\sum_{j \in \mathcal{K}} \gamma_j\right)\left(\boldsymbol{\gamma}^{\mathcal{K}}\right)'\left(\ln \boldsymbol{\gamma}^{\mathcal{K}} - \ln \boldsymbol{b}^{\mathcal{K}}\right)\right)$.*

The Proposition shows that among sectors in which the planner can allocate resources, the constrained-optimal resource allocation is proportional to the unconstrained-optimal allocation $\boldsymbol{\gamma}$. For the welfare sufficient statistic, note that the relative entropy of $\boldsymbol{b}^{\mathcal{K}}$ from $\boldsymbol{\gamma}^{\mathcal{K}}$, $\left(\boldsymbol{\gamma}^{\mathcal{K}}\right)'\left(\ln \boldsymbol{\gamma}^{\mathcal{K}} - \ln \boldsymbol{b}^{\mathcal{K}}\right)$, summarizes the distance relative to the first-best allocation among sectors in $\mathcal{K}$. Relative to the welfare formula (15) for the unconstrained optimal allocation, the new term $\sum_{j \in \mathcal{K}} \gamma_j \leq 1$ (with equality when $\mathcal{K}$ includes all sectors) reflects the fact that there is less to be gained when the planner can reallocate resources across fewer sectors.

Proof. Let $s^{\mathcal{K}} \equiv \bar{s} - \sum_{k \notin \mathcal{K}} s_k$ denote the available resource the planner can allocate among sectors in $\mathcal{K}$, and let $\gamma_i^{\mathcal{K}}$ denote the constrained-optimal share of $s^{\mathcal{K}}$ allocated to sector i. That $\gamma_i^{\mathcal{K}}$ is time-invariant follows from the same proof as Proposition 3. $\boldsymbol{\gamma}^{\mathcal{K}}$ is thus the solution to

$$\boldsymbol{\gamma}^{\mathcal{K}} = \arg\max_{\{\delta_i\}_{i \in \mathcal{K}}} \sum_{i \in \mathcal{K}} \gamma_i \left(\ln \delta_i - \ln b_i\right) \quad \text{s.t.} \sum_{i \in \mathcal{K}} \delta_i = 1.$$

It is thus immediate that $\gamma_i^{\mathcal{K}} = \frac{\gamma_i}{\sum_{j \in \mathcal{K}} \gamma_j}$. By Proposition 3, the welfare gains from adopting the constrained optimal allocation is

$$\frac{\psi\lambda}{\rho^2}\left(\sum_{i \in \mathcal{K}} \gamma_i \left(\ln \gamma_i^{\mathcal{K}} \left(\sum_{i \in \mathcal{K}} b_i\right) - \ln b_i^{\mathcal{K}} \left(\sum_{i \in \mathcal{K}} b_i\right)\right) + \sum_{i \notin \mathcal{K}} \gamma_i \left(\ln b_i - \ln b_i\right)\right),$$

the consumption-equivalent gains then simplifies to the formula in the Proposition. □

The Proposition also holds in an environment with foreign spillovers, which we state below.

Proposition 10. *Consider an open economy with R&D self-sufficiency ξ and given paths of foreign knowledge and relative import prices $\left\{\boldsymbol{q}_t^f, p_t^f\right\}_{t=0}^{\infty}$. Suppose R&D allocations in sectors $k \notin \mathcal{K}$ are given exogenously and that the planner can only choose R&D allocations in sectors $k \in \mathcal{K}$*

when solving the planning problem in (20). Along the entire equilibrium path, the constrained optimal R&D allocation is $s_i = \gamma_i^{\mathcal{K}} \left(\bar{s} - \sum_{k \notin \mathcal{K}} s_k \right)$ *for* $i \in \mathcal{K}$. *The consumption-equivalent welfare gains from adopting the constrained-optimal R&D allocation (instead of allocation **b**) is* $\mathcal{L}^{\mathcal{K}}(\boldsymbol{b}) = \exp\left(\frac{\psi \lambda}{\rho} \xi \left(\sum_{j \in \mathcal{K}} \gamma_j \right) \left(\boldsymbol{\gamma}^{\mathcal{K}} \right)' \left(\ln \boldsymbol{\gamma}^{\mathcal{K}} - \ln \boldsymbol{b}^{\mathcal{K}} \right) \right)$.

B.6 Optimal R&D Allocation in Large Open Economies

In the open economy environment presented in the main text, we studied the problem of a domestic planner who takes the paths of import prices and foreign knowledge as given. In this appendix section, we construct an environment in which a domestic planner internalizes the impact of domestic allocations on foreign variables. This analysis is empirically relevant for studying the R&D allocation in the U.S., a country that generates significant knowledge spillovers to other economies.

Consider an environment with two economies, home (U.S.) and foreign (rest of the world). The home consumer has preferences

$$V = \int_0^\infty e^{-\rho t} \left(\sigma^h \ln c_t^{hh} + \left(1 - \sigma^h\right) \ln c_t^{hf} \right) \mathrm{d}t, \tag{A16}$$

where c_t^{hh} is the home consumption of home goods and c_t^{hf} is the home consumption of foreign goods. Home goods is a Cobb-Douglas aggregator over sectoral composite goods, which are aggregations of intermediate varieties produced from labor (equations 2 and 3). We can simplify the home production functions as

$$\ln y_t^h = \sum_i \beta_i \left(\psi \ln q_{it}^h + \ln \ell_{it}^h \right). \tag{A17}$$

Home can import the foreign goods c_t^{hf} by exporting unconsumed home goods $\left(y_t^h - c_t^{hh}\right)$. Home innovation production function follows

$$n_{it}^h = \eta_i^h s_{it}^h \chi_{it}^h, \qquad \text{where } \chi_{it}^h = \prod_{j=1}^K \left[\left(q_{jt}^h\right)^{x_{ij}^h} \left(q_{jt}^f\right)^{1 - x_{ij}^h} \right]^{\omega_{ij}}, \tag{A18}$$

and the law of motion for home knowledge stock is

$$\frac{\mathrm{d} \ln q_{it}^h}{\mathrm{d}t} = \lambda \ln \left(n_{it}^h / q_{it}^h \right). \tag{A19}$$

Home is endowed with workers $\bar{\ell}^h$ and scientists $\bar{s}^h$. The foreign economy has analogous preferences and technologies, swapping superscripts h and f.

We study the home planner's problem of allocating workers and scientists to maximize home welfare, while taking the time path of foreign allocations $\left\{ \boldsymbol{\ell}_t^f, \boldsymbol{s}_t^f \right\}$ as given and decentralizing international trade. Given home and foreign output y_t^h, y_t^f, Cobb-Douglas preferences imply that the home consumer spends $\left(1-\sigma^h\right)$ fraction of income on home imports, and that the foreign consumer spends $\left(1-\sigma^f\right)$ fraction of income on home exports. Trade balance therefore implies that home consumption of foreign goods is $\left(1-\sigma^f\right) y_t^f$. Hence, given flow output y_t^h, y_t^f, the home consumer's flow utility is

$$\sigma^h \ln c_t^{hh} + \left(1-\sigma^h\right) \ln c_t^{hf} = \sigma^h \ln \sigma^h y_t^h + \left(1-\sigma^h\right) \ln \left(1-\sigma^f\right) y_t^f.$$

Substituting into (A16), we can write the home planning problem as

$$V^*\left(\left\{\boldsymbol{\ell}_t^f, \boldsymbol{s}_t^f\right\}_{t=0}^{\infty}\right) \equiv \max_{\left\{s_{it}^h, \ell_{it}^h\right\}} \int_0^\infty e^{-\rho t}\left(\sigma^h \ln y_t^h + \left(1-\sigma^h\right) \ln y_t^f\right) \mathrm{d}t, \tag{A20}$$

subject to the innovation production functions (A19 and A18), goods production function (A17), and the corresponding foreign innovation and goods production functions

$$\frac{\mathrm{d} \ln q_{it}^f}{\mathrm{d}t} = \ln \eta_i^f + \ln s_{it}^f + \sum_{j=1}^K \omega_{ij}\left(x_{ij}^f \ln q_{jt}^f + \left(1-x_{ij}^f\right) \ln q_{jt}^h\right),$$

$$\ln y_t^f = \sum_i \beta_i \left(\ln q_{it}^f + \ln \ell_{it}^f\right),$$

with market clearing conditions $\sum_i s_{it}^h = \bar{s}^h$ and $\sum_i \ell_{it}^h = \bar{\ell}^h$.

To solve the home planner's problem, first consider a hypothetical world as an integrated economy in which resources can freely move across countries, and where the home planner can choose worker and scientist allocations in both economies; then, our closed economy analysis in Section 2.2 exactly applies: the solution would be characterized exactly by our closed economy results in Lemma 1 and Proposition 1, recognizing that there are $K \times 2$ sectors in both economies, with home's consumption elasticity captured by

$$\hat{\boldsymbol{\beta}} \equiv \left[\sigma^h \boldsymbol{\beta}', \left(1-\sigma^h\right) \boldsymbol{\beta}'\right], \tag{A21}$$

and the innovation network captured by

$$\hat{\boldsymbol{\Omega}} \equiv \begin{bmatrix} \boldsymbol{\Omega} \circ \boldsymbol{X}^h & \boldsymbol{\Omega} - \boldsymbol{\Omega} \circ \boldsymbol{X}^h \\ \boldsymbol{\Omega} - \boldsymbol{\Omega} \circ \boldsymbol{X}^f & \boldsymbol{\Omega} \circ \boldsymbol{X}^f \end{bmatrix}. \tag{A22}$$

Optimal worker allocation should follow $\hat{\boldsymbol{\beta}}$, and optimal R&D allocation should follow

$$\hat{\gamma}' \equiv \frac{\rho}{\rho+\lambda}\left(\boldsymbol{I}_{2K \times 2K} - \frac{\hat{\boldsymbol{\Omega}}}{1+\rho/\lambda}\right)^{-1}. \tag{A23}$$

Next, recognize that the actual home planner's problem (A20) is essentially the same as in the hypothetical integrated economy, but with the additional constraint that the home planner can only allocate resources domestically. We can apply the result in Section B.5 to get the following Proposition.

Proposition 11. *The optimal resource allocation for an open economy planner who takes the path of foreign allocations $\left\{\boldsymbol{\ell}_t^f, \boldsymbol{s}_t^f\right\}$ as given and solves the problem in (A20) is to allocate workers according to $\hat{\boldsymbol{\beta}}^{\mathcal{K}}$ (i.e., $\ell_{it}^h/\bar{\ell}^h = \hat{\beta}_i^{\mathcal{K}}$) and R&D resources according to $\hat{\boldsymbol{\gamma}}^{\mathcal{K}}$ (i.e., $s_{it}^h/\bar{s}^h = \hat{\gamma}_i^{\mathcal{K}}$), where $\mathcal{K}$ is the set of domestic sectors, and*

$$\hat{\beta}_i^{\mathcal{K}} = \frac{\hat{\beta}_i}{\sum_{j\in\mathcal{K}}\hat{\beta}_j}, \qquad \hat{\gamma}_i^{\mathcal{K}} = \frac{\hat{\gamma}_i}{\sum_{j\in\mathcal{K}}\hat{\gamma}_j}.$$

The consumption-equivalent welfare gains from adopting the optimal domestic R&D allocation (instead of allocation $\boldsymbol{b}$) is $\mathcal{L}^{\mathcal{K}}(\boldsymbol{b}) = \exp\left(\frac{\psi\lambda}{\rho}\left(\sum_{j\in\mathcal{K}}\hat{\gamma}_j\right)\left(\hat{\boldsymbol{\gamma}}^{\mathcal{K}}\right)'\left(\ln\hat{\boldsymbol{\gamma}}^{\mathcal{K}} - \ln\boldsymbol{b}\right)\right)$.

B.7 General Functional Forms and Endogenous Innovation Network with Foreign Spillovers

We now extend our analysis in Section 2.6 to incorporate general functional forms, thereby endogenizing the degree to which domestic innovation benefits from foreign spillovers. We show, analogous to our closed-economy analysis in Section 2.5, that Proposition 7 in the main text continues to hold, as a first-order approximation around a balanced growth path, to the welfare impact of adopting the optimal R&D allocation.

For completeness, we provide all equations to this economic environment:

$$V\left(\left\{q_{jt}^f, p_t^f\right\}\right) = \int_0^\infty e^{-\rho t}\ln\mathcal{C}\left(c_t^d, c_t^f\right)\,\mathrm{d}t,$$

$$p_t^f c_t^f = y_t - c_t^d.$$

$$y_t = \mathcal{Y}\left(\left\{q_{it}^{\psi}\ell_{it}\right\}\right)$$

$$\mathrm{d}\ln q_{it}/\,\mathrm{d}t = \lambda\cdot\left(\ln\left(\eta_i b_{it}\bar{s}\right) + \ln\mathcal{X}_i\left(\left\{q_{jt}, q_{jt}^f\right\}\right)\right)$$

The first equation represents consumer welfare; the second equation is trade balance; the third equation is the production function; the last equation is the law of motion for sectoral knowledge stock. The function $\mathcal{X}_i\left(\left\{q_{jt}, q_{jt}^f\right\}\right)$ captures how domestic innovation in sector i benefits from domestic and foreign knowledge; it is a generalization of the Cobb-Douglas functional form in equation (19). We assume $\mathcal{C}$ and $\mathcal{Y}$ are constant-returns-to-scale, and that $\mathcal{X}_i(\cdot)$ is homogeneous-of-degree-zero, $\partial\ln\mathcal{X}_i(\cdot)/\partial\ln q_{jt} \geq 0\,\forall i\neq j$, $\partial\ln\mathcal{X}_i(\cdot)/\partial\ln q_{jt}^f\,\forall i,j$, and $\left|\partial\ln\mathcal{X}_i(\cdot)/\partial\ln q_{jt}\right| \leq 1\forall i,j$.

Consider the economy initially at $t = 0$ in a BGP with R&D allocation $\boldsymbol{b}$, where foreign knowledge q_{jt}^f grows at exogenous rate g in all sectors, and p_t^f is time-invariant. Define

$$\beta_i \equiv \left.\frac{\partial \ln \mathcal{Y}\left(\{y_{it}\}\right)}{\partial \ln y_{it}}\right|_{t=0}, \qquad \theta_{ij} \equiv \begin{cases} \left.\frac{\partial \ln \mathcal{X}_i\left(\left\{q_{it}, q_{jt}^f\right\}\right)}{\partial \ln q_{jt}}\right|_{t=0} & \text{if } i = j \\ 1 + \left.\frac{\partial \ln \mathcal{X}_i\left(\left\{q_{it}, q_{jt}^f\right\}\right)}{\partial \ln q_{jt}}\right|_{t=0} & \text{otherwise.} \end{cases}$$

$\boldsymbol{\beta} \equiv [\beta_i]$ and $\boldsymbol{\Theta} \equiv [\theta_{ij}]$ are the consumption and innovation spillover elasticities with respect to domestic knowledge stock evaluated in the initial BGP. Note that $\boldsymbol{I} - \boldsymbol{\Theta}$ is generically invertible, as the economy features aggregate decreasing-returns-to-scale with respect to domestic knowledge stock.

The Gateaux derivative of welfare with respect to log R&D allocation in the direction of $\boldsymbol{h}$ is $\frac{\lambda}{\rho}\xi\boldsymbol{\gamma}'\boldsymbol{h}$; the proof parallels that of Proposition 5.

B.8 Sector-Specific Innovation Step Size

We now introduce a theoretical extension allowing for sector-specific innovation step size λ_i. Let $\boldsymbol{\Lambda} \equiv \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_K \end{bmatrix}$ denote the diagonal matrix with λ_i along the diagonal, and let $\boldsymbol{\lambda} \equiv [\lambda_i]$ denote the vector of λ_i's. We show the optimal R&D allocation $\boldsymbol{\gamma}$ should follow (scaled so that $\boldsymbol{\gamma}$ sums to one)

$$\boldsymbol{\gamma}' \propto \boldsymbol{\beta}' \left(\boldsymbol{I} - \boldsymbol{\Omega} + \rho\boldsymbol{\Lambda}^{-1}\right)^{-1}$$

and the consumption-equivalent welfare impact of adopting the optimal allocation is

$$\mathcal{L}\left(\boldsymbol{b}\right) = \exp\left(\psi\boldsymbol{\beta}' \left(\boldsymbol{I} - \boldsymbol{\Omega} + \rho\boldsymbol{\Lambda}^{-1}\right)^{-1} \left(\ln\boldsymbol{\gamma} - \ln\boldsymbol{b}\right)\right).$$

Specifically, the social planner's problem is

$$\max_{\{\gamma_t\} \text{ s.t. } \gamma_t'1=1\forall t} \int_0^\infty e^{-\rho t}\boldsymbol{\beta}' \ln \boldsymbol{q}_t \, \mathrm{d}t$$
$$\text{s.t. } \mathrm{d}\ln\boldsymbol{q}_t/\mathrm{d}t = \boldsymbol{\Lambda}\left(\ln\boldsymbol{\eta} + \ln\bar{s}\boldsymbol{1} + \ln\boldsymbol{\gamma}_t + \left(\boldsymbol{\Omega} - \boldsymbol{I}\right)\ln\boldsymbol{q}_t\right) \tag{A24}$$

The control variable is $\boldsymbol{\gamma}_t$ and the state variable is $\boldsymbol{q}_t$. Denote the co-state variables as $\boldsymbol{\mu}_t$. The current-value Hamiltonian is

$$H(\gamma_t, \boldsymbol{q}_t, \boldsymbol{\mu}_t, \zeta) = \boldsymbol{\beta}'\ln\boldsymbol{q}_t + \boldsymbol{\mu}_t'\boldsymbol{\Lambda}\left[\ln\boldsymbol{\eta} + \ln\bar{s}\boldsymbol{1} + \ln\boldsymbol{\gamma}_t + \left(\boldsymbol{\Omega} - \boldsymbol{I}\right)\ln\boldsymbol{q}_t\right] + \zeta\left(1 - \boldsymbol{\gamma}_t'\boldsymbol{1}\right).$$

For notational simplicity we suppress dependence on time for control, state, and co-state variables:

$$H(\{\gamma_i\},\{q_i\},\{\mu_i\},\zeta,t)=\sum_i\beta_i\ln q_i+\sum_i\mu_i\lambda_i\left(\ln\eta_i+\ln\bar{s}+\ln\gamma_i+\sum_j\omega_{ij}\ln q_j-\ln q_i\right)+\zeta(1-\sum_i\gamma_i).$$

By the maximum principle

$$H_{\gamma_i}=0\Longleftrightarrow\frac{\lambda_i\mu_i}{\gamma_i}=\zeta\quad\forall i \tag{A25}$$

$$H_{\ln q_i}=\rho\mu_i-\dot{\mu}_i\Longleftrightarrow\beta_i-\lambda_i\mu_i+\sum_j\lambda_j\mu_j\omega_{ji}=\rho\mu_i-\dot{\mu}_i \tag{A26}$$

Similar to the proof of Proposition 1, we can show $\dot{\mu}_i=0$ for all i; hence,

$$\boldsymbol{\gamma}'\propto\boldsymbol{\mu}'\boldsymbol{\Lambda},$$

$$\begin{aligned}\boldsymbol{\beta}' &= \boldsymbol{\mu}'\left((\rho+\boldsymbol{\Lambda})\boldsymbol{I}-\boldsymbol{\Lambda}\boldsymbol{\Omega}\right)\\ &= \boldsymbol{\mu}'\boldsymbol{\Lambda}\left(\rho\boldsymbol{\Lambda}^{-1}+\boldsymbol{I}-\boldsymbol{\Omega}\right)\\ &\propto \boldsymbol{\gamma}'\left(\rho\boldsymbol{\Lambda}^{-1}+\boldsymbol{I}-\boldsymbol{\Omega}\right)\end{aligned}$$

Hence

$$\boldsymbol{\gamma}'=\boldsymbol{\beta}'\left(\boldsymbol{I}-\boldsymbol{\Omega}+\rho\boldsymbol{\Lambda}^{-1}\right)^{-1}.$$

To derive the welfare impact of R&D reallocation, let $g_i^q\equiv\frac{\mathrm{d}\ln q_{it}}{\mathrm{d}t}$ be the growth rate of knowledge stock in sector i along the BGP. We know

$$\boldsymbol{g}^q=\boldsymbol{\Lambda}\left(\ln\boldsymbol{\eta}+\ln\bar{s}\mathbf{1}+\ln\boldsymbol{b}+(\boldsymbol{\Omega}-\boldsymbol{I})\ln\boldsymbol{q}_t\right)$$

Take derivative with respect to time,

$$\mathbf{0}=\boldsymbol{\Lambda}\left(\boldsymbol{\Omega}-\boldsymbol{I}\right)\frac{\mathrm{d}\ln\boldsymbol{q}_t}{\mathrm{d}t}$$

So that

$$\boldsymbol{g}^q=\boldsymbol{\Omega}\boldsymbol{g}^q$$

We know the only right-Perron eigenvector of $\boldsymbol{\Omega}$ is the constant vector; hence all sectors must grow at the same rate g^q, satisfying

$$g^q\mathbf{1}=\boldsymbol{\Lambda}\left(\ln\boldsymbol{\eta}+\ln\bar{s}\mathbf{1}+\ln\boldsymbol{b}+(\boldsymbol{\Omega}-\boldsymbol{I})\ln\boldsymbol{q}_t\right)$$

$$\Longrightarrow\ g^q\boldsymbol{a}'\boldsymbol{\Lambda}^{-1}\mathbf{1}=\boldsymbol{a}'\left(\ln\boldsymbol{\eta}+\ln\bar{s}\mathbf{1}+\ln\boldsymbol{b}+(\boldsymbol{\Omega}-\boldsymbol{I})\ln\boldsymbol{q}_t\right)$$

$$\implies g^q = \frac{\boldsymbol{a}'\left(\ln\boldsymbol{\eta} + \ln\bar{s}\mathbf{1} + \ln\boldsymbol{b}\right)}{\boldsymbol{a}'\boldsymbol{\Lambda}^{-1}\mathbf{1}}$$

Let $\boldsymbol{A} \equiv \frac{\mathbf{1}\boldsymbol{a}'\boldsymbol{\Lambda}^{-1}}{\boldsymbol{a}'\boldsymbol{\Lambda}^{-1}\mathbf{1}}$. Note $(\boldsymbol{I}-\boldsymbol{A})\ln\bar{s}\mathbf{1} = \mathbf{0}$, and that

$$\begin{aligned}
(\boldsymbol{\Omega}-\boldsymbol{I}) &= (\boldsymbol{\Omega}-\boldsymbol{I})(\boldsymbol{I}-\boldsymbol{A}) \\
&= -(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})(\boldsymbol{I}-\boldsymbol{A})
\end{aligned}$$

Let $\widetilde{\ln\boldsymbol{q}_t} \equiv (\boldsymbol{I}-\boldsymbol{A})\ln\boldsymbol{q}_t$; then

$$\begin{aligned}
(\boldsymbol{I}-\boldsymbol{A})\,\mathrm{d}\ln\boldsymbol{q}_t/\mathrm{d}t &= \left(\boldsymbol{\Lambda} - \frac{\mathbf{1}\boldsymbol{a}'}{\boldsymbol{a}'\boldsymbol{\Lambda}^{-1}\mathbf{1}}\right)\left(\ln\boldsymbol{\eta}+\ln\bar{s}\mathbf{1}+\ln\boldsymbol{b}\right) + \boldsymbol{\Lambda}(\boldsymbol{\Omega}-\boldsymbol{I})\ln\boldsymbol{q}_t \\
&= (\boldsymbol{I}-\boldsymbol{A})\boldsymbol{\Lambda}\left(\ln\boldsymbol{\eta}+\ln\boldsymbol{b}\right) - \boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})(\boldsymbol{I}-\boldsymbol{A})\ln\boldsymbol{q}_t
\end{aligned}$$

$$\frac{\mathrm{d}\widetilde{\ln\boldsymbol{q}_t}}{\mathrm{d}t} = (\boldsymbol{I}-\boldsymbol{A})\boldsymbol{\Lambda}\left(\ln\boldsymbol{\eta}+\ln\boldsymbol{b}\right) - \boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})(\boldsymbol{I}-\boldsymbol{A})\widetilde{\ln\boldsymbol{q}_t}$$

We can integrate the ODE system:

$$\begin{aligned}
\widetilde{\ln\boldsymbol{q}_t} &= e^{-\boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t}\left[\widetilde{\ln\boldsymbol{q}_0} + \int_0^t e^{\boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})s}(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{\Lambda}\left(\ln\boldsymbol{\eta}+\ln\boldsymbol{b}\right)\,\mathrm{d}s\right] \\
&= e^{-\boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t}\widetilde{\ln\boldsymbol{q}_0} + \boldsymbol{\Lambda}^{-1}(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})^{-1}\left[\boldsymbol{I}-e^{-\boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t}\right](\boldsymbol{I}-\boldsymbol{A})\boldsymbol{\Lambda}\left(\ln\boldsymbol{\eta}+\ln\boldsymbol{b}\right)
\end{aligned}$$

We know

$$\begin{aligned}
\boldsymbol{A}\frac{\mathrm{d}\ln\boldsymbol{q}_t}{\mathrm{d}t} &= \frac{\mathbf{1}\boldsymbol{a}'\boldsymbol{\Lambda}^{-1}}{\boldsymbol{a}'\boldsymbol{\Lambda}^{-1}\mathbf{1}}\boldsymbol{\Lambda}\left(\ln\boldsymbol{\eta}+\ln\bar{s}\mathbf{1}+\ln\boldsymbol{b}+(\boldsymbol{\Omega}-\boldsymbol{I})\ln\boldsymbol{q}_t\right) \\
&= \frac{\mathbf{1}\boldsymbol{a}'}{\boldsymbol{a}'\boldsymbol{\Lambda}^{-1}\mathbf{1}}\left(\ln\boldsymbol{\eta}+\ln\bar{s}\mathbf{1}+\ln\boldsymbol{b}\right) \\
&= \boldsymbol{A}\boldsymbol{\Lambda}\left(\ln\boldsymbol{\eta}+\ln\bar{s}\mathbf{1}+\ln\boldsymbol{b}\right)
\end{aligned}$$

Hence

$$\boldsymbol{A}\ln\boldsymbol{q}_t(\boldsymbol{b}) = \boldsymbol{A}\ln\boldsymbol{q}_0 + \boldsymbol{A}\boldsymbol{\Lambda}\left(\ln\boldsymbol{\eta}+\ln\bar{s}\mathbf{1}+\ln\boldsymbol{b}\right)t$$

Now consider starting from the same initial knowledge stock $\boldsymbol{q}_0$ but with two different time-invariant R&D allocations $\widetilde{\boldsymbol{b}}$ and $\boldsymbol{b}$,

$$\boldsymbol{A}\ln\boldsymbol{q}_t\left(\widetilde{\boldsymbol{b}}\right) - \boldsymbol{A}\ln\boldsymbol{q}_t(\boldsymbol{b}) = \boldsymbol{A}\boldsymbol{\Lambda}\left(\ln\widetilde{\boldsymbol{b}} - \ln\boldsymbol{b}\right)t$$

we have the following difference in knowledge stock over time:

$$\begin{aligned}
\ln\boldsymbol{q}_t\left(\widetilde{\boldsymbol{b}}\right) - \ln\boldsymbol{q}_t(\boldsymbol{b}) &= \boldsymbol{A}\ln\boldsymbol{q}_t\left(\widetilde{\boldsymbol{b}}\right) - \boldsymbol{A}\ln\boldsymbol{q}_t(\boldsymbol{b}) \\
&\quad + \widetilde{\ln\boldsymbol{q}_t}\left(\widetilde{\boldsymbol{b}}\right) - \widetilde{\ln\boldsymbol{q}_t}(\boldsymbol{b}) \\
&= \left[\boldsymbol{A}\boldsymbol{\Lambda}t + \boldsymbol{\Lambda}^{-1}(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})^{-1}\left[\boldsymbol{I}-e^{-\boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t}\right](\boldsymbol{I}-\boldsymbol{A})\boldsymbol{\Lambda}\right]\left(\ln\widetilde{\boldsymbol{b}} - \ln\boldsymbol{b}\right)
\end{aligned}$$

The difference in consumer welfare under two time-invariant paths of R&D allocations is

$$
\begin{aligned}
& V\left(\boldsymbol{q}_0;\{\boldsymbol{\ell}_t\},\widetilde{\boldsymbol{b}}\right)-V\left(\boldsymbol{q}_0;\{\boldsymbol{\ell}_t\},\boldsymbol{b}\right) \\
= \; & \psi\boldsymbol{\beta}'\int_0^\infty e^{-\rho t}\left[\ln\boldsymbol{q}_t\left(\widetilde{\boldsymbol{b}}\right)-\ln\boldsymbol{q}_t\left(\boldsymbol{b}\right)\right]\,\mathrm{d}t \\
= \; & \psi\boldsymbol{\beta}'\int_0^\infty e^{-\rho t}\left[\boldsymbol{A}\boldsymbol{\Lambda}t+\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A}\right)^{-1}\left[\boldsymbol{I}-e^{-\boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t}\right]\left(\boldsymbol{I}-\boldsymbol{A}\right)\boldsymbol{\Lambda}\right]\,\mathrm{d}t\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right) \\
= \; & \frac{\psi}{\rho^2}\boldsymbol{\beta}'\boldsymbol{A}\boldsymbol{\Lambda}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right) \\
& +\psi\boldsymbol{\beta}'\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A}\right)^{-1}\left[\frac{1}{\rho}\boldsymbol{I}-\int_0^\infty\left(e^{-((\rho\boldsymbol{I}+\boldsymbol{\Lambda})\boldsymbol{I}-\boldsymbol{\Lambda}(\boldsymbol{\Omega}-\boldsymbol{A}))t}\right)\,\mathrm{d}t\right]\left(\boldsymbol{I}-\boldsymbol{A}\right)\boldsymbol{\Lambda}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right) \\
= \; & \frac{\psi}{\rho}\boldsymbol{\beta}'\left\{\frac{1}{\rho}\boldsymbol{A}\boldsymbol{\Lambda}+\left(\left[\left(\rho\boldsymbol{I}+\boldsymbol{\Lambda}\right)-\boldsymbol{\Lambda}\left(\boldsymbol{\Omega}-\boldsymbol{A}\right)\right]\right)^{-1}\left(\boldsymbol{I}-\boldsymbol{A}\right)\boldsymbol{\Lambda}\right\}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right) \\
= \; & \frac{\psi}{\rho}\boldsymbol{\beta}'\left(\left[\left(\rho\boldsymbol{I}+\boldsymbol{\Lambda}\right)-\boldsymbol{\Lambda}\left(\boldsymbol{\Omega}-\boldsymbol{A}\right)\right]\right)^{-1}\left(\left[\boldsymbol{I}+\frac{1}{\rho}\boldsymbol{\Lambda}\left(\boldsymbol{I}-\left(\boldsymbol{\Omega}-\boldsymbol{A}\right)\right)\right]\right)\boldsymbol{A}\boldsymbol{\Lambda}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right) \\
& +\frac{\psi}{\rho}\boldsymbol{\beta}'\left(\left[\left(\rho\boldsymbol{I}+\boldsymbol{\Lambda}\right)-\boldsymbol{\Lambda}\left(\boldsymbol{\Omega}-\boldsymbol{A}\right)\right]\right)^{-1}\left(\boldsymbol{I}-\boldsymbol{A}\right)\boldsymbol{\Lambda}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right) \\
= \; & \frac{\psi}{\rho^2}\boldsymbol{\beta}'\left(\left[\left(\rho\boldsymbol{I}+\boldsymbol{\Lambda}\right)-\boldsymbol{\Lambda}\left(\boldsymbol{\Omega}-\boldsymbol{A}\right)\right]\right)^{-1}\left(\rho\boldsymbol{I}+\boldsymbol{\Lambda}\boldsymbol{A}\right)\boldsymbol{\Lambda}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right) \\
= \; & \frac{\psi}{\rho^2}\boldsymbol{\beta}'\left(\left[\boldsymbol{I}-\boldsymbol{\Omega}+\rho\boldsymbol{\Lambda}^{-1}+\boldsymbol{A}\right]\right)^{-1}\left(\rho\boldsymbol{I}+\boldsymbol{A}\boldsymbol{\Lambda}\right)\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right) \\
= \; & \frac{\psi}{\rho^2}\boldsymbol{\beta}'\left(\left(\rho\boldsymbol{I}+\boldsymbol{A}\boldsymbol{\Lambda}\right)^{-1}\left[\boldsymbol{I}-\boldsymbol{\Omega}+\rho\boldsymbol{\Lambda}^{-1}+\boldsymbol{A}\right]\right)^{-1}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right)
\end{aligned}
$$

Let $\alpha\equiv\boldsymbol{a}'\boldsymbol{\Lambda}^{-1}\mathbf{1}$. Note

$$
\left(\rho\boldsymbol{I}+\boldsymbol{A}\boldsymbol{\Lambda}\right)^{-1}=\frac{1}{\rho}\left(\boldsymbol{I}-\frac{\alpha}{1+\alpha\rho}\boldsymbol{A}\boldsymbol{\Lambda}\right)
$$

To see this,

$$
\begin{aligned}
& \left(\rho\boldsymbol{I}+\boldsymbol{A}\boldsymbol{\Lambda}\right)\frac{1}{\rho}\left(\boldsymbol{I}-\frac{\alpha}{1+\alpha\rho}\boldsymbol{A}\boldsymbol{\Lambda}\right) \\
= \; & \frac{1}{\rho}\left(\rho\boldsymbol{I}+\boldsymbol{A}\boldsymbol{\Lambda}-\left(\rho\boldsymbol{I}+\boldsymbol{A}\boldsymbol{\Lambda}\right)\frac{\alpha}{1+\alpha\rho}\boldsymbol{A}\boldsymbol{\Lambda}\right) \\
= \; & \boldsymbol{I}+\frac{1}{\rho}\left(\frac{\mathbf{1}\boldsymbol{a}'}{\alpha}-\left(\rho\boldsymbol{I}+\frac{\mathbf{1}\boldsymbol{a}'}{\alpha}\right)\frac{1}{1+\alpha\rho}\mathbf{1}\boldsymbol{a}'\right) \\
= \; & \boldsymbol{I}+\frac{1}{\rho}\left(\frac{1}{\alpha}-\frac{1+\alpha\rho}{\alpha}\frac{1}{1+\alpha\rho}\right)\mathbf{1}\boldsymbol{a}' \\
= \; & \boldsymbol{I}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& V\left(\boldsymbol{q}_0;\{\boldsymbol{\ell}_t\},\widetilde{\boldsymbol{b}}\right)-V\left(\boldsymbol{q}_0;\{\boldsymbol{\ell}_t\},\boldsymbol{b}\right)\\
&=\frac{\psi}{\rho^2}\boldsymbol{\beta}'\left(\frac{1}{\rho}\left(\boldsymbol{I}-\frac{\alpha}{1+\alpha\rho}\boldsymbol{A}\boldsymbol{\Lambda}\right)\left[\boldsymbol{I}-\boldsymbol{\Omega}+\rho\boldsymbol{\Lambda}^{-1}+\boldsymbol{A}\right]\right)^{-1}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right)\\
&=\frac{\psi}{\rho}\boldsymbol{\beta}'\left(\left[\boldsymbol{I}-\boldsymbol{\Omega}+\rho\boldsymbol{\Lambda}^{-1}+\frac{\mathbf{1}\boldsymbol{a}'\boldsymbol{\Lambda}^{-1}}{\alpha}\right]-\frac{\mathbf{1}\boldsymbol{a}'}{1+\alpha\rho}\left[\boldsymbol{I}-\boldsymbol{\Omega}+\rho\boldsymbol{\Lambda}^{-1}+\frac{\mathbf{1}\boldsymbol{a}'\boldsymbol{\Lambda}^{-1}}{\alpha}\right]\right)^{-1}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right)\\
&=\frac{\psi}{\rho}\boldsymbol{\beta}'\left(\left[\boldsymbol{I}-\boldsymbol{\Omega}+\rho\boldsymbol{\Lambda}^{-1}+\frac{\mathbf{1}\boldsymbol{a}'\boldsymbol{\Lambda}^{-1}}{\alpha}\right]-\frac{\mathbf{1}\boldsymbol{a}'\boldsymbol{\Lambda}^{-1}}{\alpha}\right)^{-1}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right)\\
&=\frac{\psi}{\rho}\boldsymbol{\beta}'\left(\boldsymbol{I}-\boldsymbol{\Omega}+\rho\boldsymbol{\Lambda}^{-1}\right)^{-1}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right)
\end{aligned}
$$

Following the proof of Proposition 4, the consumption-equivalent welfare impact of adopting the optimal allocation is thus

$$
\mathcal{L}(\boldsymbol{b})=\psi\boldsymbol{\beta}'\left(\boldsymbol{I}-\boldsymbol{\Omega}+\rho\boldsymbol{\Lambda}^{-1}\right)^{-1}\left(\ln\boldsymbol{\gamma}-\ln\boldsymbol{b}\right).
$$

B.9 Innovation Network with Heterogeneous Row-Sums

Our baseline specification of $\boldsymbol{\Omega}$ assumes that each row sums to one (i.e., $\boldsymbol{\Omega}\mathbf{1}=\mathbf{1}$, so that $\boldsymbol{\Omega}$ is a row-stochastic Markov matrix). Because the spectral radius of any Markov matrix is equal to one, our baseline model is one with endogenous growth. The specification also motivates our measurement of the innovation network based on patent citations, $\omega_{ij}\equiv\frac{Cites_{ij}}{\sum_{k=1}^{K}Cites_{ik}}$.

In general, the knowledge spillover network is inherently difficult to measure. A reason alternative specification is to construct the network as $\omega_{ij}\propto Cites_{ij}$. This specification results in an innovation network matrix $\boldsymbol{\Omega}$ with heterogeneous row-sums ($\sum_j\omega_{ij}$ varies with i). The proportionality constant maps monotonically into the spectral radius of $\boldsymbol{\Omega}$. The model features endogenous (semi-endogenous) growth if the spectral radius is equal to (less than) one.[14]

Propositions 1 extends directly to the case where the spectral radius of $\boldsymbol{\Omega}$ is ≤ 1, as the proof does not make use of the fact that $\boldsymbol{\Omega}$ is row-stochastic. We now show Propositions 4 and 4 both hold in the endogenous growth case. Analogous results can be derived (but omitted here) in the semi-endogenous growth case as well.

Let $\boldsymbol{v}$ denote the right-Perron eigenvector of $\boldsymbol{\Omega}$, scaled so that $\boldsymbol{a}'\boldsymbol{v}=1$. Let $\boldsymbol{A}\equiv\boldsymbol{v}\boldsymbol{a}'$. We adapt the derivations in the proof of Proposition 3 to this setting, replacing $\boldsymbol{A}\equiv\mathbf{1}\boldsymbol{a}'$ in the baseline proof to $\boldsymbol{A}\equiv\boldsymbol{v}\boldsymbol{a}'$. Note that in the baseline setting where $\boldsymbol{\Omega}$ is row-stochastic, $\boldsymbol{v}=\mathbf{1}$, so the derivation below is a strict generalization.

[14]The model features explosive growth if the spectral radius of $\boldsymbol{\Omega}$ is greater than one.

For time-invariant R&D allocation $\boldsymbol{b}$, the law of motion of sectoral knowledge stock implies

$$\boldsymbol{a}'\frac{\mathrm{d}\ln\boldsymbol{q}}{\mathrm{d}t}=\lambda\boldsymbol{a}'\left(\ln\boldsymbol{\eta}+\ln\bar{s}\cdot\mathbf{1}+\ln\boldsymbol{b}\right)$$

Hence, $\boldsymbol{a}'\ln\boldsymbol{q}_t$ always grows at a constant rate (and it equals to the rate of growth along a BGP) and can be solved in closed-form:

$$\boldsymbol{a}'\ln\boldsymbol{q}_t=\boldsymbol{a}'\ln\boldsymbol{q}_0+\lambda\boldsymbol{a}'\left(\ln\boldsymbol{\eta}+\ln\bar{s}\cdot\mathbf{1}+\ln\boldsymbol{b}\right)t$$

Note that $\boldsymbol{\Omega}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{\Omega}=\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}=\boldsymbol{A}$. Hence

$$\left(\boldsymbol{I}-\boldsymbol{A}\right)\left(\boldsymbol{\Omega}-\boldsymbol{I}\right)=\left(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A}\right)\left(\boldsymbol{I}-\boldsymbol{A}\right)$$

Left-multiply the law of motion by $(\boldsymbol{I}-\boldsymbol{A})$, substitute the above, and let $\widetilde{\ln\boldsymbol{q}_t}\equiv\left(\boldsymbol{I}-\boldsymbol{A}\right)\ln\boldsymbol{q}_t$, we get

$$\frac{\mathrm{d}\widetilde{\ln\boldsymbol{q}_t}}{\mathrm{d}t}=\lambda\left(\boldsymbol{I}-\boldsymbol{A}\right)\left(\ln\boldsymbol{\eta}+\ln\boldsymbol{b}\right)-\lambda\left(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A}\right)\widetilde{\ln\boldsymbol{q}_t}$$

Following the proof of Proposition 3,

$$\begin{aligned}\widetilde{\ln\boldsymbol{q}_t} &= e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t}\left[\widetilde{\ln\boldsymbol{q}_0}+\lambda\int_0^t e^{\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})s}\left(\boldsymbol{I}-\boldsymbol{A}\right)\left(\ln\boldsymbol{\eta}+\ln\boldsymbol{b}\right)\,\mathrm{d}s\right]\\ &= e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t}\widetilde{\ln\boldsymbol{q}_0}+\left(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A}\right)^{-1}\left(\boldsymbol{I}-e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t}\right)\,\mathrm{d}s\left(\boldsymbol{I}-\boldsymbol{A}\right)\left(\ln\boldsymbol{\eta}+\ln\boldsymbol{b}\right)\end{aligned}$$

$$\begin{aligned}\ln\boldsymbol{q}_t &= \widetilde{\ln\boldsymbol{q}_t}+\boldsymbol{A}\ln\boldsymbol{q}_t\\ &= \boldsymbol{A}\ln\boldsymbol{q}_0+\lambda\boldsymbol{A}\left(\ln\boldsymbol{\eta}+\ln\bar{s}\cdot\mathbf{1}+\ln\boldsymbol{b}\right)t\\ &\quad +e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t}\widetilde{\ln\boldsymbol{q}_0}+\left(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A}\right)^{-1}\left(\boldsymbol{I}-e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t}\right)\,\mathrm{d}s\left(\boldsymbol{I}-\boldsymbol{A}\right)\left(\ln\boldsymbol{\eta}+\ln\boldsymbol{b}\right)\end{aligned}$$

Starting from the same initial knowledge stock $\boldsymbol{q}_0$ but with two different time-invariant R&D allocations $\widetilde{\boldsymbol{b}}$ and $\boldsymbol{b}$, we have the following difference in knowledge stock over time:

$$\begin{aligned}\ln\boldsymbol{q}_t\left(\widetilde{\boldsymbol{b}}\right)-\ln\boldsymbol{q}_t\left(\boldsymbol{b}\right) &= \boldsymbol{A}\ln\boldsymbol{q}_t\left(\widetilde{\boldsymbol{b}}\right)-\boldsymbol{A}\ln\boldsymbol{q}_t\left(\boldsymbol{b}\right)\\ &\quad +\widetilde{\ln\boldsymbol{q}_t}\left(\widetilde{\boldsymbol{b}}\right)-\widetilde{\ln\boldsymbol{q}_t}\left(\boldsymbol{b}\right)\\ &= \left[\boldsymbol{A}\lambda t+\left(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A}\right)^{-1}\left(\boldsymbol{I}-e^{-\lambda(\boldsymbol{I}-\boldsymbol{\Omega}+\boldsymbol{A})t}\right)\left(\boldsymbol{I}-\boldsymbol{A}\right)\right]\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right)\end{aligned}$$

The difference in consumer welfare under two time-invariant paths of R&D allocations is (derivation follows from the proof of Proposition 3)

$$\begin{aligned}
& V\left(\boldsymbol{q}_0;\{\boldsymbol{\ell}_t\},\widetilde{\boldsymbol{b}}\right)-V\left(\boldsymbol{q}_0;\{\boldsymbol{\ell}_t\},\boldsymbol{b}\right) \\
= \; & \psi\boldsymbol{\beta}'\int_0^{\infty} e^{-\rho t}\left[\ln\boldsymbol{q}_t\left(\widetilde{\boldsymbol{b}}\right)-\ln\boldsymbol{q}_t\left(\boldsymbol{b}\right)\right]\,\mathrm{d}t \\
= \; & \frac{\psi}{\rho}\frac{\lambda}{\rho+\lambda}\boldsymbol{\beta}'\left(\boldsymbol{I}-\frac{\lambda}{\rho+\lambda}\boldsymbol{\Omega}\right)^{-1}\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right) \\
= \; & \frac{\psi\lambda}{\rho^2}\boldsymbol{\gamma}'\left(\ln\widetilde{\boldsymbol{b}}-\ln\boldsymbol{b}\right).
\end{aligned}$$

which establishes Proposition 3 in this setting where $\boldsymbol{\Omega}$ is not row-stochastic (but has spectral radius equal to one). Proposition 4 follows immediately.

B.10 Resource Mobility Between Production and R&D

In the closed economy analysis in the main text, we assumed the endowments of production workers $\bar{\ell}$ and scientists $\bar{s}$ are both exogenous. We now argue that the optimal allocation shares $\ell_{it}/\bar{\ell}$ and $s_{it}/\bar{s}$ characterized in Lemma 1 and Proposition 1 continue to hold even if agents in the economy can endogenously choose to become workers or scientists.

First, note that the proofs of Lemma 1 and Proposition 1 continue to hold even if the exogenous endowments of workers and scientists are time-varying. Let $V\left(\boldsymbol{q}_0;\left\{\bar{\ell}_t\right\},\left\{\bar{s}_t\right\}\right)$ denote the planner's value function, where the masses of workers and scientists are both exogenous along the entire growth path. The value function (7) in the main text corresponds to the special case where $\bar{\ell}_t=\bar{\ell}$ and $\bar{s}_t=\bar{s}$.

Now assume the economy is endowed with a unit mass of agents who can freely choose to become workers or scientists, $\bar{\ell}_t+\bar{s}_t=1$. The value function that solves the relaxed problem, where $\bar{\ell}_t$ and $\bar{s}_t$ are endogenous, can be written as

$$V\left(\boldsymbol{q}_0\right)=\max_{\left\{\bar{\ell}_t,\bar{s}_t\right\}} V\left(\boldsymbol{q}_0;\left\{\bar{\ell}_t\right\},\left\{\bar{s}_t\right\}\right) \quad \text{s.t. } \bar{\ell}_t+\bar{s}_t=1.$$

Since the optimal allocation shares of workers ($\ell_{it}/\bar{\ell}$) and scientists ($s_{it}/\bar{s}$) are invariant to the total mass of workers and scientists, it follows directly that the solution characterized in Section 2.2 continues to hold in the relaxed problem.

C Details on Data Construction

In this appendix, we provide details on data collection and harmonization and robustness of our approach.

C.1 U.S. Patent Data

U.S. patent data are obtained from the United States Patent and Trademark Office (USPTO).[15] The data include information on patent inventors and patent assignee, allowing us to identify the geographic locations of the innovation (e.g., identifying cases in which a Chinese firm is granted a USPTO patent). We also observe the timing of the patents including the application and grant year. Each patent record also provides information about the invention itself, including—important for our research—its technology classifications based on the International Patent Classification (IPC) system and the citations it makes to prior inventions.

C.2 Global Patent Data

Data Source To capture global innovation, we use global patent data collected from Google Patents. The data set contains information on more than 36 million patents from the more than 40 main patent authorities around the world, over the period 1976–2020, including the USPTO, the European Patent Office (EPO), the Japanese Patent Office (JPO), and the Chinese National Intellectual Property Administration, among others. For each patent, Google Patents provides similar information as in the USPTO data described above.

Google Patents data are obtained from the DOCDB (EPO worldwide bibliographic data), the same underlying source as the more widely used PATSTAT data. We choose to use Google Patents as our main global innovation data source because it is public and accessible to all researchers free of charge. In Appendix D, we discuss specific differences between Google Patents and PATSTAT data. We show that these databases have only minor differences in their coverages and definitions of key variables and that all our empirical results are robust to both.

Identifying Patenting Locations Filing a patent in a country or patent office does not necessarily mean the underlying invention is created in the same geographic unit (e.g., Chinese firms file USPTO patents, Korean firms file patents with the Chinese National Intellectual Property Administration). These "global patenting" activities pose two important challenges for our empirical analysis. First, we need to properly determine the geographical location of the innovating activities. We assign each patent to a geographical unit according to the country of residence of its

[15] We obtain the patent data from the USPTO PatentsView platform, accessible at https://www.patentsview.org/download/.

inventor(s). When a given patent is associated with multiple inventors from different countries or territories, we assign these inventors equal weight (e.g., N inventors each obtaining $1/N$ credit). If this information is not available (as in 31% of the global patent sample),[16] we use the country of the assignee(s) instead. For 8% of patents with no easily accessible geographic location data, we assign the country of the patent office.

Identifying a Unique Invention Behind Multiple and Multinational Patents The second challenge is to de-duplicate multiple patents filed with different patent authorities for the same underlying invention. This is common practice for IP protection reasons, but may lead to double counting. To overcome this challenge, we use patent family information. We assign a set of patents to the same family if they have: (1) the same application number; or (2) the same PCT number; or (3) the same Google-provided patent family ID; or (4) at least one priority application number in common. Using patent family information, we can make sure a single invention is not counted more than once even when multiple patents are filed based on it. We also can use the earliest filing date to properly identify the timing of the underlying invention.

Cross-country Citations Importantly, patent citation information is global too—that is, we observe citations made by a patent filed by a U.S. firm with the USPTO to a patent owned by a German firm filed at the EPO. This allows us to track the innovation network at the global scale. In our sample, the proportion of citations a patent makes to foreign patents is 38%, and this number has been growing over the years.

C.3 Connecting Patent Data with Sectoral Data

Patent data are classified into International Patent Classification (IPC) classes based on the technological content of the invention. The IPC system provides a uniform and hierarchical system of language-independent symbols for the classification of patents and utility model according to the different areas of technology to which they pertain. The IPC classification system does not naturally map to the sector classifications in either the WIOD data nor the BLS data on sectoral output and linkages. Specifically, each sector could patent in multiple IPC classes, while many sectors could patent in each single IPC class. Patent data need to be mapped to sectoral data (on value-added, R&D expenditures, employment, intermediate inputs, etc.) for our empirical analysis in different sections of our paper. This includes: (1) constructing sectoral measures of innovation activities, and (2) projecting sectoral measures into technology class levels.

[16] Patent observations with only the country of the patent office as geographic location are mainly historical U.S. patents (51%) and historical patents originating from France, Germany, and the Soviet Union (each accounting for about 10%).

Measuring Innovation at the Sector Level To construct innovation output for each country-sector-year and the country-sector-pair-wise innovation network, we need to map innovation activities to industrial sectors. We rely on our ability to observe innovation activities at the level of firms, for which we observe their industry classifications. Starting with U.S. domestic data—we link the USPTO patent database to Compustat using the bridge file provided by the NBER (up to the year 2006) and KPSS's data repository.[17] For later years, we complete the link using a fuzzy matching method based on company name, basic identity information, and innovation profiles, similar to Ma (2020) and Ma (2021). Firms' sectoral classifications are defined by North American Industry Classification System (NAICS) codes, which are then mapped to BLS sectors using the crosswalk file provided by the BLS website.[18] For each sector, we can aggregate all innovation activities including patent numbers, citation-adjustment patent counts, and total R&D expenditures, conducted by U.S. firms in that specific sector..

The connection between international patent and sectoral data implements a similar logic but uses more complicated data collection and matching processes. We assemble information on global firms from Worldscope and Datastream databases accessed through Wharton Research Data Services (WRDS). The raw data sets cover more than 109,000 global firms located in 160 countries all over the world. The process is similar to that described above for U.S. data. The standard industry classifications in these databases are based on the International Standard Industrial Classification (ISIC), and can therefore be accurately mapped to the WIOD, which is also organized using the ISIC system.

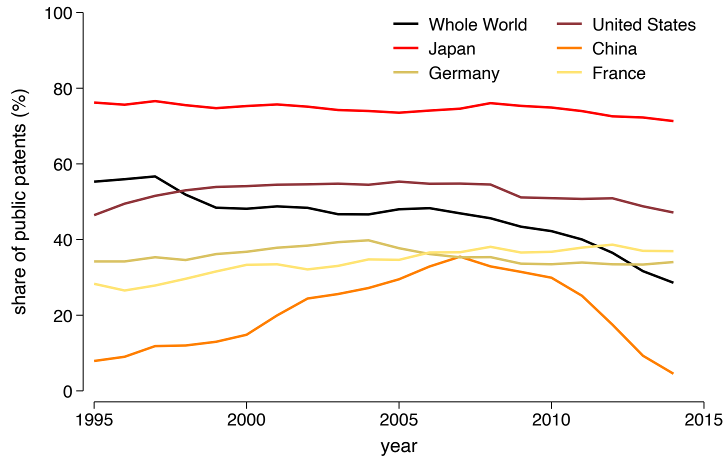
The benefit of using information on firms to accurately link innovation to industrial sectors warrants the question of how representative those firms' innovation are. We find that firms in our data set produce about half of all patents in each country—for example, our sample of firms covers 44% of patents in the U.S., and 65% in Japan, two countries with the largest number of patents. Figure A.1 shows the time trend of patent shares from firms covered in our databases in the whole world and in different countries. The similarity of industry distribution between patents from covered firms and all patents in the USPTO is 0.97 when we compare the share of patents in each of the 131 3-digit IPCs for all patents and for patents from firms covered in our firm-level databases.

Projecting Sectoral Measures to Technology Classes When the unit of analysis is an IPC class (in a certain country-year), the key challenge is to project sectoral measures, such as value-

[17] The extended data for KPSS can be accessed at https://github.com/KPSS2017/Technological-Innovation-Resource-Allocation-and-Growth-Extended-Data.

[18] Accessed at https://www.bls.gov/ces/naics/.

Figure A.1. Comovements of Public Patent Sample and Whole Sample



Notes. This table documents the time trend of the patent shares for firms covered in our firm-level databases across the world and in different countries.

added, to technology classes. We use the sector-IPC mapping provided in Lybbert and Zolas (2014). Using this mapping, we decompose each sectoral measure with proper weights to relevant IPC classes, and then aggregate the measures into the IPC level.

C.4 Constructing Cross-Sector R&D Allocation Data

Our quantitative analysis uses data on R&D allocation across different technology classes in each country. There is no standard database to exhaustively measure such information. Our primary measure relies on aggregating firm-level R&D expenditures to the country-sector-year level, based on three widely used firm-level data sets: Compustat, Worldscope, and Datastream. Combined, these data cover more than 100,000 global firms located in 160 countries and account for over 95% of the world's total market capitalization. For multinationals, we first attribute the firm-level R&D expenditures to IPC-country level in proportion to each firm's shares of patents in each IPC-country, following Griffith, Harrison, and Van Reenen (2006), and then aggregate to IPC-country-year level.

This primary measure of sectoral R&D has the advantage of covering more country-years compared to alternative approaches such as the OECD ANBERD Database. It also allows us to attribute R&D expenditures of multi-sector and multinational firms more explicitly and in a more transparent fashion. However, the primary measure of sectoral R&D is imperfect, as the firm-level data sets oversample large firms and have potentially different reporting standard across countries; we also miss R&D inputs from public sectors. Nevertheless, it is important to note that, as our theory concerns the cross-sector R&D allocation, what matters for our quantitative analysis later is the allocation shares of R&D resources across sectors in each country and not the aggregate R&D levels; any mismeasurement that affects all sectors proportionally should have no quantitative impacts.

As robustness checks, we show that our primary measure of R&D allocation shares correlates strongly with two independent sources of R&D data, thereby giving us confidence in using our measure for quantitative analysis. We first provide a robustness check using the OECD Analytical Business Enterprise Research and Development (ANBERD) Database (Machin and Van Reenen, 1998), which has country-sector-level R&D information. Relative to our primary R&D measure, the ANBERD Database has more limited country-year coverage and relies more on imputations from firm-level surveys. Our primary R&D measure also allows us to explicitly and transparently attribute R&D of multi-sector or multinational enterprises to different sectors and countries.

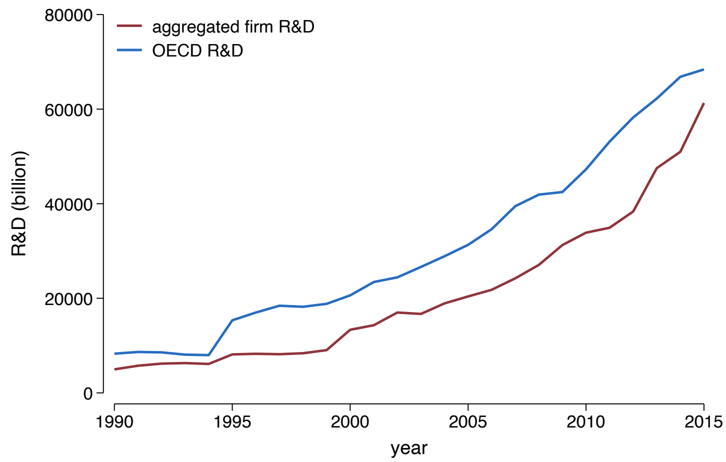
For all the major economies in both data sets, R&D allocation from ANBERD is highly correlated with our primary measure. In the subsample of country-year observations covered in both data sources, we show that R&D expenses calculated from our firm-level data represents a significant proportion of R&D estimated by the ANBERD data, and they follow a very similar aggregate trend (Figure A.2).

The second robustness check calculates the cross-sector R&D allocation using the innovation output (which is better measured) rather than input: the number of patents produced in each country-IPC (or country-sector) divided by total number of patents produced in that specific country.

Table A.1 shows the correlation among R&D allocation measures used in our empirical analysis—R&D expense shares using R&D expenditures aggregated from firm-level data; R&D expenditures surveyed and imputed in the OECD ANBERD database; and patent shares. The correlations are calculated using 20 top patenting countries in 2010 and their R&D allocation measures across 3-digit IPC categories. The top panel first aggregate sectoral R&D expenditures across all countries and then calculate correlation of the sectoral R&D shares. The bottom panel calculate a country-specific sectoral R&D allocation correlations and then average the correlations across different countries. In each panel, the bottom half of the table shows the Pearson correlations; the top half of the table shows Spearman's rank correlation, which is equal to the Pearson correlation of the rank values.

These three proxies for R&D allocations are highly correlated. For example, in Panel A, the correlation between R&D allocations aggregated from firm-level data and from the OECD scores above 0.9. The correlation between input shares and the patent output shares is slightly lower, but still above 0.8. The high correlations among these three measures of R&D allocation shares translate into the robustness of our quantitative results, as illustrated in Section E.4 of the Online Appendix.

Figure A.2. Comovements of Public Patent Sample and Whole Sample



Notes. This table documents the time trend of total R&D expenditures calculated from aggregating firm-level R&D from Compustat, Worldscope, and Datastream and those calculated from aggregating country-sector information from OECD ANBERD data. For each year, we cover countries that are covered in both databases.

Table A.1. Different Measures of Cross-Sector R&D Allocation Are Highly Correlated

Panel A	Share of Aggregated Firm R&D	Share of Patents	Share of OECD R&D
Share of Aggregated Firm R&D		0.83	0.97
Share of Patents	0.86		0.82
Share of OECD R&D	0.93	0.78	

Panel B	Share of Aggregated Firm R&D	Share of Patents	Share of OECD R&D
Share of Aggregated Firm R&D		0.74	0.91
Share of Patents	0.74		0.76
Share of OECD R&D	0.74	0.69	

Notes: This table shows the correlation of R&D allocation measures used in our empirical analysis–R&D expense shares using R&D expenditures aggregated from firm-level data; R&D expenditures surveyed and imputed in the OECD ANBERD database; and patent shares. The correlations are calculated using 20 top patenting countries in 2010 and their R&D allocation measures across 3-digit IPC categories. The top panel first aggregate sectoral R&D expenditures across all countries and then calculate correlation of the sectoral R&D shares. The bottom panel calculate a country-specific sectoral R&D allocation correlations and then average the correlations across different countries. In each panel, the bottom half of the table shows the Pearson correlations; the top half of the table shows Spearman's rank correlation, which is equal to the Pearson correlation of the rank values.

D Cross-checking Google Patents with PATSTAT

This appendix compares data from Google Patents (accessible to all researchers free of charge) and the widely used commercial database PATSTAT. These exercises will compare their data coverage, key variable definitions, and the robustness of empirical analyses in those two databases.

D.1 Basic Data Structure and Coverage

Google Patents and PATSTAT share nearly identical data structure. Both databases have three levels of innovation units: publication, application, and family.

- **Application:** The central unit is an innovation application, which is a request filed to a patent office for patent protection for an invention (which may or may not be granted later).
- **Publication (most basic unit):** After an application is filed, various publications could be issued.[19] These publications can be disclosed patent filings (often 18 months after the initial filing date), granted patent specification, corrections, etc. In simple terms, publications help identify key events over an application's life cycle. The basic units of both Google Patents and PATSTAT are innovation "publications."
- **Family:**[20] Applications that cover the same underlying invention are grouped into families. This often happens when the same invention is filed with multiple patent offices, sometimes simultaneously, for protections in different countries. All applications (and publications tracking their life cycle events) in the same family thus have the same priorities, and their technical content is often regarded as identical or almost identical. Patent family counting allows us to track unique inventions across different economies.

Figure A.3 presents the sample coverage of publications, the most basic units, for both Google Patents and PATSTAT in the time series. The coverages of the two data sets are virtually identical.

D.2 Identifying Granted Patents

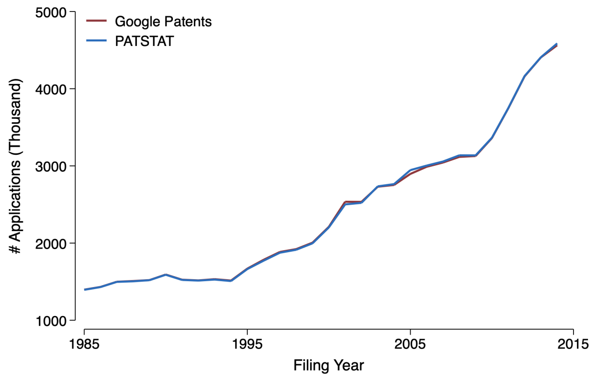
Publications represent the most comprehensive set of innovation-related documents, yet many of them are irrelevant for studying innovation—some publications are associated with denied applications, some are design patents unrelated to scientific or technological progress, etc. As a result, it is useful to identify granted patents related to new technologies (e.g., utility patents

[19] In cases that generate no publications (i.e., the invention is treated with absolute confidentiality), the invention would not be accessible in any database.

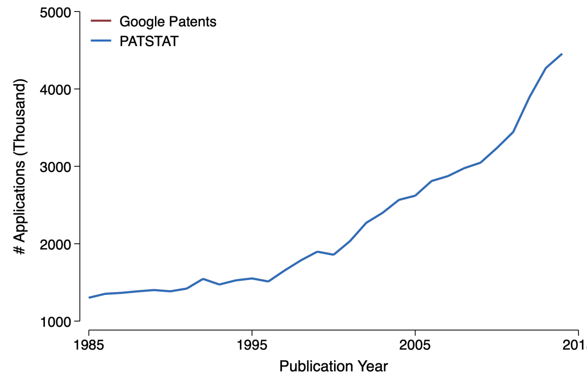
[20] In our paper, we consider the more widely accepted definition of simple family, also called the DOCDB family or Espacenet patent family.

Figure A.3. Google Patents v.s. PATSTAT by Year

(a) Applications by Filing Year



(b) Applications by Publication Year



in the USPTO system). The two database handle this process largely identically, yielding very comparable patent sets. However, there are three noticeable differences:

1. Identifying whether a patent is granted mainly relies on the kind code of the patent, which is defined by the patent office and will change with the reform of the patent system of the patent office.[21] For example, the kind code of patent "US-10001017-B2" is "B2." The rules used to identify granted patents differ somewhat in Google Patents vs. PATSTAT.

2. Because PATSTAT uses additional legal event data to identify granted patents, patents granted by some small patent offices can be identified.

3. Other minor differences include missing filing dates or issue dates.

Table A.2 shows the comparison of granted patents between Google Patents and PATSTAT and list the sources of coverage differences.

[21]For the detailed meaning of difference kind codes in different patent offices, we refer readers to the document of format concordance of publication numbers in EPO (see https://www.epo.org/searching-for-patents/data/coverage/regular.html).

Table A.2. Difference of Granted Patents Between Google Patents and PATSTAT

Panel (A): For Granted Patents in PATSTAT

	# Patents			
Patents granted in 1985—2014	19,923,292	100.00%		
Overlapped with Google Patents	17,135,611	86.01%		
Non-overlapped with Google Patents	2,787,681	13.99%	100.00%	
1. Additional patent office data from legal event data	1,456,242		52.24%	100.00%
(1) For patent office ZA	175,317			12.04%
(2) For patent office MX	125,298			8.60%
(3) For patent office PL	125,246			8.60%
(4) For patent office UA	95,956			6.59%
(5) For patent office PT	82,533			5.67%
(6) For patent office DD	79,171			5.44%
(7) For patent office NO	65,312			4.48%
(8) For patent office BR	62,447			4.29%
(9) For patent office HU	61,707			4.24%
(10) For patent office IL	57,165			3.93%
(11) Other patent offices including BG, BY, CH, CO, CS, CU, CZ, EA, EE, GE, GR, HK, HR, ID, IE, IN, IS, KE, LT, LV, MA, MC, MD, ME, MN, MT, MY, NI, OA, PE, PH, RO, RS, SA, SE, SG, SI, SK, SM, SV, TJ, TR, UY, VN, YU, ZW	526,090			36.13%
2. Additional rules used to identify granted patents	1,331,439		47.76%	100.00%
(1) For patent office AT, patents with kind code in [T]	543,805			40.84%
(2) For patent office DE, patents with kind code in [T2]	468,202			35.17%
(3) For patent office KR, patents with kind code in [A]	65,237			4.90%
(4) For patent office DK, patents with kind code in [T3]	58,520			4.40%
(5) For patent office ES, patents with kind code in [A1, A6]	47,354			3.56%
(6) For patent office AU, patents with kind code in [A1, A8]	32,835			2.47%
(7) For patent office FI, patents with kind code in [C]	31,907			2.40%
(8) For patent office CN, patents with kind code in [A]	28,928			2.17%
(9) For patent office AR, patents with kind code in [A1]	24,865			1.87%
(10) For patent office US, patents with kind code in [E]	16,366			1.23%
(11) Other patent offices	13,420			1.01%

Panel (B): For Granted Patents in Google Patents

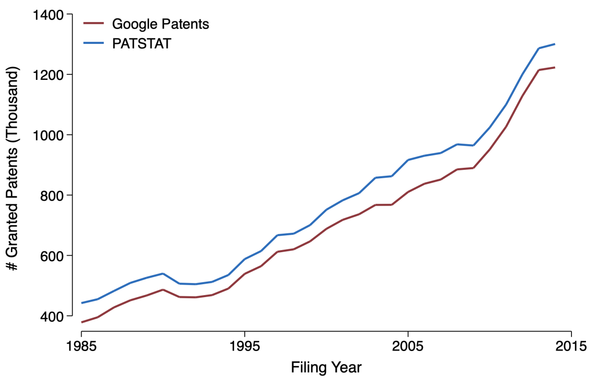
	# Patents			
Patents granted in 1985—2014	18,144,529	100.00%		
Overlapped with PATSTAT	17,135,612	94.44%		
Non-overlapped with PATSTAT	1,008,917	5.56%	100.00%	
1. Additional patent office data from legal event data	0		0.00%	
2. Additional rules used to identify granted patents	1,008,917		100.00%	100.00%
(1) For patent office DE, patents with kind code in [D1]	883,482			87.57%
(2) For patent office DK, patents with kind code in [T3]	58,118			5.76%
(3) For patent office FI, patents with kind code in [B]	31,585			3.13%
(4) For patent office BE, patents with kind code in [A3, A4, A5, A6, A7]	20,797			2.06%
(5) For patent office KR, patents with kind code in [B1]	6,546			0.65%
(6) For patent office ES, patents with kind code in [B1]	2,399			0.24%
(7) For patent office DZ, patents with kind code in [A1]	1,755			0.17%
(8) For patent office AU, patents with kind code in [B2]	1,458			0.14%
(9) For patent office EP, patents with kind code in [B1]	1,344			0.13%
(10) For patent office SU, patents with kind code in [A1]	932			0.09%
(11) Other patent offices	501			0.05%

Notes. This table compares coverages of granted patents between Google Patents and PATSTAT and the reasons for discrepancies.

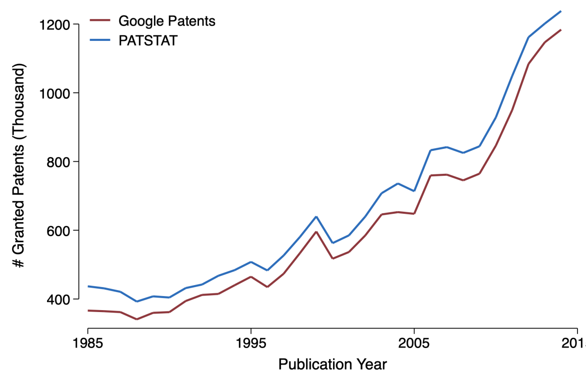
Despite those differences, Google Patents and PATSTAT agree on roughly 95% of the identified granted patents. In Figure A.4, we present the numbers of granted patents in Google Patents and PATSTAT. We also show this difference across various patent offices and countries of origin.

Figure A.4. Google Patents v.s. PATSTAT Coverage

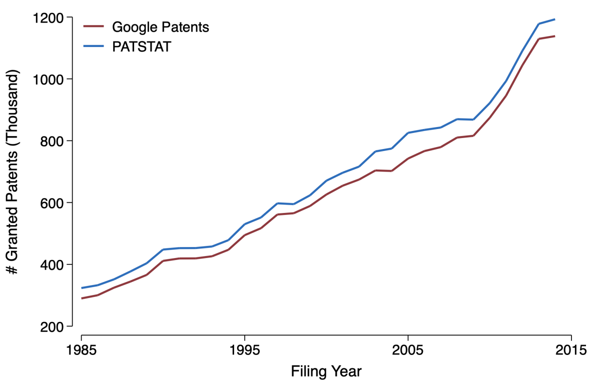
(a) Granted Patents by Filing Year



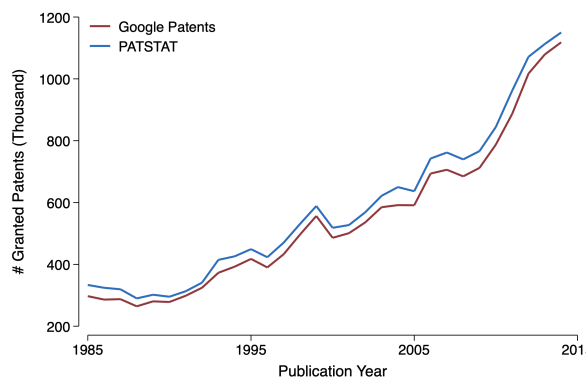
(b) Granted Patents by Publication Year



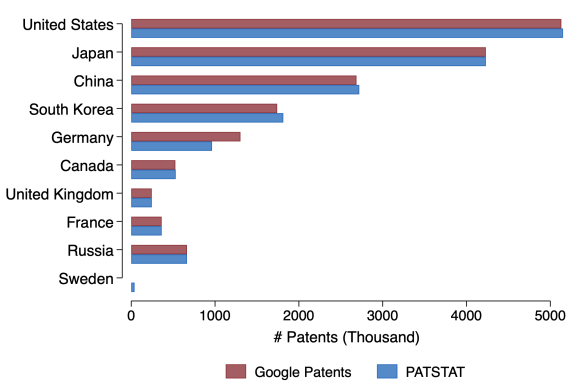
(c) Granted Patents by Filing Year (43 WIOD Countries)



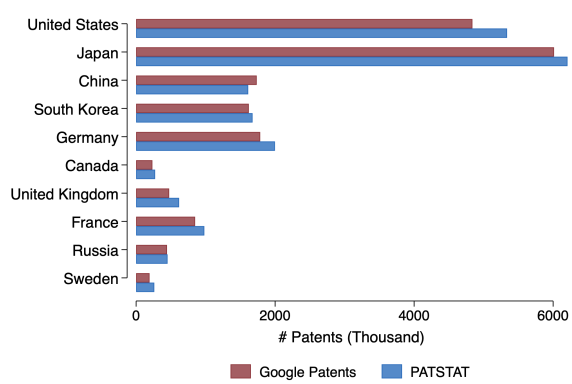
(d) Granted Patents by Publication Year (43 WIOD Countries)



(e) Google Patents vs. PATSTAT By Patent Offices



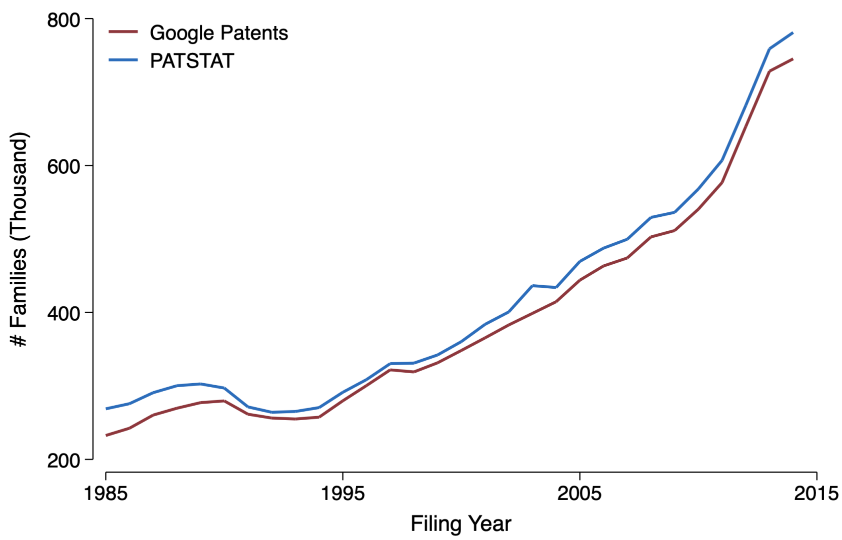
(f) Google Patents vs. PATSTAT By Invention Origin



D.3 Patent Family

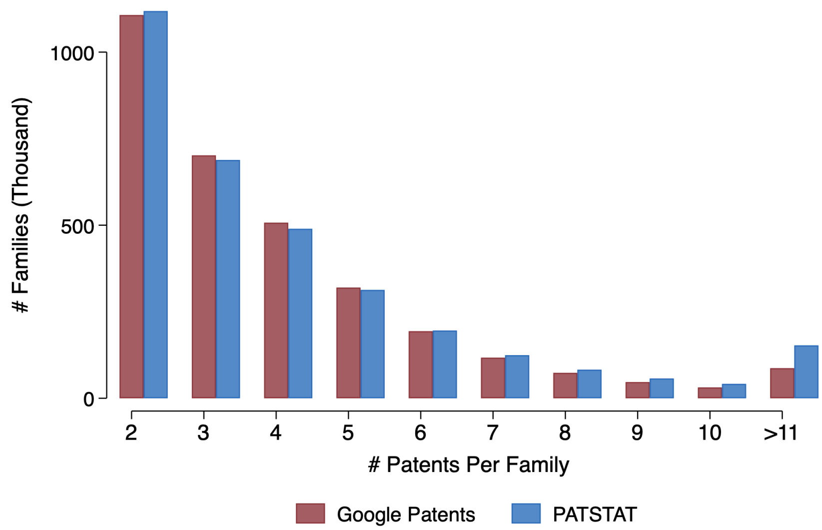
Defining patent family involves the use of information regarding priority dates and priority patents in the global patent database, among others. Figure A.5 presents the number of patent families identified in both data sets. They are very comparable to each other, and the minor gap can be explained by the differences in the number of identified patents described in the previous section.

Figure A.5. Google Patents v.s. PATSTAT: Patent Families by Year



To further check this consistency, in Figure A.6 we show the distribution of the number of patents in each family in Google Patents and PATSTAT, which again are quite comparable. In Google Patents, there are 11,693,980 patent families between 1985 and 2014. Among these families, 3,184,884 contain at least two patents, and on average, these families contain 3.99 patents. In PATSTAT, there are 12,344,446 patent families between 1985 and 2014. Among those families, 3,263,376 of them contain at least two patents, and on average, these families contain 4.34 patents.

Figure A.6. Google Patents v.s. PATSTAT: Distribution of Number of Patents in Each Family



We next perform a family-to-family comparison between the two databases. First, we focus on families that only contain one patent: 98.74% of these families in Google Patents are consistent with that in PATSTAT, and 97.79% of those families in PATSTAT are consistent with those in Google Patents. For patent families with two patents, the share of patents in PATSTAT that are consistent with Google Patents is 94.11%; the share of patents in Google Patents that is consistent with PATSTAT is 94.38%. Overall, patent families seem to be consistently defined across the databases at a very high rate.

D.4 Robustness of Results Using Google Patents and PATSTAT

In this section, we present results from using PATSTAT patent data as the base for innovation measurement and innovation network construction. The overall takeaway is that the results using PATSTAT are virtually identical to results using Google Patents.

D.4.1 Innovation Network

Results in this subsection show that innovation networks constructed using PATSTAT and Google Patents are highly correlated (Table A.3), and they have virtually identical properties such as centrality (Figure A.7) and visualizations (Figure A.8).

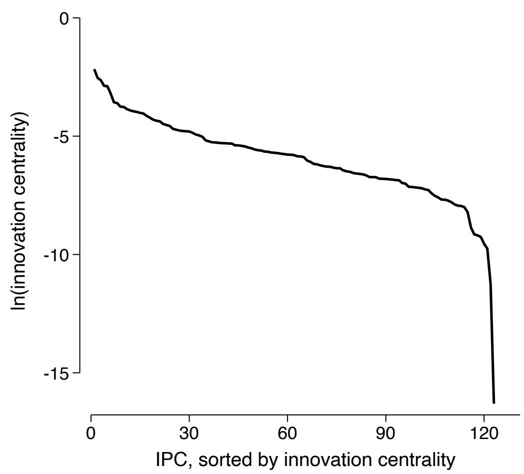
Table A.3. Correlations of Between the Innovation Network from Google Patents and PATSTAT

All	U.S.	Japan	China	Korea	Germany	Canada	UK	France	Russia	Sweden
0.997	0.998	0.945	0.987	0.975	0.979	0.986	0.989	0.966	0.887	0.934

Notes. This is the correlation between the innovation networks calculated using Google Patents and PATSTAT data.

Figure A.7. Innovation Centrality and Key Sectors for PASTAT

(a) Innovation Centrality Across IPCs



(b) Top Ten IPCs by Innovation Centrality a_i

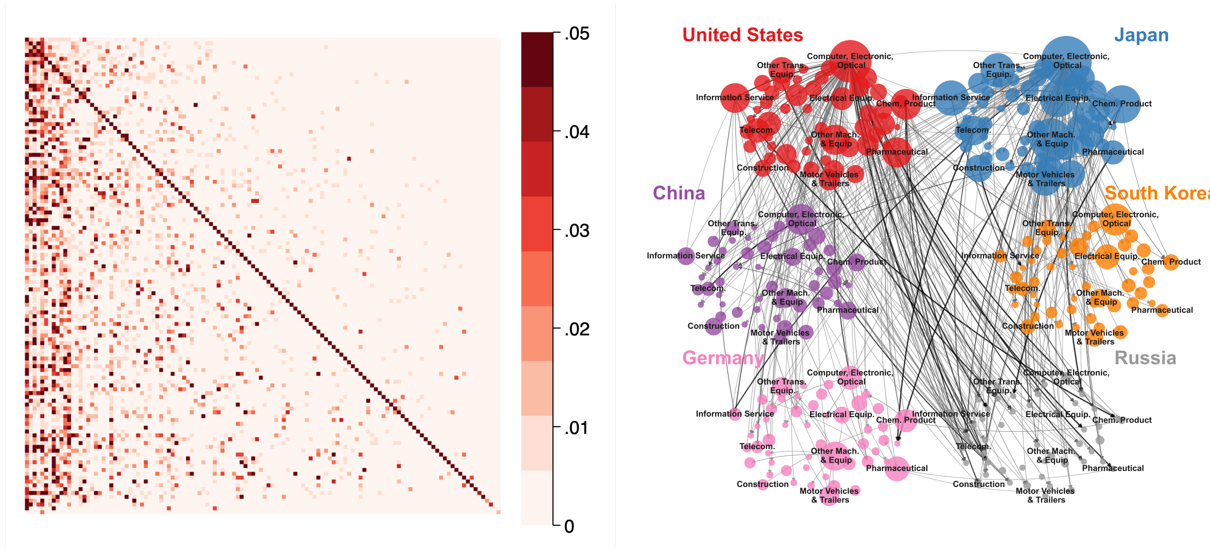
1	A61	medical or veterinary science; hygiene
2	G06	computing; calculating or counting
3	H01	basic electric elements
4	G01	measuring; testing
5	H04	electric communication technique
6	B60	vehicles in general
7	G02	optics
8	B01	physical or chemical processes or apparatus in general
9	C08	organic macromolecular compounds; their preparation or chemical working-up; compositions based thereon
10	F16	engineering elements or units; general measures for producing and maintaining effective functioning of machines or installations; thermal insulation in general

Notes. This figure reproduces Figure 2 in the paper using PATSTAT data. This figure presents the innovation centrality of different technology classes categorized using IPCs. Panel (a) plots $\log(a_i)$, and the sectors are ranked in descending order based on a_i. Panel (b) lists the top ten IPCs by their innovation centrality.

Figure A.8. Visualizing the Innovation Network for PATSTAT

(a) IPC-to-IPC (131×131) Network Ω

(b) Global Innovation Network Across Country-Sectors



Notes. This figure reproduces Figure 1 in the paper using PATSTAT data. The left panel visualizes the IPC-to-IPC network $\boldsymbol{\Omega}$ as a heatmap, with darker colors representing larger matrix entries; sectors are ordered according to their innovation centrality. The right panel visualizes the global innovation network. Each node is a country-sector, with size drawn in proportion to patent output. Arrows represent knowledge flows, with width drawn in proportion to citation shares.

D.4.2 Knowledge Spillovers

This subsection reproduces results to confirm the mechanism of sectoral innovation activities being influenced by innovation from global upstream sectors.

Table A.4. Evidence of the Global Innovation Network for Knowledge Spillovers
Based on WIOD - PATSTAT

$Y=$	ln(Patents)			ln(Cites)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Knowledge^{Up}_{mit}$	0.181***	0.193***	0.174***	0.285***	0.325***	0.275***
	(0.053)	(0.056)	(0.054)	(0.080)	(0.082)	(0.081)
$ln(R\&D)_{mi,t-1}$	0.031***	0.031***	0.030***	0.036***	0.038***	0.036**
	(0.010)	(0.010)	(0.010)	(0.014)	(0.014)	(0.014)
$Knowledge^{Down}_{mit}$		-0.035			-0.113***	
		(0.030)			(0.039)	
$Knowledge^{Up,IO}_{mit}$			0.054			-0.036
			(0.067)			(0.071)
R^2	0.968	0.968	0.969	0.943	0.943	0.944
No. of Country x Sectors	564	564	550	564	564	550
No. of Obs	10549	10549	10315	10549	10549	10315
Fixed Effects	Country x Sector, Country x Year, Sector x Year					

Notes. This table reproduces Table 3 in the paper using PATSTAT data. This table tests the relation between innovation in a focal sector and past innovation in connected sectors through the innovation network, in an international setting. We restrict the sample to country-sectors that have at least ten patents over the full sample period. To measure innovation production (Y), we use the number of patents and total number of citations. The key variable of interest, $\text{Knowledge}^{Up}_{it}$, is the knowledge from upstream, defined in (28). Fixed effects at the country-sector, country-year, and sector-year levels are included as controls. Columns (2) and (5) include downstream knowledge as a control. Columns (3) and (6) include knowledge accumulated from upstream sectors in the production network as a control. Standard errors in parentheses are clustered at the country-sector level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels respectively.

E Supplementary Results

In this section, we provide additional empirical results.

E.1 Innovation Networks Are Stable Over Time and Across Countries

We first document that innovation networks are stable over time and across innovative countries. We construct time-varying measures of the innovation network, following the formula in (24) but using citations made by patents filed during specific time periods, from all countries in our sample. For the innovation network time-stamped at t, we use new patents and their citations between $t-10$ and $t-1$ to construct the network. Table A.5 shows the correlations between our baseline, time-invariant measure ω_{ij} of the innovation network and these other measures ω_{ijt} constructed using patents filed in specific years t. The bottom half of the table shows the Pearson correlations; the top half of the table shows Spearman's rank correlation, which is equal to the Pearson correlation of the rank values and can be more revealing of network similarities than the Pearson correlation of values (Liu, 2019). Table A.5 shows that the innovation network is highly stable over time; the time-varying measures exhibit above 0.8 correlations even when measured using citation data that are three decades apart, and all year-specific measures are strongly correlated with our time-invariant baseline measure.

Table A.5. The Innovation Network Is Highly Correlated Over Time

Time Period	All years	2020	2010	2000	1990	1980
All years		0.98	0.98	0.97	0.90	0.89
2020	0.95		0.97	0.93	0.86	0.85
2010	0.96	0.97		0.96	0.88	0.87
2000	0.93	0.92	0.96		0.92	0.90
1990	0.90	0.80	0.84	0.90		0.91
1980	0.81	0.77	0.81	0.87	0.89	

Notes: This table shows the correlation of innovation networks calculated using different vintages of patent data. For each decade, all global patents in that decade are included when constructing the innovation network. The bottom half of the table shows the Pearson correlations; the top half of the table shows Spearman's rank correlation, which is equal to the Pearson correlation of the rank values.

Second, we construct country-specific innovation networks. Specifically, we use the same formula (24) but restrict the sample to all patents from each country. Table A.6 shows the correlations between our baseline, location-invariant measure and the country-specific measures for the ten countries with the most patents in our sample; Pearson correlations are again shown in the bottom half of the table whereas Spearman's rank correlations are shown in the top half. Innovation networks are highly stable across countries. In particular, our baseline measure, which is constructed using patents pooled from around the world, has a correlation coefficient of above 0.98 with the network implied by U.S. patents and is also highly correlated (>0.8 rank correlation) with the innovation networks in Japan, China, Germany, Canada, the U.K., and France. The only exception is Russia, whose innovation network is less perfectly correlated with the measures, but the correlation is still substantial (about 0.6).

Table A.6. The Innovation Network Is Highly Correlated Across Countries

Countries	All	US	Japan	China	South Korea	Germany	Russia	France	UK	Canada	Netherlands
All		0.98	0.87	0.87	0.84	0.89	0.63	0.86	0.92	0.88	0.81
US	0.95		0.84	0.86	0.82	0.88	0.64	0.85	0.92	0.88	0.80
Japan	0.86	0.83		0.88	0.89	0.85	0.63	0.87	0.86	0.84	0.83
China	0.85	0.86	0.87		0.88	0.85	0.66	0.85	0.87	0.86	0.82
South Korea	0.78	0.77	0.83	0.84		0.84	0.64	0.84	0.85	0.82	0.84
Germany	0.85	0.87	0.81	0.80	0.72		0.64	0.83	0.87	0.83	0.81
Russia	0.62	0.63	0.62	0.62	0.55	0.61		0.65	0.64	0.64	0.66
France	0.91	0.86	0.79	0.77	0.72	0.82	0.57		0.86	0.85	0.83
UK	0.87	0.89	0.85	0.85	0.80	0.86	0.64	0.80		0.88	0.82
Canada	0.86	0.88	0.79	0.81	0.71	0.81	0.59	0.80	0.81		0.81
Netherlands	0.84	0.85	0.79	0.82	0.75	0.79	0.58	0.78	0.79	0.81	

Notes: This table shows the correlation of innovation networks calculated using patents in the top ten innovative countries ranked by patent outputs between 2010—2014. When calculating this country-specific innovation network, all patents of the country across all years are included. The bottom half of the table shows the Pearson correlations; the top half of the table shows Spearman's rank correlations, which are equal to the Pearson correlation of the rank values.

E.2 Knowledge Spillovers Through Innovation Networks—Robustness

This subsection provides additional robustness analyses on innovation diffusion through innovation networks, echoing Section 4.2 in the paper. The main results supporting the important role of innovation networks in knowledge spillovers are provided in Tables 2 and 3 in the paper. Below, we present tests to show the robustness of these results. Specifically, these analyses incorporate changing U.S. BLS Sectors to IPC (International Patent Classification) classes as the node in innovation networks (Table A.7), additional measures of innovation output (Table A.11), and different time horizons to calculate upstream innovation (Tables A.9 and A.12). Finally, we revisit the dynamic prediction of our key law of motion (25), that upstream knowledge from the more distant past has less effect on patent output, in Figure (A.9). The figure shows an obsolescence-like pattern (Ma, 2021) in which past upstream knowledge's effect on subsequent innovation weakens over time, precisely as our theory predicts.

Table A.7. U.S. and Global Evidence of Knowledge Spillover Through Innovation Networks
Based on IPC

$Y=$	US		Global	
	ln(Patents)	ln(Cites)	ln(Patents)	ln(Cites)
	(1)	(2)	(3)	(4)
$Knowledge^{Up}_{it}$	0.519***	0.548***	0.050***	0.075***
	(0.083)	(0.106)	(0.011)	(0.015)
$ln(R\&D)_{i,t-1}$	0.298***	0.254**	0.006*	0.004
	(0.075)	(0.113)	(0.003)	(0.005)
$Knowledge^{Down}_{it}$	-0.243***	-0.348***	-0.032***	-0.025**
	(0.064)	(0.090)	(0.008)	(0.011)
R^2	0.959	0.947	0.947	0.905
No. of Sectors	431	431		
No. of Country x Sectors			4,560	4,560
No. of Obs	8,620	8,620	8,2977	8,2977
Fixed Effects	Sector, Year		Country x Sector Country x Year Sector x Year	

Notes. This table reproduces Tables 2 and 3 in the paper. The key difference is this table uses the country by detailed 4-digit IPC (international patent classification) class as the unit of nodes instead of country by (BLS or WIOD) industrial sectors.

Table A.8. U.S. Evidence of Knowledge Spillover Through Innovation Networks
Adding The Impact of Own Sector

$Y=$	ln(Patents)			ln(Cites)			ln(Patent Value)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$Knowledge^{Up}_{it}$	0.742***	0.789***	0.675***	0.914***	0.814***	0.860***	1.005***	1.036**	0.961***
	(0.130)	(0.170)	(0.138)	(0.193)	(0.279)	(0.203)	(0.309)	(0.395)	(0.305)
$Knowledge^{Own}_{it}$	0.656***	0.648***	0.583***	0.554***	0.571***	0.491***	0.294***	0.289***	0.230
	(0.050)	(0.050)	(0.083)	(0.077)	(0.073)	(0.124)	(0.105)	(0.105)	(0.201)
$ln(R\&D)_{i,t-1}$	0.142***	0.140***	0.139***	0.139*	0.143*	0.135*	0.248**	0.247**	0.241**
	(0.036)	(0.036)	(0.037)	(0.077)	(0.078)	(0.079)	(0.095)	(0.095)	(0.095)
$Knowledge^{Up,IO}_{it}$			0.006			-0.019			-0.094
			(0.125)			(0.183)			(0.225)
$Knowledge^{Own,IO}_{it}$			0.094			0.085			0.097
			(0.079)			(0.107)			(0.200)
$Knowledge^{Down}_{it}$		-0.085			0.180			-0.056	
		(0.142)			(0.278)			(0.299)	
R^2	0.938	0.939	0.939	0.912	0.912	0.912	0.888	0.888	0.888
No. of Sectors	93	93	93	93	93	93	93	93	93
No. of Obs	1,840	1,840	1,840	1,840	1,840	1,840	1,840	1,840	1,840
Fixed Effects		Sector, Year			Sector, Year			Sector, Year	

Notes. This table reproduces Table 2 in the paper by incorporating patenting activities from past innovation from own sector.

Table A.9. U.S. Evidence of Knowledge Spillover Through Innovation Networks
Different Knowledge Periods

Panel (a): $\tau = 5$

$Y=$	ln(Patents)			ln(Cites)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Knowledge_{it}^{Up,\tau=5}$	0.454***	0.471***	0.392***	0.699***	0.730***	0.653***
	(0.152)	(0.176)	(0.149)	(0.163)	(0.173)	(0.164)
$ln(R\&D)_{i,t-1}$	0.288***	0.288***	0.291***	0.273***	0.273***	0.275***
	(0.062)	(0.061)	(0.058)	(0.084)	(0.084)	(0.084)
$Knowledge_{it}^{Down,\tau=5}$		-0.038			-0.068	
		(0.157)			(0.095)	
$Knowledge_{it}^{Up,IO}$			0.392**			0.296
			(0.172)			(0.205)
R^2	0.914	0.914	0.916	0.901	0.901	0.902
No. of Sectors	94	94	94	94	94	94
No. of Obs	1,847	1,847	1,847	1,847	1,847	1847
Fixed Effects		Sector, Year			Sector, Year	

Panel (b): $\tau = 20$

$Y=$	ln(Patents)			ln(Cites)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Knowledge_{it}^{Up,\tau=20}$	0.697***	0.714***	0.611***	0.871***	0.903***	0.808***
	(0.183)	(0.206)	(0.179)	(0.209)	(0.223)	(0.200)
$ln(R\&D)_{i,t-1}$	0.270***	0.269***	0.274***	0.254***	0.253***	0.258***
	(0.062)	(0.061)	(0.059)	(0.086)	(0.085)	(0.086)
$Knowledge_{it}^{Down,\tau=20}$		-0.038			-0.071	
		(0.164)			(0.101)	
$Knowledge_{it}^{Up,IO}$			0.338*			0.249
			(0.173)			(0.204)
R^2	0.916	0.916	0.917	0.901	0.901	0.902
No. of Sectors	94	94	94	94	94	94
No. of Obs	1,847	1,847	1,847	1,847	1,847	1847
Fixed Effects		Sector, Year			Sector, Year	

Notes. This table reproduces Table 2 in the paper. The key difference is using different τ periods to calculate knowledge accumulated through the innovation network. Table 2 uses $\tau = 10$, while this table uses alternative values of $\tau = 5$ and $\tau = 10$.

Table A.10. U.S. Evidence of Knowledge Spillover Through Innovation Networks
Exponential Knowledge Discounting

$Y=$	ln(Patents)			ln(Cites)			ln(Patent Value)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$Knowledge_{it}^{Up,depreciation}$	0.539***	0.556***	0.466***	0.779***	0.814***	0.725***	0.935***	0.933***	0.919***
	(0.177)	(0.202)	(0.171)	(0.189)	(0.204)	(0.184)	(0.306)	(0.318)	(0.304)
$ln(R\&D)_{i,t-1}$	0.279***	0.278***	0.282***	0.260***	0.259***	0.263***	0.308***	0.308***	0.309***
	(0.064)	(0.062)	(0.060)	(0.087)	(0.087)	(0.086)	(0.099)	(0.099)	(0.099)
$Knowledge_{it}^{Down,depreciation}$		-0.036			-0.072			0.003	
		(0.163)			(0.101)			(0.120)	
$Knowledge^{Up,IO}\_it$			0.378**			0.282			0.084
			(0.173)			(0.205)			(0.219)
R^2	0.915	0.915	0.916	0.901	0.901	0.902	0.885	0.885	0.886
No. of Sectors	94	94	94	94	94	94	94	94	94
No. of Obs	1,847	1,847	1,847	1,847	1,847	1,847	1,847	1,847	1,847
Fixed Effects		Sector, Year			Sector, Year			Sector, Year	

Notes. This table reproduces Table 2 in the paper by incorporating exponential discounting of knowledge stocks.

Table A.11. U.S. Evidence of Knowledge Spillover Through Innovation Networks
Additional Innovation Measure

$Y=$	ln(Patent Value)			
	(1)	(2)	(3)	(4)
$Knowledge_{it}^{Up}$	0.929***	0.926***	0.914***	1.545***
	(0.316)	(0.329)	(0.316)	(0.447)
$ln(R\&D)_{i,t-1}$	0.307***	0.307***	0.307***	0.332***
	(0.100)	(0.099)	(0.100)	(0.124)
$Knowledge_{it}^{Down}$		0.008		
		(0.117)		
$Knowledge_{it}^{Up,IO}$			0.073	
			(0.223)	
Specification	OLS	OLS	OLS	IV 2nd Stage
R^2	0.885	0.885	0.885	0.152
No. of Sectors	94	94	94	94
No. of Obs	1,847	1,847	1,847	1,113
Fixed Effects		Sector, Year		

Notes. This table reproduces Table 2 in the paper with the additional innovation measure of patent value from (Kogan et al., 2017) based on the stock market reaction to patent approval.

Table A.12. Global Evidence of Knowledge Spillover Through Innovation Networks
Different Knowledge Periods

Panel (a): $\tau = 5$

Y=	ln(Patents)			ln(Cites)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Knowledge_{mit}^{Up,\tau=5}$	0.105**	0.125**	0.099*	0.227***	0.280***	0.224***
	(0.050)	(0.055)	(0.051)	(0.078)	(0.087)	(0.080)
$ln(R\&D)_{mi,t-1}$	0.034***	0.034***	0.032***	0.047***	0.048***	0.047***
	(0.010)	(0.010)	(0.010)	(0.014)	(0.014)	(0.014)
$Knowledge_{mit}^{Down,\tau=5}$		-0.031			-0.085*	
		(0.036)			(0.048)	
$Knowledge_{mit}^{Up,IO}$			0.080			-0.037
			(0.064)			(0.071)
R^2	0.969	0.969	0.969	0.944	0.944	0.944
No. of Country x Sectors	564	564	550	564	564	550
No. of Obs	10,552	10,552	10,318	10,552	10,552	10,318
Fixed Effects	Country x Sector, Country x Year, Sector x Year					

Panel (b): $\tau = 20$

Y=	ln(Patents)			ln(Cites)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Knowledge_{mit}^{Up,\tau=20}$	0.199***	0.249***	0.198***	0.426***	0.499***	0.426***
	(0.059)	(0.064)	(0.059)	(0.085)	(0.092)	(0.087)
$ln(R\&D)_{mi,t-1}$	0.033***	0.034***	0.032***	0.045***	0.047***	0.045***
	(0.010)	(0.010)	(0.010)	(0.014)	(0.013)	(0.014)
$Knowledge_{mit}^{Down,\tau=20}$		-0.138***			-0.202***	
		(0.048)			(0.066)	
$Knowledge_{mit}^{Up,IO}$			0.081			-0.034
			(0.064)			(0.070)
R^2	0.969	0.969	0.969	0.944	0.945	0.945
No. of Country x Sectors	564	564	550	564	564	550
No. of Obs	10,552	10,552	10,318	10,552	10,552	10,318
Fixed Effects	Country x Sector, Country x Year, Sector x Year					

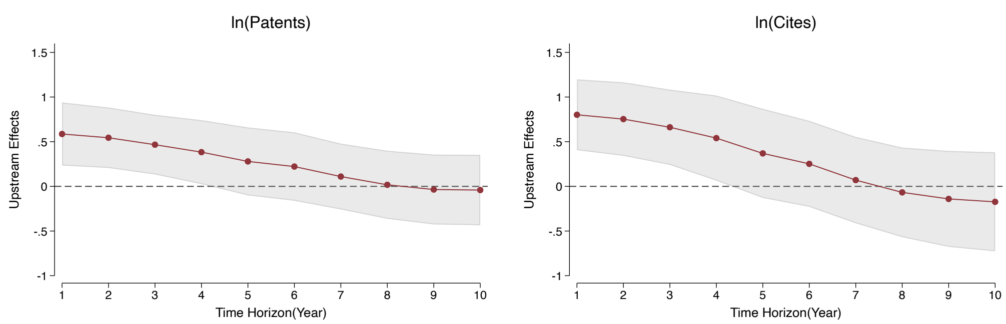
Notes. This table reproduces Table 3 in the paper. The key difference is this table uses different τ periods to calculate knowledge accumulated through the innovation network. Table 3 uses $\tau = 10$, while this table uses alternative values of $\tau = 5$ and $\tau = 10$.

Table A.13. U.S. Evidence of Knowledge Spillover Through Innovation Networks
Exploring the I-O Linkages

$Y=$	ln(Patents)		ln(Cites)		ln(Patent Value)	
	(1)	(2)	(3)	(4)	(5)	(6)
$Knowledge_{it}^{Up}$	0.508***		0.743***		0.914***	
	(0.174)		(0.196)		(0.316)	
$ln(R\&D)_{i,t-1}$	0.279***	0.304***	0.261***	0.297***	0.307***	0.351***
	(0.060)	(0.057)	(0.086)	(0.083)	(0.100)	(0.098)
$Knowledge_{it}^{Up,IO}$	0.363**	0.462***	0.268	0.413*	0.073	0.251
	(0.173)	(0.174)	(0.205)	(0.209)	(0.223)	(0.238)
R^2	0.917	0.914	0.902	0.898	0.885	0.880
No. of Sectors	94	94	94	94	94	94
No. of Obs	1,847	1,847	1,847	1,847	1,847	1,847
Fixed Effects	Sector, Year		Sector, Year		Sector, Year	

Notes. This table reproduces Table 2 in the paper by incorporating standalone knowledge spillovers from the I-O network in columns (2), (4), and (6).

Figure A.9. Dynamic Responses of Innovation Output to Upstream Knowledge



Notes. This figure presents how the focal sector's innovations dynamically respond to past innovations from upstream sectors in the innovation network. The coefficients are from regressions of focal sectors' innovations at times $t+1$ through $t+10$ on upstream knowledge measured at time-t. We control for log R&D with time-1 lag as well as sector and year fixed effects. The half-life of the dynamic effects is about 4 years.

E.3 Using R&D Tax Credit as an Instrument for Upstream R&D

Our analysis on the impact of upstream innovation (i.e., Tables 2 and 3) is subject to the concern of common shocks: a group of sectors connected to each other via citation linkages may face similar demand, supply, and investment opportunities, leading to co-movements of innovation activities. Serial correlations in these common shocks would lead to a positive coefficient β_1 in regression (27) even without cross-sector knowledge spillovers. This is a classic version of the "reflection problem" documented in Manski (1993) and, more relevant to our setting, in Bloom et al. (2013). As noted in Bloom et al. (2013), since knowledge spillovers through the innovation network are entered lagged at least one year (and up to ten years), and because fixed effects and other controls are included in the estimation, the potential bias is likely small. Nevertheless, to further resolve this issue, we consider an instrumental variable strategy based on R&D tax credits, a method widely used in innovation literature. Here we present only the basic framework and how we adapt the strategy to our setting. We refer readers to a classic use case in Bloom et al. (2013) and the Online Appendix of the paper.

This instrumental variable strategy shocks R&D activities using the user cost of R&D capital, which in turn is often closely tied to tax policies and subsidies like R&D tax credit. User cost of R&D is affected by two types of R&D tax credit, federal tax rules that interact with different firms differently (e.g., based on past R&D expenses, etc.), and state-level tax credits, depreciation allowances, and corporation taxes that affect firms differently based on the location of R&D activities.

- For state-level tax credits, we obtain the state-by-year R&D tax price data, available for 1970 to 2006, from Wilson (2009). These data are further aggregated to sector-year-level tax price of R&D by calculating the weighted sectoral average, which is weighted using the total number of inventors in a sector who work in each state (ten-year average of inventor shares). In other words, if a sector has more inventor weight in a high tax credit state (thus the user cost is lower), the sector will have a lower user cost of R&D in our aggregation. Using inventor shares is common practice in this literature as R&D labor cost is often the key target of R&D tax policies.

- For the federal tax component, which is shown to be less powerful for explaining sector-level R&D activities in our setting, we follow the approach in Bloom et al. (2013) and construct a firm-year level federal tax-driven user cost of R&D. This firm-year-level measure is then further aggregated to sector-year level by weighting each firm according to its size measured using the number of inventors.

The R&D user cost can also be calculated at the country-sector-year level. For this purpose, we obtain data from Thomson (2017), who provides the user cost estimates for different types of R&D input, in particular labor and capital, in different country-years. Following Thomson (2017), we calculate the tax price at the country-sector-year level using the weight-average tax price of different expenditure types with lagged expenditure share on those types as weights. For example, the "Apparel, dressing, and dyeing of fur" industry has a capital-labor R&D composition ratio of 92% to 8%, then the R&D user cost is a weighted average using those ratios. This estimate covers 25 WIOD countries from 1980 to 2006.

We implement the empirical strategy by first projecting sectoral innovation on the instrument. Table A.14 demonstrates that the instruments have power in predicting sectoral innovation output both in the U.S. (column 1) and globally (column 2). In both models, we control for fixed effects at the cross-section and in the time series. From these models, we calculate sectoral innovation predicted by these tax credits, $\ln n_{it}^{TAX}$.

Table A.14. Predicting Sectoral Patent Count Using R&D Tax Credits

	United States	Global
$Y=$	ln(Patents) (1)	ln(Patents) (2)
$ln(User\ Cost\ of\ R\&D\ Capital)$	-11.774***	-0.288**
	(4.041)	(0.134)
Fixed Effects		
Sector	Yes	
Year	Yes	
Country x Sector		Yes
Country x Year		Yes
Sector x Year		Yes
R^2	0.866	0.969
No. of Sectors	158	
No. of Country x Sectors		1,242
No. of Obs	4,615	18,799

Notes. This table presents evidence that the user cost of R&D capital predicts patent output. Standard errors are clustered at the sector and year levels.

In the main 2SLS analysis, for each sector, we calculate upstream knowledge using the same equation as in (26), replacing the realized sectoral innovation with the fitted values $\ln n_{it}^{TAX}$. We denote this fitted value of the knowledge as $\text{Knowledge}_{it}^{Up,TAX}$. The variable $\text{Knowledge}_{it}^{Up,TAX}$ is then used as an instrument in the analysis in (27). We report the first-stage regressions in Table A.15, and domestic and global versions of the knowledge diffusion results in Tables A.16 and A.17, corresponding to Tables 2 and 3 in the paper.

Table A.15. Predicting Sectoral Patent Count Using R&D Tax Credits

	United States	Global
$Y=$	$Knowledge_{it}^{Up}$ (1)	$Knowledge_{mit}^{Up}$ (2)
$Knowledge_{it}^{Up,IV}$	1.110***	
	(0.051)	
$Knowledge_{mit}^{Up,IV}$		0.540***
		(0.045)
$ln(R\&D)_{i,t-1}$	0.034***	0.001
	(0.012)	(0.004)
Fixed Effects		
Sector	Yes	
Year	Yes	
Country x Sector		Yes
Country x Year		Yes
Sector x Year		Yes
F-statistics	465.9	146.2
R^2	0.983	0.982
No. of Sectors	94	
No. of Country x Sectors		280
No. of Obs	1,113	4,467

Notes. The first-stage regression, instrumental variable is the fitted value of upstream innovation accumulated through the innovation network. Standard errors are clustered at the sector and year levels.

Table A.16. US Evidence of Knowledge Spillovers Through Innovation Networks–*Second-Stage IV Results*

$Y=$	ln(Patents)			ln(Cites)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Knowledge_{it}^{Up}$	0.679**	0.683**	0.687**	0.974***	0.995***	0.980***
	(0.266)	(0.268)	(0.263)	(0.279)	(0.281)	(0.277)
$ln(R\&D)_{i,t-1}$	0.269***	0.271***	0.257***	0.174**	0.184**	0.163*
	(0.070)	(0.074)	(0.068)	(0.082)	(0.086)	(0.084)
$Knowledge_{it}^{Down}$		-0.013			-0.074	
		(0.131)			(0.109)	
$Knowledge_{it}^{Up,IO}$			0.337			0.319
			(0.393)			(0.439)
R^2	0.152	0.151	0.152	0.099	0.100	0.094
No. of Sectors	94	94	94	94	94	94
No. of Obs	1,113	1,113	1,113	1,113	1,113	1,113
Fixed Effects		Sector, Year			Sector, Year	

Notes. Second-stage regression. Same setting as in Table 2.

Table A.17. Global Evidence of Knowledge Spillovers Through Innovation Networks–*Second-Stage IV Results*

$Y=$	ln(Patents)			ln(Cites)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Knowledge_{mit}^{Up}$	0.202*	0.206*	0.222*	0.405***	0.413***	0.421***
	(0.113)	(0.109)	(0.116)	(0.147)	(0.147)	(0.151)
$ln(R\&D)_{mi,t-1}$	0.066***	0.066***	0.066***	0.072***	0.072***	0.072***
	(0.014)	(0.014)	(0.014)	(0.022)	(0.022)	(0.022)
$Knowledge_{mit}^{Down}$		-0.008			-0.018	
		(0.084)			(0.118)	
$Knowledge_{mit}^{Up,IO}$			-0.130			-0.057
			(0.380)			(0.378)
R^2	0.040	0.040	0.032	0.031	0.032	0.029
No. of Country x Sectors	280	280	275	280	280	275
No. of Obs	4,467	4,467	4,412	4,467	4,467	4,412
Fixed Effects	Country x Sector, Country x Year, Sector x Year					

Notes. Second-stage regression. Same setting as in Table 3.

E.4 Additional Results on R&D Misallocation

This subsection presents additional results that quantify R&D misallocation, supplementing Section 5.

- Tables A.18 and A.19 present cross-country and time-series correlations of optimal R&D allocation γ.
- Figure A.10 presents analysis using alternative parameters of ρ/λ; Figure A.11 presents analysis using data from different years.
- Figures A.12, A.13, and A.14 present analysis using patent outputs and OECD R&D expenditure shares as innovation allocation measures, supplementing analysis using R&D expenditure shares (aggregated from firm-level data) in the paper.
- Figure A.15 provides additional analysis on the time series of R&D misallocation and implied welfare cost.
- Table A.20 summarizes the robustness of our quantitative analysis across different specifications of Ω, ρ, and λ.
- Figure A.16 presents evidence on R&D misallocation within 1-digit IPC patent classes.

Table A.18. Unilaterally Optimal R&D Allocations Across Countries

Countries	US	Japan	China	South Korea	Germany	Russia	France	UK	Canada	Netherlands	EU
US		0.97	0.90	0.93	0.95	0.84	0.94	0.94	0.92	0.95	0.95
Japan	0.91		0.93	0.94	0.96	0.87	0.94	0.94	0.93	0.94	0.96
China	0.87	0.93		0.95	0.91	0.91	0.91	0.90	0.91	0.90	0.94
South Korea	0.85	0.89	0.84		0.92	0.83	0.90	0.90	0.88	0.89	0.92
Germany	0.77	0.89	0.79	0.82		0.85	0.97	0.96	0.94	0.97	0.99
Russia	0.70	0.76	0.86	0.60	0.57		0.84	0.82	0.90	0.86	0.86
France	0.81	0.89	0.87	0.73	0.73	0.76		0.98	0.94	0.97	0.98
UK	0.84	0.89	0.86	0.73	0.73	0.76	0.99		0.94	0.97	0.98
Canada	0.78	0.88	0.88	0.72	0.71	0.84	0.97	0.96		0.95	0.95
Netherlands	0.83	0.89	0.87	0.74	0.72	0.76	0.98	0.97	0.96		0.97
EU	0.87	0.96	0.91	0.82	0.90	0.74	0.95	0.95	0.93	0.94	

Notes. This table shows the pair-wise correlations of optimal R&D allocations γ across countries using country-specific statistics as of 2010. The lower triangular panel shows the Pearson correlation coefficients; the upper triangular panel shows Spearman's rank correlation.

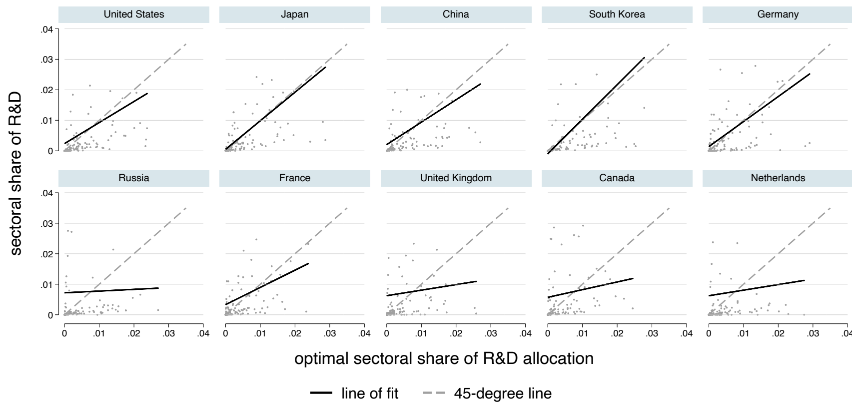
Table A.19. Unilaterally Optimal US R&D Allocations of Across Time

Time Period	2020	2010	2000	1990	1980
2020		1.00	0.99	0.98	0.98
2010	0.99		0.99	0.98	0.98
2000	0.97	0.97		1.00	0.99
1990	0.96	0.94	0.99		1.00
1980	0.94	0.93	0.99	1.00	

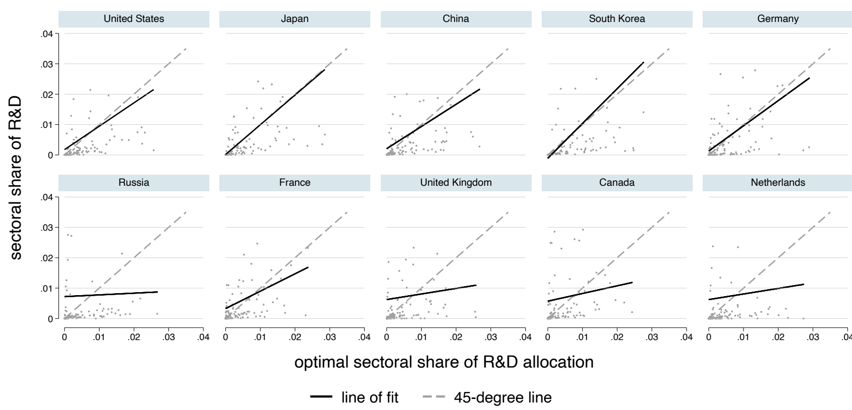
Notes. This table shows the pair-wise correlations of optimal R&D allocations γ across different time periods using U.S. statistics during the specific year. The lower triangular panel shows the Pearson correlation coefficients; the upper triangular panel shows Spearman's rank correlation.

Figure A.10. Alignment Between Real Allocation and Optimal Allocation Across Countries
Using Alternative Parameter Values

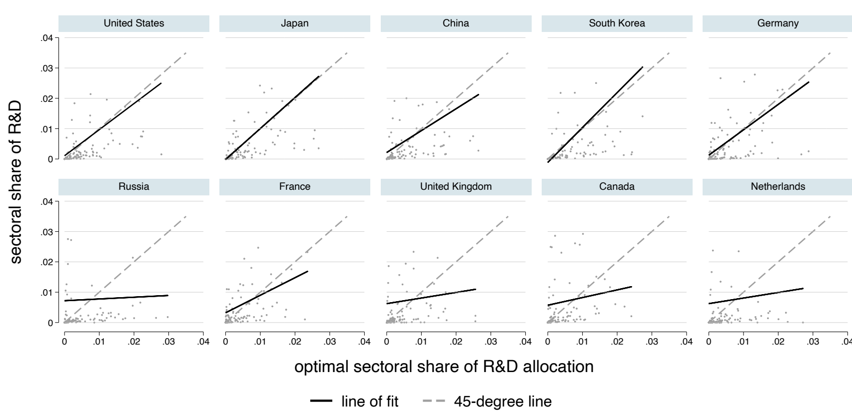
Panel (a): Use $(1+\rho/\lambda)^{-1} = 0.6$



Panel (b): Use $(1+\rho/\lambda)^{-1} = 0.7$



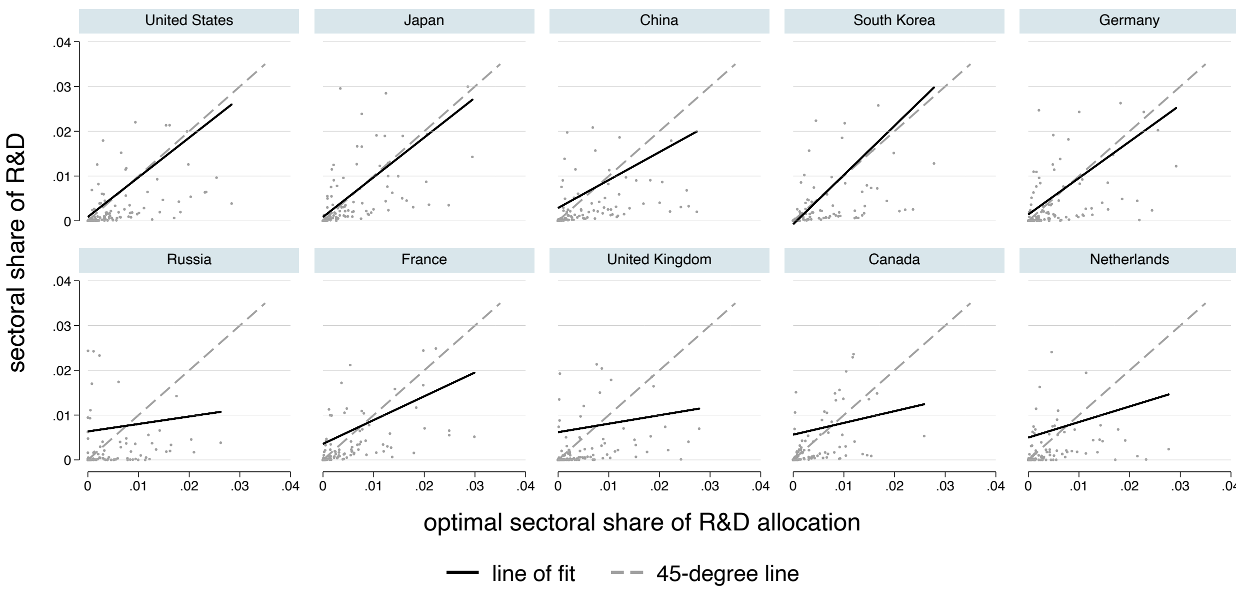
Panel (c): Use $(1+\rho/\lambda)^{-1} = 0.8$



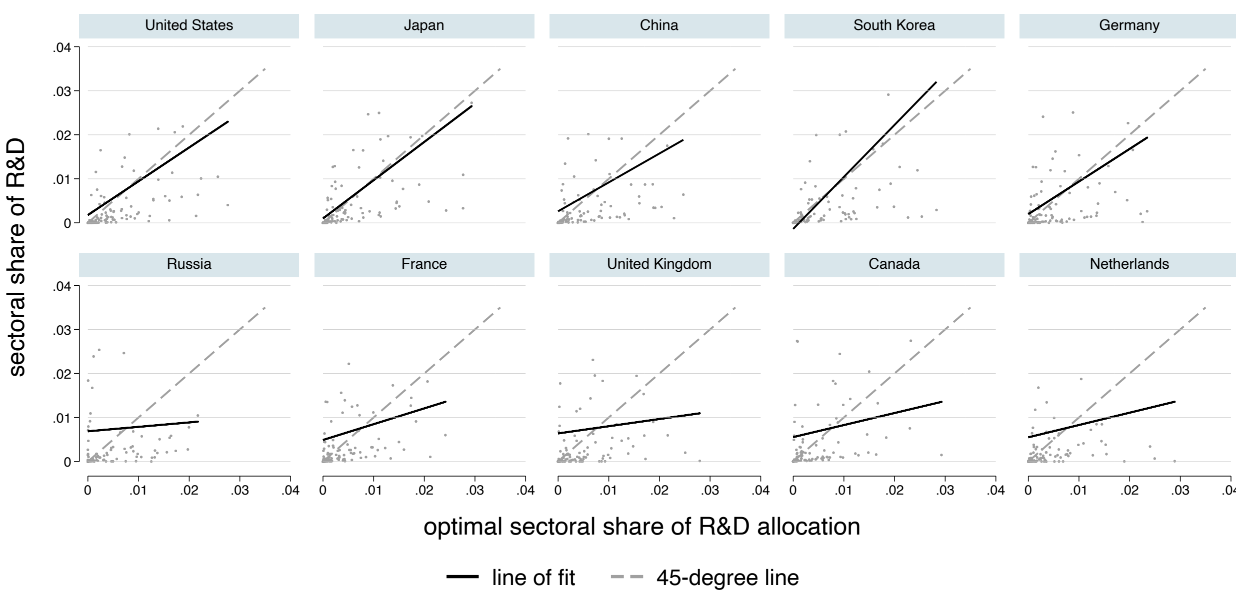
Notes. This table reproduces Figure 6 in the paper with alternative parameter values of ρ/λ.

Figure A.11. Alignment Between Real Allocation and Optimal Allocation Across Countries
Different Years

Panel (a): 2000

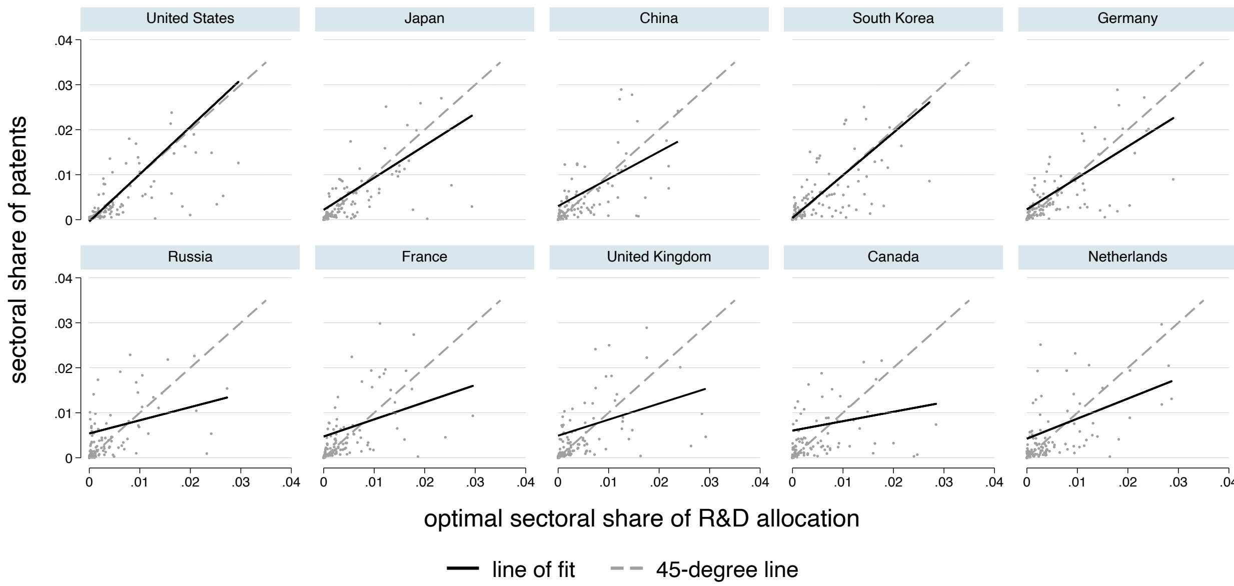


Panel (b): 2005



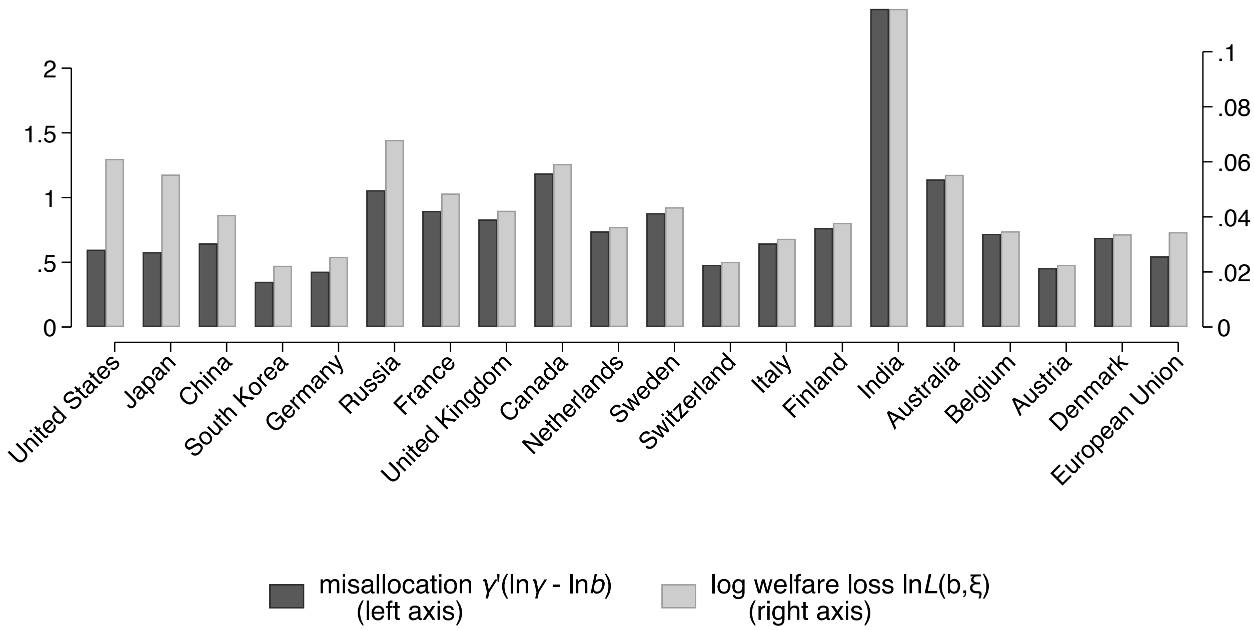
Notes. This figure reproduces Figure 6 in the paper using data from alternative years. The figure shows scatter plots of sectoral R&D expenditure share in total national R&D expenditures against the optimal sectoral share of R&D allocation for top ten innovative countries. The solid line is the linear fit; the dashed line is the 45-degree line.

Figure A.12. Alignment Between Real Allocation and Optimal Allocation Across Countries
Using Sectoral Share of Patents as Real Allocation



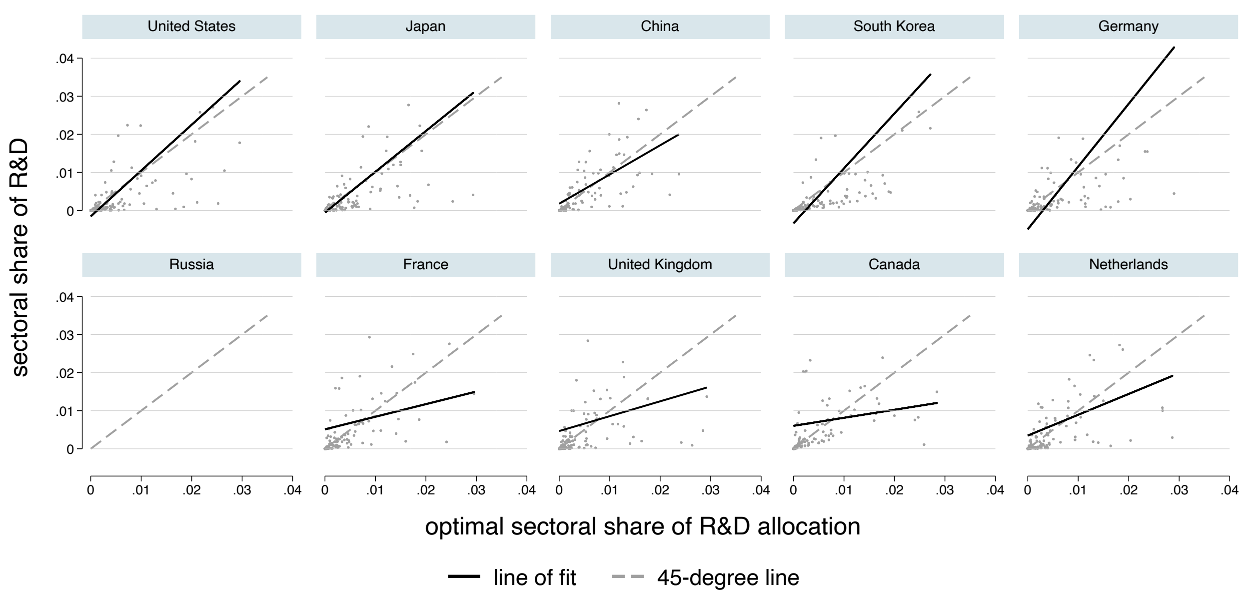
Notes. This figure reproduces Figure 6 in the paper. The figure shows scatter plots of sectoral patent output share in total patent output in the country against the optimal sectoral share of R&D allocation for top ten innovative countries in 2010. The solid line is the linear fit; the dashed line is the 45-degree line.

Figure A.13. Country-Level Welfare Loss from Misallocation
Using Sectoral Share of Patents as Real Allocation



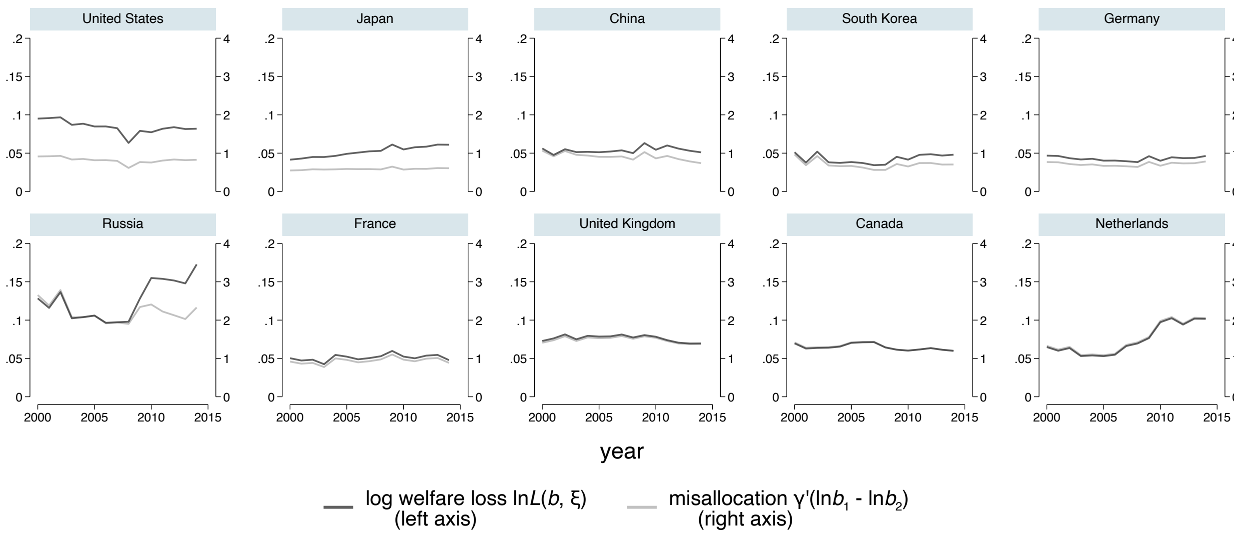
Notes. This table shows the level of R&D misallocation and associated welfare cost during 2010—2014. The table reproduces Figure 7 in the paper, but uses sectoral share of patents, rather than R&D expenditure shares, as the real allocation.

Figure A.14. Alignment Between Real Allocation and Optimal Allocation Across Countries
Using OECD R&D Shares as Real Allocation



Notes. This figure reproduces Figure 6 in the paper. The figure shows scatter plots of sectoral R&D share as reported in the OECD ANBERD database against the optimal sectoral share of R&D allocation for top ten innovative countries in 2010. The solid line is the linear fit; the dashed line is the 45-degree line.

Figure A.15. R&D Misallocation and Welfare Cost Across Countries and Over Time



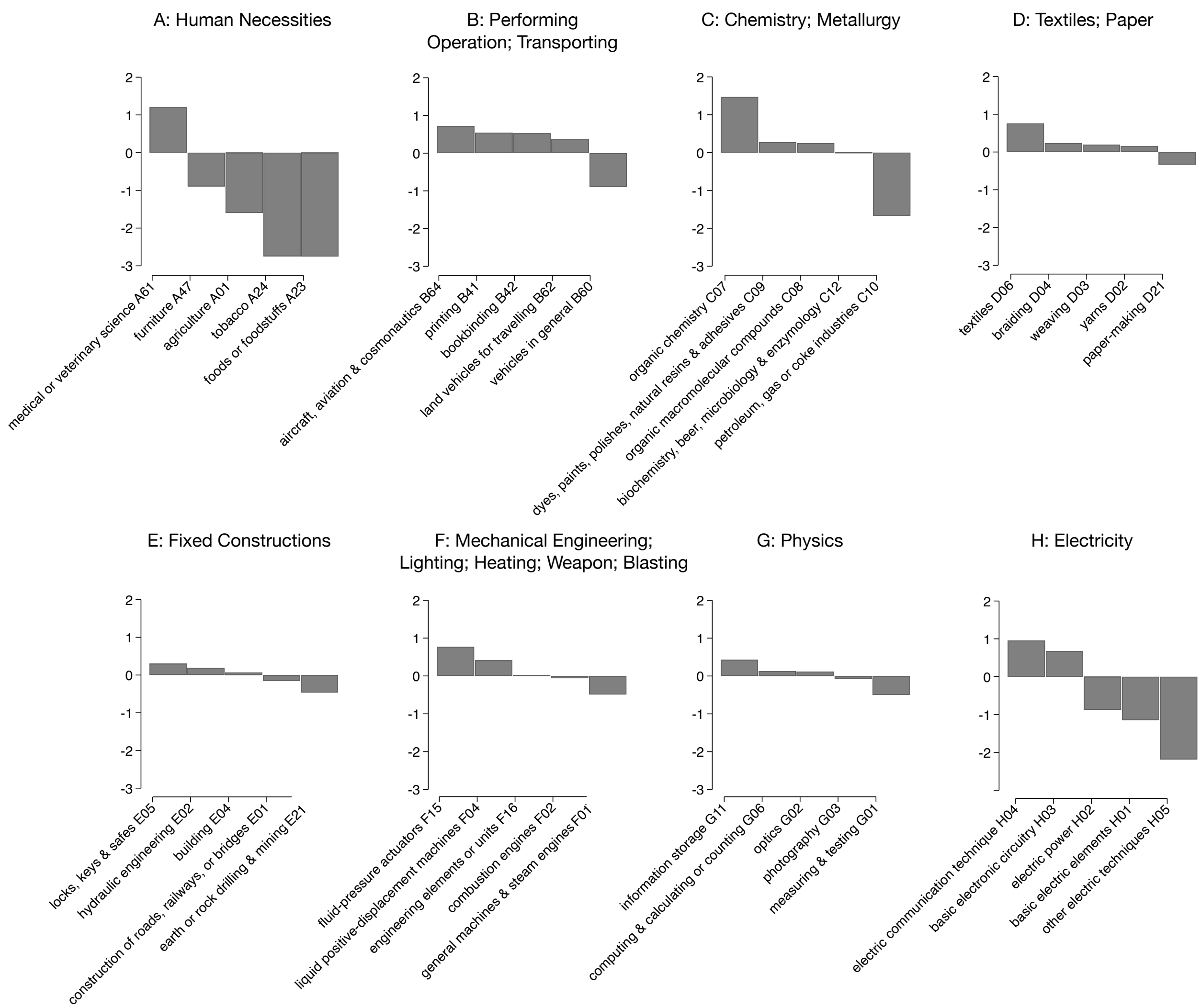
Notes. This figure plots the level of misallocation and welfare cost across countries over time. The calculation focuses on misallocation in top 50 IPC classes by total patents.

Table A.20. Robustness of $\boldsymbol{\gamma}$ and Centrality Across Specification of $\boldsymbol{\Omega}$, ρ, and λ

	Specifications	Average Correlation With Baseline Case			
		Optimal Allocation γ		Centrality a	
		Pearson's r	Spearman's ρ	Pearson's r	Spearman's ρ
A. Alternative Specifications of Ω					
A1	Forward-citation weighted Ω	0.9974	0.9994	0.8916	0.9864
A2	Backward-citation weighted Ω	0.9999	0.9999	0.9974	0.9968
A3	Scaled Ω	0.9955	0.9959	0.5228	0.9327
B. Alternative Values of ρ *and* λ					
B1	Using $(1+\rho/\lambda)^{-1} = 0.4$	0.9976	0.9984	-	-
B2	Using $(1+\rho/\lambda)^{-1} = 0.5$	0.9986	0.9990	-	-
B3	Using $(1+\rho/\lambda)^{-1} = 0.6$	0.9993	0.9996	-	-
B4	Using $(1+\rho/\lambda)^{-1} = 0.7$	0.9999	0.9998	-	-
B5	Using $(1+\rho/\lambda)^{-1} = 0.8$	1.0000	1.0000	-	-
B6	Using $(1+\rho/\lambda)^{-1} = 0.9$	0.9994	0.9997	-	-
B7	Using $(1+\rho/\lambda)^{-1} = 0.95$	0.9988	0.9994	-	-
C. Industry-Specific λ					
C1	ROA (median = 0.1747, s.d. = 0.0268)	0.9982	0.9994	-	-
C2	Gross Profit Margin (median = 0.2242, s.d. = 0.0460)	0.9985	0.9995	-	-
D. Injecting Measurement Errors into Ω					
D1	Adding log-N(0.02, 0.02) noise to Ω	0.9936	0.9934	0.8900	0.8166
D2	Adding log-N(0.04, 0.04) noise to Ω	0.9936	0.9934	0.8199	0.6607
D3	Adding N(0.02, 0.02) noise to Ω	0.9962	0.9944	0.9105	0.8192
D4	Adding N(0.04, 0.04) noise to Ω	0.9951	0.9931	0.8417	0.6696
D5	Adding $\max\{$N(0.02, 0.02), 0$\}$ noise to Ω	0.9963	0.9945	0.9118	0.8299
D6	Adding $\max\{$N(0.04, 0.04), 0$\}$ noise to Ω	0.9952	0.9932	0.8506	0.6941
D7	Adding U[0, 0.02] noise to Ω	0.9978	0.9967	0.9453	0.9287
D8	Adding U[0, 0.04] noise to Ω	0.9965	0.9953	0.9252	0.8861
D9	Adding Exp(0.02) noise to Ω	0.9964	0.9942	0.9068	0.8059
D10	Adding Exp(0.04) noise to Ω	0.9952	0.9928	0.8320	0.6597

Notes. This table reports the average Pearson and Spearman-rank correlation of $\boldsymbol{\gamma}$ and centrality of the innovation networks between the benchmark specification and different sets of alternative innovation network constructions for $\boldsymbol{\Omega}$ (Panel A), alternative values of ρ and λ (Panels B and C), and $\boldsymbol{\Omega}$ with injected errors (Panel D). In rows A1 and A2, we weigh each cite in $\boldsymbol{\Omega}$ construction (24) by the quality (total forward citations received) of either the citing or the cited patent. In row A3, we construct $\omega_{ij} \propto Cites_{i\to j}$ to scale directly with the total citations totally across or ij-pairs (rather than normalized by the citations from i), and we choose the proportionality constant so that the spectral radius of $\boldsymbol{\Omega}$ is equal to one, ensuring endogenous growth as in our baseline model. Rows B1 to B4 consider a range of alternative values for ρ and λ. Changing ρ/λ affects, across all countries, the magnitude of the welfare impact of R&D reallocation, but the cross-country welfare impact still correlates highly with our baseline specification. Panel C considers a specification with sector-specific innovation step size λ_i. Motivated by the decentralized economy constructed in Section 2.7.3, where the step size corresponds to the profit share, we measure λ_i using each sector's median ROA (return on assets) calculated from our firm-level datasets, and calculate the corresponding $\boldsymbol{\gamma}$ and welfare impact of R&D reallocation using the theoretical extension in Section B.8. Finally, in Panel D, we show our quantitative analysis is robust to introducing additional, simulated random errors to $\boldsymbol{\Omega}$. For each of the listed distribution, in each simulation, we add to each element in $\boldsymbol{\Omega}$ random and independent terms drawn from the distribution rescale $\boldsymbol{\Omega}$ to ensure row sum to be one. For each distribution, we simulate the exercise for 10,000 times, and the correlations are reported as average of the benchmark with each of the simulated $\boldsymbol{\Omega}$.

Figure A.16. U.S. R&D Misallocation within 1-digit IPC Classes



Notes. This figure shows U.S. R&D misallocation across 3-digit IPC classes within each 1-digit IPC class. For each 1-digit IPC class, $\ln \boldsymbol{b} - \ln \boldsymbol{\gamma}$ is shown for the top five 3-digit IPC classes ranked by R&D expenditures.