Innovation and Development over the (Very) Long Run*

Francisco Buera[†] Martin Kornejew[‡] Yueran Ma[§]
Ezra Oberfield[¶] Nicholas Trachter[‖]

July 2026

Abstract

We study the spatial distribution of innovation and production over the long run. We assemble a dataset of technological innovations around the world throughout human history from Wikipedia articles, and use large language models to record the time, location, and other background information of each innovation. The data reveal several key facts. First, innovation is far more concentrated than population or output. A single continent or a few countries account for over half of the world's innovations through much of history. Relatedly, innovations per capita vary widely across countries, even among advanced economies. Second, the center of innovation shifts across continents and countries over the millennia. In the last century, innovations per capita have been stagnant in many European countries while rising steadily in the United States. Third, a country's growth increases following more innovation of its own as well as more innovation from the global leader, consistent with diffusion. Motivated by these facts, we develop a model of innovation, knowledge flows, and technology adoption that gives a simple account of the evolution of the global distribution of innovation and output over the last 500 years. The model helps explain the concentration of innovation relative to output, the persistence in the leadership of innovation, and the great divergence in output per capita. Our estimated model shows that differential access to world innovations contributes to most of modern-time cross-country output gaps, and sheds light on the rise and fall of innovation leaders.

*We thank Jingyi Jia, Howard Guankun Li, and Marco Zhang for extraordinary research assistance. The views expressed herein are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

†Washington University in St. Louis and NBER (fjbuera@wustl.edu).

‡Bocconi University (martin.kornejew@unibocconi.it).

§University of Chicago Booth School of Business and NBER (yueran.ma@chicagobooth.edu).

¶Cornell University and NBER (ezraoberfield@cornell.edu).

‖Federal Reserve Bank of Richmond (trachter@gmail.com).

1 Introduction

Technological innovations are fundamental for economic growth and development. Yet we do not have a comprehensive picture of their emergence around the world over the long run, and how it has shaped economic development. Where do significant innovations come from throughout human history, and how concentrated are they? Is the volume of innovation proportional to population, or are some places disproportionately innovative? Do centers of innovation persist, or do they rise and fall? And how much do innovation activities and their concentration affect output in different parts of the world?

These questions have been difficult to analyze on a global scale over the long run, in large part due to data constraints. Patent data, the most commonly used measure of inventive activity, can provide information for specific countries, and more broadly in recent decades, but are difficult to compare across countries and over time. Similarly, R&D expenditures and researcher counts are available only for recent decades and a subset of countries. Economic historians have a long tradition of studying technological developments, which has produced rich qualitative accounts for the world in general as well as for specific countries and episodes (Usher, 1929; Needham, 1954; Landes, 1969; Mokyr, 2002; Gordon, 2016), but broad-based, long-run quantitative analyses of innovations have been limited.

In this paper, we assemble a dataset of significant technological innovations throughout human history, and use it to document a set of facts about the emergence of innovations around the world in the long run. We then develop a framework of innovation, knowledge flows, and technology adoption to interpret these facts, and take the framework to the data.

Our dataset is built from the universe of Wikipedia articles. Following the insights of Laouenan et al. (2022), Asirvatham (2024), and Kalyani et al. (2025), we use large language models to filter Wikipedia articles for those describing technological innovations, and extract their background information such as the time and place of invention, the type of inventor, and the domain of application. After filtering and validation, the final dataset records 54,079 innovations since the beginning of human history around 5 million years ago to 1999 (we end in the 20th century to make sure innovations have had enough time to be cataloged in Wikipedia). Our current analyses focus on innovations since year 0 at the continent level and since 1500 at the country level (output data are sparse before 1500 which also limits our subsequent model estimation to this time period). The data capture the canonical innovations of successive eras (e.g., the Newcomen atmospheric engine, the Watt steam engine, the telephone, the automobile, the airplane, the transistor, the microprocessor, and the mobile phone). We cross-check the

Wikipedia-based measure against patent data available in recent decades, and verify that the two align with each other.

The data reveal several facts. First, innovation is highly concentrated geographically, more than population or economic output. For example, a single continent has almost always accounted for over half of the world's innovations. Over the last three centuries, it took only one or two countries to account for half of the world's innovations, while these countries contribute less than 10% of the world's population and a minority of the world's GDP. The gap between innovation and output is present even among high-income countries, which stands in contrast to a simple narrative in which frontier countries all innovate (while laggards imitate).

Second, the geographic center of innovation shifts over time. In early history, the center was Africa (e.g., Egypt) and Asia (e.g., Mesopotamia) at the beginning of human civilizations, then Europe around Greek and Roman times, and East Asia afterwards. Later, Europe emerged as the dominant center during the Scientific Revolution and maintained this position through the Industrial Revolution. Since the 19th century, the United States has risen steadily to become the single largest source of technological innovations by the early 20th century. Within these regions, individual countries rose and fell at a faster, though still gradual, pace.

Third, innovations per capita in the world have increased tremendously over the very long run, while individual countries may rise and fall or stagnate. For instance, the United Kingdom had high innovations per capita during and after the First Industrial Revolution, but its innovation intensity has stagnated since then—even though the country did not experience obvious deterioration in institutional quality. A similar pattern holds for several other European countries, most notably Germany after the 1930s. Meanwhile, the United States has experienced a substantial increase in innovations per capita over the past two centuries. Across countries, the heterogeneity in innovations per capita is large even among advanced economies, and a country's innovation trajectory can diverge substantially from its output path. For example, Australia, Canada, and many European countries (e.g., Austria, Belgium, Greece, and Spain), with the exception of Switzerland, have always had limited innovations per capita, yet have obtained a decent level of GDP per capita. Countries such as postwar Japan and China have also achieved rapid growth with modest homegrown innovation. High output, in other words, can be attained with either high volumes of innovation (e.g., the United States) or limited homegrown innovation (e.g., many parts of the developed world).

Fourth, subsequent economic growth increases following both a country's own innovation and a rise in the global leader's innovation. Countries with more innovations over the past decade experience higher future growth in real GDP per capita, controlling for current output and recent growth trajectories. Moreover, the global innovation leader's activity has a notable positive relationship with subsequent growth in *other* countries, suggesting that the benefits of innovation diffuse across borders. Diffusion can help explain how output remains relatively dispersed even as innovation stays geographically concentrated—countries that do not innovate much themselves can still grow by adopting technologies developed elsewhere.

These facts are informative about the economic forces shaping innovation and growth around the globe. The heterogeneity in innovations per capita—even among countries with similar populations and output—suggests that innovations are not simply a function of scale. The scarcity or stagnation of innovation in countries with stable institutions points to the limitation of purely institutional accounts. And the evidence on technology diffusion suggests that even as innovation concentrates, its benefits spread increasingly broadly.

Motivated by these facts, we develop a simple framework incorporating innovation and diffusion. Individuals at each location choose among three activities: i) research, which creates new ideas; ii) adoption, which commercializes existing ideas into varieties; and iii) production. In the model, innovation refers to ideas whose productivity substantially exceeds the existing frontier; diffusion refers to the flow of ideas across country borders, which occurs stochastically mediated by frictions. The model provides simple examples that speak to our empirical findings. First, innovations per capita are more geographically concentrated than output per capita because diffusion allows countries to produce using mixtures of ideas created around the world, compressing variation in output. Second, leadership in innovation among continents is more persistent than leadership among countries within a continent because ideas diffuse more strongly within continents than across them. Third, holding diffusion frictions fixed, a higher growth rate of innovation leads to more dispersion in output per capita across countries. This is consistent with the great divergence, which followed a rise in the growth rate of innovation.

The model delivers two estimating equations that we take to data on innovation, output, and population for 63 countries from 1500 to 1999. We include countries with GDP coverage for at least ten decades during this period. Using likelihood-based methods, we estimate the model's parameters and recover diffusion frictions. The model distinguishes between two types of frictions: barriers to adoption, which delay producers' access to ideas developed elsewhere, and barriers to research spillovers, which delay innovators' access to foreign ideas. The estimates reveal substantial cross-country variation in both barriers, and that barriers to research spillovers tend to be stronger than barriers to adoption.

We use the estimated model to ask what drives the geography of innovation and output. A structural variance decomposition shows that barriers to adoption account for the bulk of the modern cross-country output gap, explaining nearly nine-tenths of the variance of output per capita by 1900. Barriers to research spillovers play a smaller role in the dispersion of innovation for much of the sample, but their importance rises in the twentieth century, when they account for roughly half of cross-country variation in innovation. Removing barriers to adoption would compress the distribution of output; removing barriers to research spillovers raises the world's overall level of output without substantially reshaping its distribution. Finally, a case study of the rise and fall of the United Kingdom and the United States shows that the model attributes their ascents primarily to domestic research productivity rather than unusually low barriers to research spillovers. These spillovers account for a growing share of their lead only in the modern era, when faster diffusion and larger global idea stocks make external knowledge more valuable.

Finally, we use several additional analyses to sharpen the interpretation of the geographic concentration of innovation. These analyses ask whether concentration reflects field-specific specialization, technological spillovers, public research, intellectual-property institutions, or broader agglomeration forces. The evidence points most strongly to technological and agglomerative forces. First, we find that innovation leadership is broad-based across technological domains: countries rise and fall as innovation leaders because of changes in local innovative activity, not merely because they specialize in domains that become globally important. Spillovers are also much stronger within domains than across domains, suggesting that leader innovations transmit usable technological insights rather than only a general culture of innovation. Second, the data show that both private innovations (e.g., innovations by private companies and individuals) and public innovations (e.g., innovations by universities and governments) are concentrated, and private innovation appears, if anything, to lead public innovation, which argues against an explanation centered on government-directed research or public science leading private activity. Third, like technological innovation, mathematics and literature are far more concentrated than population or output. This common concentration across creative activities points to powerful agglomeration forces, while the high concentration of mathematics suggests that patents and formal appropriability institutions are unlikely to be the main source of innovation concentration.

In summary, with centuries of global data, our findings highlight concentrated innovations together with widespread growth. Diffusion frictions play a critical role in shaping countries' output. The production of ideas concentrates; what spreads them—or does not—is what drives prosperity around the world.

Related literature This paper makes two main contributions. First, we provide new facts about innovation and development over the very long run. A large body of work documents that innovation clusters geographically, driven by localized knowledge spillovers and agglomeration economies (Krugman, 1991; Jaffe, Trajtenberg, and Henderson, 1993; Audretsch and Feldman, 1996; Akcigit et al., 2018; Moretti, 2021; Giroud, Liu, and Mueller, 2026; Chikis, Kleinman, and Prato, 2025); see Carlino and Kerr (2015) for a survey. This literature focuses primarily on within-country variation over recent decades. Over a longer time span, Serafinelli and Tabellini (2022) show that notable creative individuals (spanning arts, humanities, science, and business) in European cities from the 11th to the 19th centuries are spatially concentrated and clustered across disciplines. Andrews and Whalley (2022) use 150 years of geolocated United States patents to document persistent but gradually declining spatial concentration of innovation. Several papers also investigate the relationship between technology and output over the long run. Comin, Easterly, and Gong (2010) find that the use of technology in 1500 can predict output today based on qualitative coding of the adoption of selected technologies, while Akcigit, Grigsby, and Nicholas (2017) document that patented inventions contribute to long-run economic growth in U.S. data since the late 19th century. Bergeaud, Gozen, and Van Reenen (2026) show that patent trajectories as measured using international patent data over the past century are positively correlated with TFP growth. Our paper uses a different data source—Wikipedia articles processed with large language models (Kalyani et al., 2025; Asirvatham, Mokski, and Shleifer, 2026)—to trace innovation concentration around the world throughout human history.

Second, we provide new theoretical and quantitative insights about the forces behind innovation concentration and the drivers of per-capita output differences across countries. Our framework builds on models of idea-driven growth (Romer, 1990; Grossman and Helpman, 1991; Aghion and Howitt, 1992), particularly the strand focusing on semi-endogenous growth (Jones, 1995; Kortum, 1997). Our analysis emphasizes the diffusion of technological innovations across countries (Eaton and Kortum, 1999; Comin and Hobijn, 2004; Klenow and Rodriguez-Clare, 2005; Comin and Hobijn, 2010; Alvarez et al., 2013; Benhabib, Perla, and Tonetti, 2014; Comin and Mestieri, 2018; Arkolakis et al., 2018; Buera and Oberfield, 2020; Cai, Li, and Santacreu, 2022; Lind and Ramondo, 2023; Liu and Ma, 2024), and in particular a tradeoff between innovation and adoption (Jovanovic and Rob, 1989; Acemoglu, Aghion, and Zilibotti, 2006; König, Lorenz, and Zilibotti, 2016; Perla, Tonetti, and Waugh, 2021; Sampson, 2023; Trouvain, 2024). Like Buera and Oberfield (2020), we model knowledge diffusion as a process in which new ideas build on insights from existing ideas rather than simply replicating them. Like Gerschenkron (1962) and Parente and Prescott (1994), we emphasize barriers to adoption. We model these barriers, following Eaton and Kortum (1999), as a stochastic delay in diffusion of ideas developed abroad, as in Cai, Li, and Santacreu (2022) and Lind and Ramondo (2023). Our estimated model shows that differential access to world innovation contributed to most of modern-time cross-country output gaps, in line with the results in Sampson (2023) and Lind and Ramondo (2024).

2 Innovations Dataset from Wikipedia Articles

We aim to measure innovations around the world over a long period of time. To do so, we build on the insights of recent work (Yu et al., 2016; Laouenan et al., 2022; Asirvatham, 2024; Kalyani et al., 2025) to collect information from Wikipedia articles using large language models (LLMs).

2.1 Main Dataset for Technological Innovations

We obtain all English Wikipedia articles classified under STEM topic tags from the Wikimedia English Wikipedia Dumps (Wikimedia Foundation, 2025). First, we filter Wikipedia article titles using LLMs to extract those about technological innovations, as explained in detail in Appendix IA1.1 (e.g., to remove articles about people, places, events, academic journals, etc. that are not about technological innovations). Second, we use LLMs to compile information about these innovations, including the time, location, inventor type (e.g., individuals, private companies, universities/nonprofits, governments), and domain of application (based on the structure of two-digit NAICS codes), as explained in detail in Appendix IA1.2. Finally, we deduplicate overlapping entries and exclude fictional or unrealized innovations, including failed proposals that never produced a working implementation; together, these two filters drop 1,914 innovations.

After these steps, the final dataset records 54,079 innovations since the beginning of human history around 5 million years ago to 1999. Currently we end in 1999 because more recent innovations may not yet be comprehensively recorded. At the continent-level, we focus on results since 0 AD to display broad long-run patterns, and also present the full history for additional information (but data on population and output are challenging to obtain very early on). At the country level, we focus on 1500 onward as innovations per country are rather sparse in ancient times. Appendix Figure IA1 shows the evolution of the global innovation count over time: the number rises sharply from the 18th century onward, consistent with the acceleration of technological progress.

For each innovation, we collect background information from the Wikipedia page content using LLMs (see Appendix IA1.2). Specifically, we record the year when the innovation first became workable and the country where it was first created.[1] We also classify the type of inventor into four categories: individual inventors working independently of large organizations, private companies, government projects (including military initiatives), and universities or nonprofits. Moreover, we assign each innovation a domain of application from 15 categories: 13 following the structure of two-digit NAICS codes, plus military and defense and basic research.

We check several key examples to verify that LLM filtering and information collection are reliable. For example, for the first industrial revolution, our dataset includes the Newcomen atmospheric engine (1712, United Kingdom), the Watt steam engine (1776, United Kingdom), and the high-pressure steam locomotive (1836, United Kingdom). For the second industrial revolution, our dataset includes the Otto engine (1876, Germany), the telephone (1876, United States), the automobile (1886, Germany), and the airplane (1903, United States). For the third industrial revolution, our dataset includes the transistor (1947, United States), the integrated circuit (1958, United States), the microprocessor (1971, United States), the mobile phone (1973, United States), and the personal computer (1974, United States).

We cross-check the Wikipedia-based innovation measure against patent data from PATSTAT (European Patent Office, 2024), which contains bibliographical data on patent applications filed at major patent offices worldwide. For each country-year pair, we count the total number of patent applications using the filing authority as a proxy for the country of origin. As shown in Appendix Figure IA2, there is a strong positive relationship between the cross-sectional percentiles of Wikipedia innovations per capita and patent applications per capita across countries over 1970 to 1999. Patent data have important limitations that motivate our use of Wikipedia: patent records are available only for recent decades and are difficult to compare across countries due to differences in patent systems, filing practices, and legal standards.

A natural question is whether English Wikipedia has a bias toward English-speaking countries, potentially overstating their share of innovation. Appendix Figure IA2 also highlights English-speaking countries (the United States, the United Kingdom, Australia, Canada, Ireland, and New Zealand) in dark circles, with all other countries shown in lighter triangles. If English Wikipedia systematically inflated innovation counts for English-speaking countries, those countries would appear above the regression line—they would have more Wikipedia innovations per patent than non-English-speaking countries. In fact, the English-speaking countries do not appear systematically above the line, suggesting that our measure does not suffer from a significant English-language bias. Appendix Figure IA3 provides a complementary check by plotting the ratio of Wikipedia innovations to patent applications (i.e., the

[1]We follow the Maddison Project Database (Bolt and van Zanden, 2024) in defining and tracking countries through time in order to merge the database's estimates of population and real GDP per capita.

number of Wikipedia innovations per 1,000 patents) from 1960 to 1999 for nine major innovating countries. The ratio for the United Kingdom is similar to that for France, Germany, and Italy. The ratio for the United States is higher, but sometimes surpassed by Italy and the Netherlands. These checks suggest that the Wikipedia data we collect do not have a strong English-language bias, and the geographic concentration we document subsequently is not just an artifact of such a bias.

2.2 Additional Dataset for Mathematical Discoveries and Literature

To compare patterns of technological inventions with achievements in other domains, we also analyze Wikipedia articles on mathematical discoveries (as a clean benchmark of pure science) and literature (as an art genre with a long history and one that has remained popular over time).[2] First, we filter Wikipedia article titles using LLMs to extract those about mathematical discoveries and literature, as explained in detail in Appendix Sections IA2.1 and IA3.1. Second, we use LLMs to compile information about the work from the Wikipedia page content, including the time and location of composition, as explained in detail in Appendix Sections IA2.2 and IA3.2.

2.3 Additional Dataset for Universities

We also construct a country-year measure of university density (see Appendix IA4 for details). We begin with the set of countries in our main country-year panel and use web-grounded LLM queries to compile country-specific lists of universities. We define universities as nationally recognized higher-education institutions that grant undergraduate degrees, including equivalent institutions such as institutes of technology or universities of applied sciences when they award bachelor-level degrees. We exclude graduate-only or research-only institutes, stand-alone training centers, community colleges without university status, and branch campuses. We deduplicate institutions by institutional lineage so that mergers and renamings are counted once.

For each country, the LLM first obtains a target count of universities using official or authoritative sources, including education ministry registries, accreditation agencies, WHED/UNESCO, and official university websites (see Appendix IA4.1). The LLM then constructs a list of university names, which

[2]Other science fields like physics, chemistry, and biology have a large number of discoveries of objects in nature such as planets, compounds, and species, which could make the number of discoveries difficult to interpret. Some of their discoveries are also more applied and more difficult to separate from technological innovations. Other art genres can be less suitable for long-run analyses—for example, film has a short history while painting has become less prevalent due to technological change.

is further enriched with information on founding year, city, official website, and source URLs (see Appendix IA4.2). When the founding year is missing from the initial sources, we search Wikipedia pages for the relevant institution, clean the page text, and use an LLM to extract the founding year only when it is clearly stated in the text. We drop universities without a usable founding year.

We then aggregate the institution-level data to the country-year level. For each country n and year t, we define the stock of universities, Universities_{nt}, as the cumulative number of universities founded in country n by year t. Years before the first observed university foundation are coded as zero. We aggregate East and West Germany into Germany and map countries to standard ISO country codes before merging the series into the main country-year panel.

3 Empirical Facts

This section presents the main empirical findings on the spatial distribution of innovation over the long run. We present results at different levels of geographic aggregation. The continent level offers the broadest view over the longest horizon, from 0 to 1999, using cumulative innovations over the preceding 100 years. The country level covers 1500 to 1999 (as the data for earlier years are more sparse), using innovations over the preceding 50 years. The trailing windows smooth the lumpiness inherent in innovation counts, particularly in the earlier parts of the sample.

3.1 The Spatial Concentration of Innovation over Time

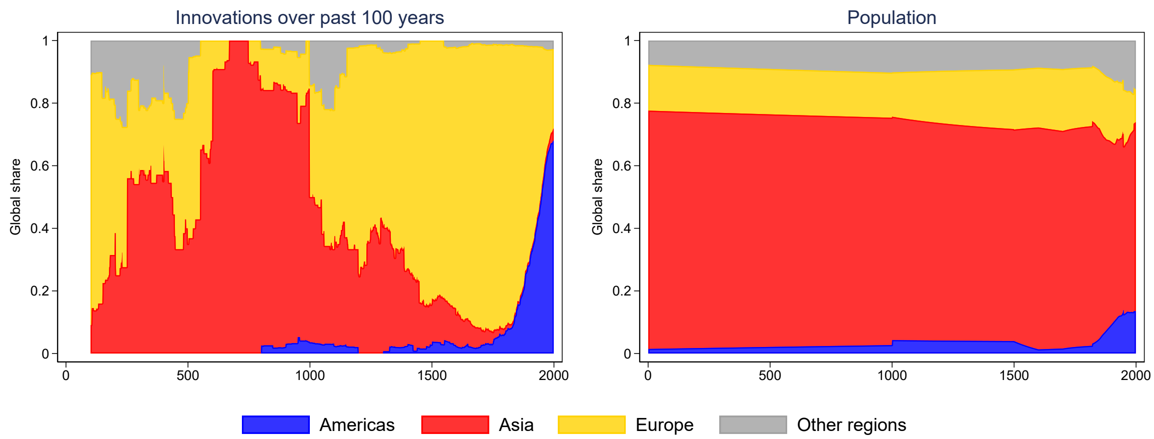
We begin by documenting the patterns of innovation around the world over the very long run. A central finding is that innovation is highly concentrated geographically—much more than population or economic output—and that the geographic center of innovation has shifted substantially through human history.

3.1.1 Innovation Shares

We first examine the share of global innovations originating from each continent and country, as well as the concentration of innovation compared with population and economic output.

Continent level Figure 1 presents the distribution of innovations across continents over the period of 0 to 1999, alongside population. The left panel shows each continent's stacked share of global inno-

Figure 1. Innovation Shares across Continents: 0 to 1999



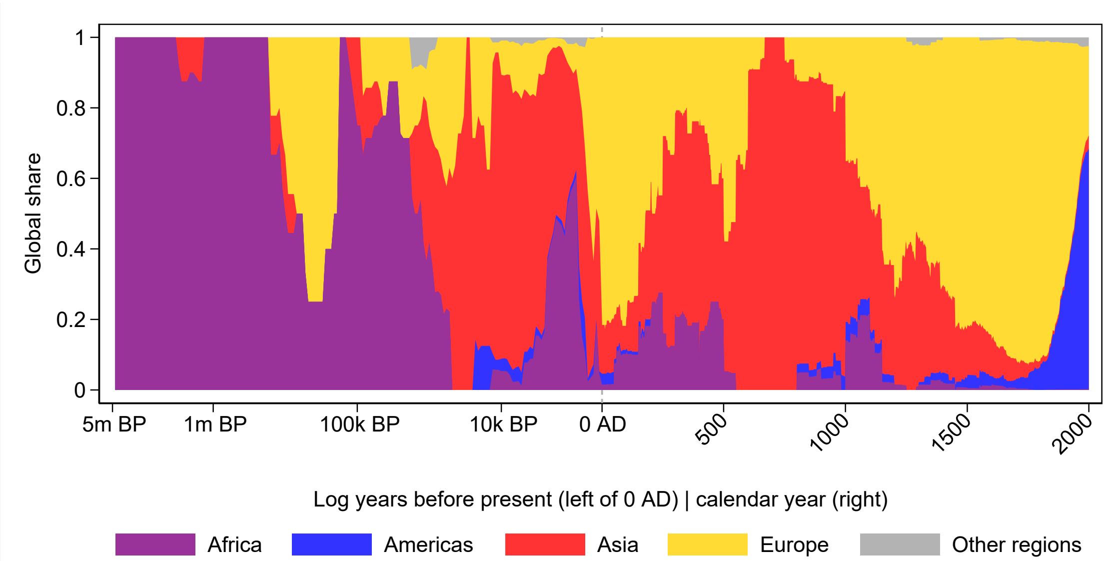
Notes: The figure shows stacked shares of global innovations (left panel) and population (right panel) by continent between 0 and 1999. Innovations are measured as the cumulative count over the past 100 years, to smooth lumpiness in the time series in the early parts of the sample. "Other regions" aggregates all continents other than the Americas, Asia, and Europe (e.g., Africa and Oceania).

vations, and reveals shifts in the geographic center of innovation over time. In the early centuries, Asia was a major source of global innovations. Starting around the Scientific Revolution, Europe emerged as the center of innovation, a position it maintained through the Industrial Revolution. Subsequently, the Americas rose to become the leading source of innovations. The right panel shows the corresponding population shares. The contrast is clear: innovation is far more concentrated than population. Europe and the Americas account for a large majority of global innovations in recent centuries, but their share of global population is much smaller.[3]

Figure 2 extends the continent-level innovation shares back to the earliest recorded innovation in our dataset at the beginning of human history (hunting weapon), around 5 million years ago. Because innovations are sparse in the earliest time periods, the horizontal axis is logarithmic in years up to 0 AD and linear in calendar year afterwards like in Figure 1, and we also use longer trailing windows before 0 AD. In addition, this figure separates out Africa given its importance for ancient innovations. Over the very very long run, the data show that Africa was essentially the sole locus of innovation through the Paleolithic—consistent with it being the cradle of humanity—after which the center of innovation moved to the Near East and Asia during the Neolithic and Bronze Age, and then to Europe in Greek and Roman times by the start of Figure 1. Even in the very early days, global innovations appear to be highly concentrated in space. Almost the entire time in human history, a single continent accounts for

[3]Appendix Figure IA4 provides further comparison using Herfindahl-Hirschman indices (HHI). The HHI for innovation is typically higher than that for population.

Figure 2. Innovation Shares across Continents over the Very Long Run



Notes: The figure shows each continent's stacked share of global innovations from the beginning of the technological record (around 5 million years ago) to 1999, with Africa separated as its own region. The horizontal axis is logarithmic in years before present (BP) to the left of 0 AD, and linear in calendar year to the right of 0 AD; the two segments meet at the dashed line. Innovations are measured as the cumulative count over a trailing window of $\max(100, (\text{BP} - 1800)/2)$ years—a 100-year window from the present back to 0 AD, as in Figure 1, widening to about half the elapsed age deep in prehistory, where a fixed 100-year window would be empty. Observations without a locatable region are excluded.

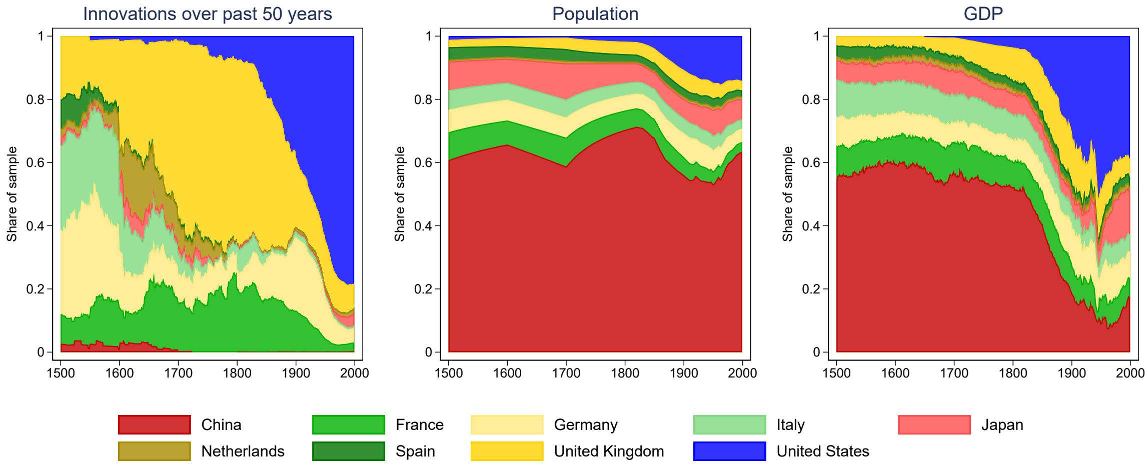
over half of the world's innovations.

Country level Figure 3 presents analogous evidence at the country level for 1500 to 1999, showing the shares of innovation, population, and GDP accounted for by nine major countries (each expressed as a share of the nine-country total). The innovation shares reveal the same kind of shift seen across continents: the United Kingdom was the leading innovator from the Industrial Revolution through the 19th century, with Germany and France also contributing significant shares, after which the United States rose steadily to become the single largest innovating nation by the early 20th century. Comparing across measures, innovation is far more concentrated than either population or GDP. Even among these high-income countries, innovation shares differ markedly from GDP or population shares, indicating that the concentration of innovation is not merely a reflection of economic size.[4]

Figure 4 provides further perspective by showing the number of countries, population share, and GDP share needed to account for 50% of global innovations. Throughout the sample, a remarkably small

[4]Appendix Figure IA5 presents HHI at the country level. The HHI of innovation is again generally higher than the HHI of population. The figure also includes the HHI of GDP (shown with a lighter color before 1800, where historical GDP data from the Maddison Project Database are sparser). Innovation concentration substantially exceeds GDP concentration, confirming that even economic output—which is itself spatially unequal—is far less concentrated than innovative activity.

Figure 3. Innovation Shares across Countries: 1500 to 1999



Notes: The figure shows the shares of innovations, population, and GDP accounted for by nine selected countries (China, France, Germany, Italy, Japan, the Netherlands, Spain, the United Kingdom, and the United States) between 1500 and 1999, each expressed as a share of the nine-country total. Innovations are measured as the cumulative count over the past 50 years, to smooth lumpiness in the early parts of the sample.

number of countries accounts for the majority of global innovations—often just one or two countries suffice to reach the 50% threshold. These countries represent a modest share of global population but a slightly larger share of GDP.

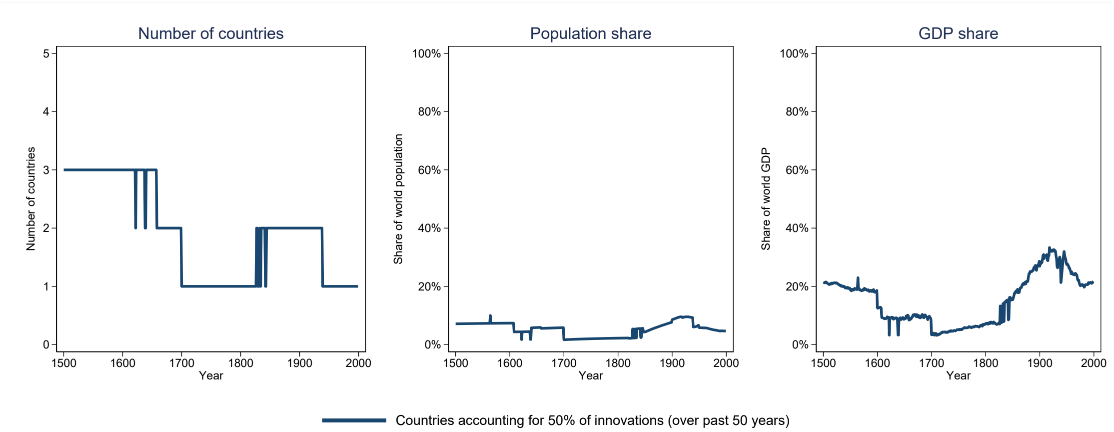
Figure 5 presents another illustration of the dispersion of innovation using the coefficient of variation (CV) across countries since 1500. The CV of innovations per capita hovers around two, which is much higher than the CV of GDP per capita that stays below one. According to this metric, technological innovation is several times more dispersed across countries than output, and this gap remains over the past centuries.

3.1.2 Innovations per Capita over Time

The results above reveal that the total volume of innovation is highly concentrated across space. We now examine innovations per capita and their evolution over time. This measure captures a region's innovation intensity, and it shows interesting patterns about the evolution of innovation intensity.

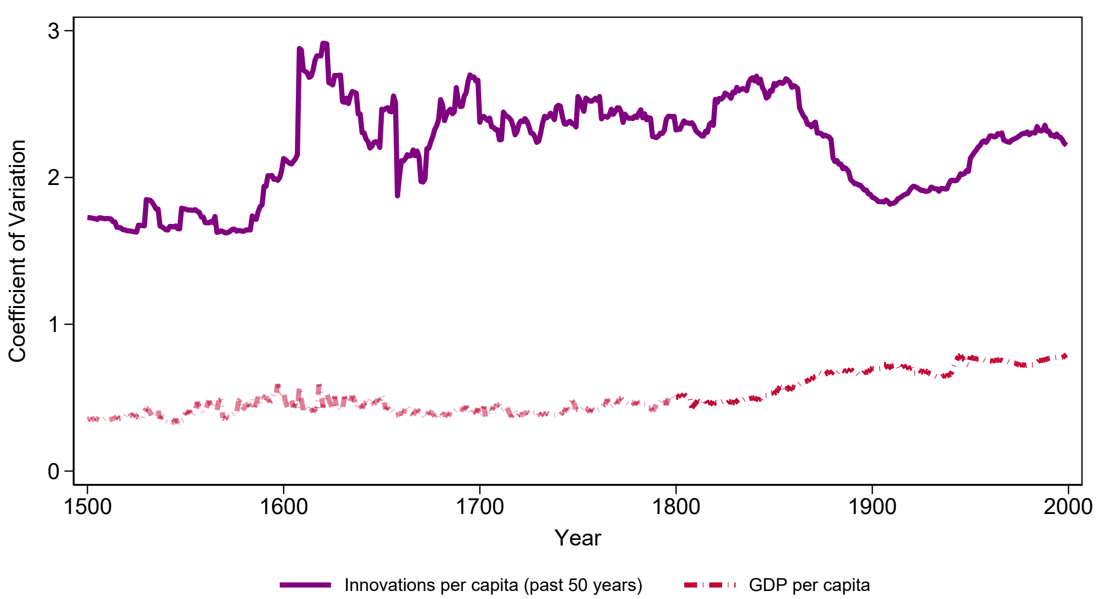
Continent level Figure 6 plots innovations per capita by continent from 0 to 1999. Europe had the highest innovations per capita for centuries, reflecting its role in the Scientific Revolution and the Industrial Revolution. In the 20th century, however, the Americas greatly surpassed Europe, driven by the rapid ascent of the United States.

Figure 4. Number of Countries for Half of World Innovations



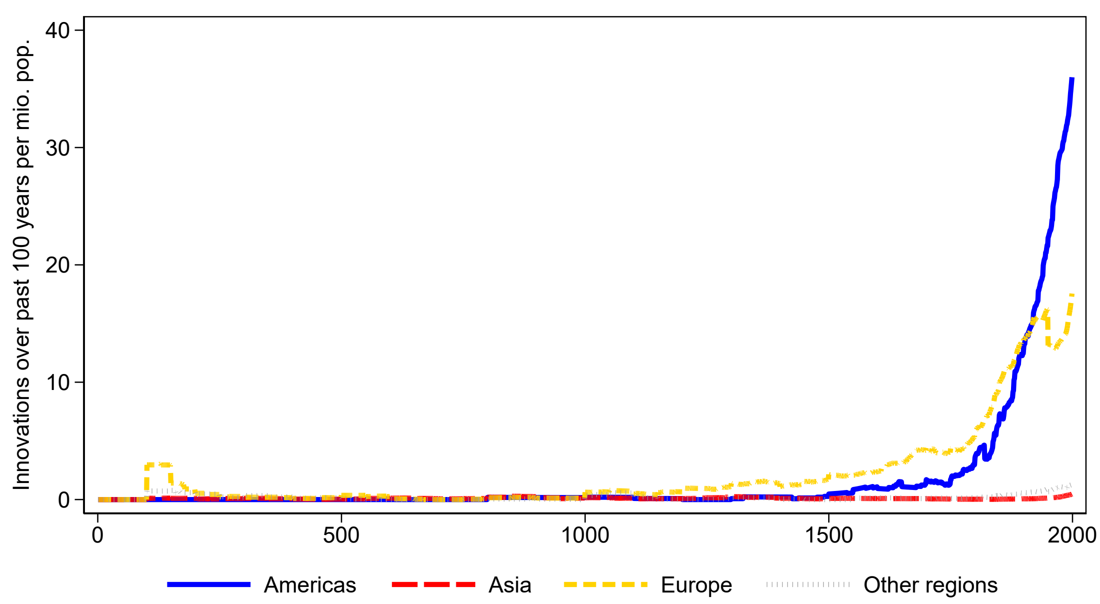
Notes: The figure shows the number of countries (left panel), the share of global population (middle panel), and the share of global GDP (right panel) needed to account for 50% of global innovations over the past 50 years. The sample covers 1500 to 1999. Population and GDP data are from the Maddison Project Database (Bolt and van Zanden, 2024).

Figure 5. Coefficient of Variation: Innovation vs. GDP per Capita



Notes: The figure plots the cross-country coefficient of variation (standard deviation divided by the mean) of innovations per capita (solid line) and of real GDP per capita (dashed line), each year from 1500 to 1999. The sample is the 63 countries with at least 10 decades of non-missing GDP per capita from 1500 to 1999 (the quantitative estimation sample of Section 5). Innovations per capita are measured as the cumulative count over the past 50 years relative to population. GDP per capita is from the Maddison Project Database (Bolt and van Zanden, 2024), shown in a lighter shade before 1800 where the historical data are sparser. A higher value indicates greater dispersion across countries.

Figure 6. Innovations per Capita by Continent

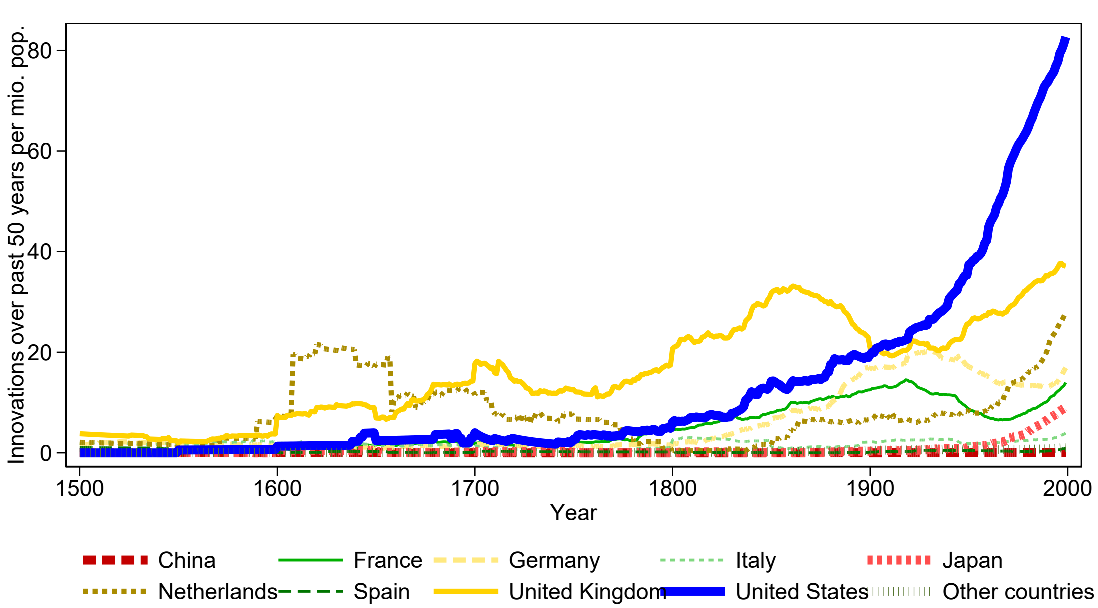


Notes: The figure plots the number of innovations from a given continent relative to that continent's population between 0 and 1999. Innovations are measured as the cumulative count over the past 100 years, to smooth lumpiness in the early parts of the sample. "Other regions" aggregates all continents other than the Americas, Asia, and Europe (e.g., Africa and Oceania).

Country level Figure 7 shows innovations per capita at the country level from 1500 to 1999. Several patterns stand out. First, while innovations per capita trend up over the long run, innovation intensity has stagnated relative to one century ago in many European countries. For the United Kingdom, the decline of innovations per capita started with the Second Industrial Revolution: it had extraordinarily high innovations per capita during and after the First Industrial Revolution, but its innovation intensity has been stagnant since then. For Germany and France, we also observe notable declines of innovation intensity after the Second World War. Second, the United States has experienced a steady increase in innovations per capita since the founding of the country, a trajectory sustained for more than two centuries.

Figure 8 presents the joint evolution of log GDP per capita and innovations per capita for 20 OECD countries, providing a country-by-country view of this relationship. The heterogeneity in innovations per capita across countries is large even among high-income nations. For example, Australia, Canada, as well as many European countries (e.g., Belgium, Greece, and Spain) except Switzerland have achieved high and rising outputs over the past century while producing relatively few innovations per capita on their own. Meanwhile, the United Kingdom and the Netherlands—countries with comparable or higher output levels—had substantially more innovations per capita historically, and the United States

Figure 7. Innovations per Capita by Country

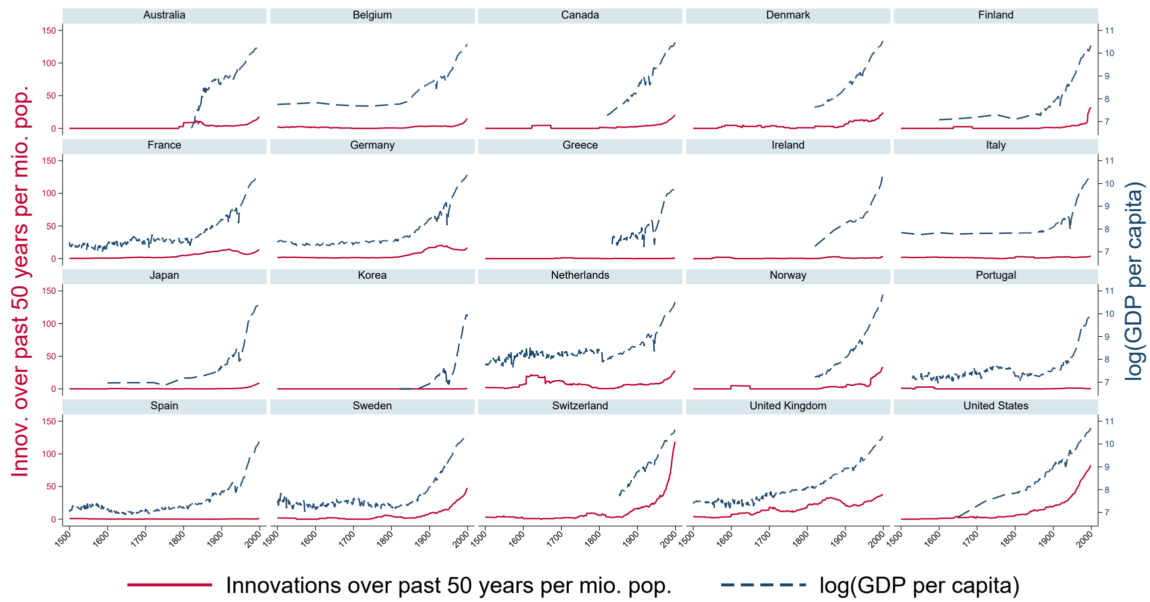


Notes: The figure plots the number of innovations from a given country relative to that country's population between 1500 and 1999. Innovations are measured as the cumulative count over the past 50 years, to smooth lumpiness in the early parts of the sample. Nine countries are shown individually: China, France, Germany, Italy, Japan, the Netherlands, Spain, the United Kingdom, and the United States. "Other countries" shows the population-weighted average of innovations per capita across all remaining countries.

has a particularly high level of innovations per capita. In other words, high output can be attained with either high or low volumes of innovation.

Figure 9 brings these threads together in a single scatterplot. Each dot is a country in a 50-year window: the horizontal axis is innovations per million inhabitants over the window, the vertical axis is the contemporaneous change in log real GDP per capita, and the color reflects the country's output tercile at the start of the window. Labeled black markers identify selected major countries at the window of their peak innovation intensity. The figure shows that countries can achieve a similar level of growth with very different amounts of innovation. Among the labeled major innovators, the United States, the United Kingdom, Germany, France, the Netherlands, Switzerland, and Sweden sit at broadly similar levels of growth despite spanning a wide range of innovation intensities—Switzerland, for instance, has by far the highest innovations per capita yet grows at a rate comparable to far less innovation-intensive economies. Conversely, some of the largest growth episodes belong to countries with very few homegrown innovations: China and Japan over 1950 to 2000 grew by 1.8 and 2.4 log points respectively with only modest innovations per capita, and postwar Italy showed a similar pattern. High growth can come either from inventing at home or from absorbing innovations from elsewhere. The

Figure 8. Innovation and GDP per Capita for Selected OECD Countries

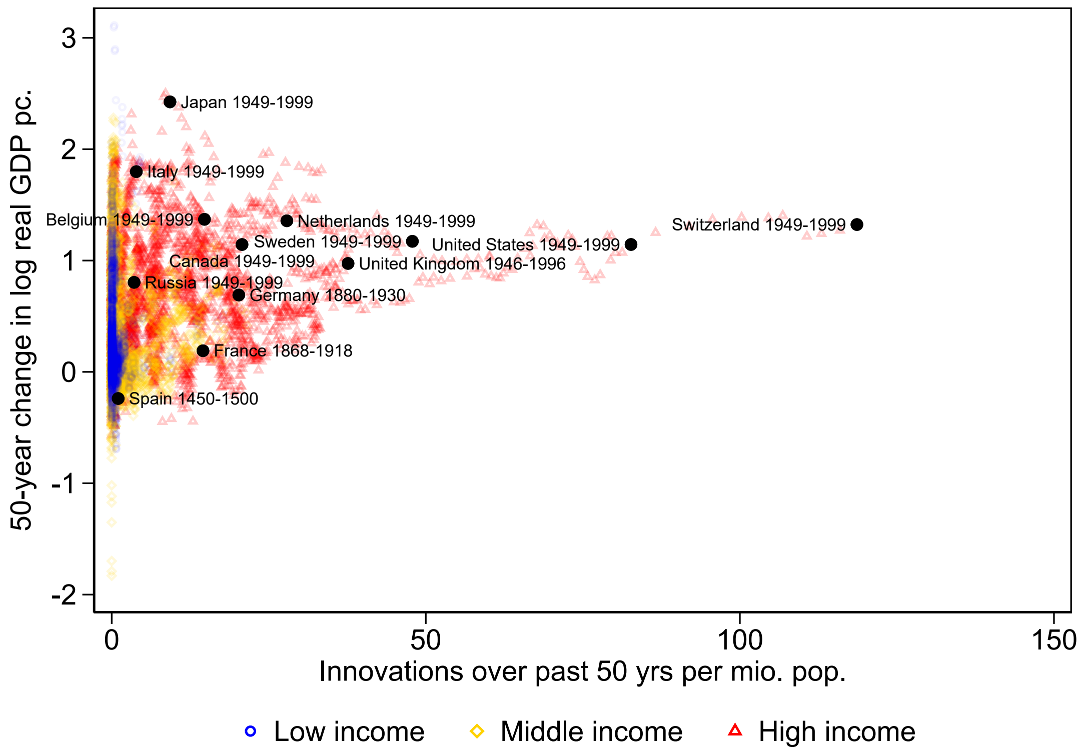


Notes: The figure shows the joint evolution of log real GDP per capita (right axis, navy dashed line) and innovations per million population (left axis, cranberry solid line) for 20 OECD countries over the period 1500 to 1999. Each panel displays one country; each country's series begins in the first year its GDP per capita data are available. GDP per capita data are the raw observations from the Maddison Project Database (Bolt and van Zanden, 2024). Innovations are measured as the cumulative count over the past 50 years relative to population (in millions).

decoupling between innovation and output within individual countries is potentially consistent with technology diffusion allowing countries to benefit from innovations produced elsewhere—a possibility we investigate further in Section 3.2.

In Appendix Table IA1, we explore which observable country characteristics are most closely associated with innovation intensity by regressing innovations per capita on development indicators for 17 advanced economies over 1870 to 1999. These regressions present the correlations among different country characteristics, with all of the usual caveats, and are not meant to demonstrate causation. Consistent with the patterns documented above, population size and real GDP per capita have limited explanatory power for innovation intensity—reinforcing the finding that innovation is not simply proportional to economic size or output. Stock market capitalization relative to GDP (data from Kuvshinov and Zimmermann, 2022) is a consistently positive correlate both across and within countries, while bank lending (data from Jordà, Schularick, and Taylor, 2017) shows no significant association. The stronger role of equity markets relative to bank lending is consistent with the idea that risk-tolerant financial capital—channeled through stock markets rather than bank loans—could be an important input for innovation.

Figure 9. Innovations per Capita and Contemporaneous Growth



Notes: Each marker is one country-year (t) observation over the full 1500 to 1999 sample. The horizontal axis measures innovations per million population over the preceding 50 years; the vertical axis measures the change in log real GDP per capita over the same backward window. Marker color and shape indicate the country's output tercile based on real GDP per capita at year $t - 50$, with terciles defined over the pooled sample. Black markers identify selected highlighted countries—the United Kingdom, France, Germany, the United States, Italy, the Netherlands, Japan, China, Russia, Switzerland, Sweden, Canada, Belgium, and Spain—each shown at the 50-year window of its peak innovation intensity, labeled with the country name and window. GDP and population data are from the Maddison Project Database (Bolt and van Zanden, 2024).

3.2 Innovation and Subsequent Output Growth

The evidence above shows that innovation is far more concentrated than output, and that a country's innovation intensity can diverge substantially from its output level. In this section, we study the relationship between innovation and subsequent growth. First, do innovations matter for a country's subsequent economic performance? Second, do countries benefit from innovations produced *elsewhere*—that is, does technology diffuse across borders? If diffusion is important, it could help explain how countries with modest innovation activity nonetheless achieve high outputs.

We examine these questions using local projections of the form

$$\Delta_h \log(\text{real GDP per capita}_{n,t+h}) = \gamma_{n,h} + \beta_h\, x_{n,t} + \boldsymbol{\psi}_h' \boldsymbol{z}_{n,t} + \epsilon_{n,t,h}, \tag{1}$$

where n indexes the country, t the year, and h the prediction horizon; a separate regression is estimated for each h. In different specifications, the variable of interest $x_{n,t}$ captures either the country's own innovation activity or the innovation activity of the global leader (over the past 10 years, normalized by world innovation stock). The control vector $\boldsymbol{z}_{n,t}$ contains 10 lags and the contemporaneous annual real GDP per capita growth, as well as the level of log real GDP per capita. The sample covers 1500 to 1999, with GDP data from the Maddison Project Database (Bolt and van Zanden, 2024). Standard errors are double clustered by country and year.

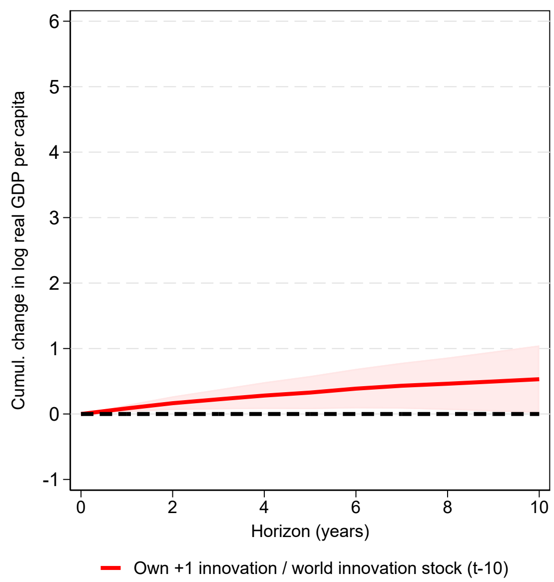
Figure 10 presents results for a country's own innovations. Panel (a) shows that countries with more innovations over the past decade experience higher subsequent growth in real GDP per capita, with the effect building over longer horizons. This relationship appears stronger in the historical period (Panel b) than in the modern one (Panel c). In the modern era, the relationship between own innovation and future local growth attenuates, consistent with the possibility that growth has become more closely linked to the diffusion of innovations elsewhere as we show next.

Figure 11 examines the complementary question: does the global innovation leader's activity predict GDP per capita growth in *other* countries? Here $x_{n,t}$ is replaced by $x_{l,t}$, the innovation leader's innovations over the past decade relative to the world innovation stock, and the control vector also includes the country's own innovation activity. The leader is defined as the country with the most total innovations over the past decade; the sample excludes the leader itself. We find that the leader's innovations positively predict growth elsewhere (Panel a), and the relationship is stronger in the modern era (Panel c) than before (Panel b). These results suggest that the economic gains from innovation have become less tied to their place of origin over time. Early on, growth follows more from local innovation relative to leader's innovation. In the modern period, growth follows less from local innovation and more from leader's innovation.

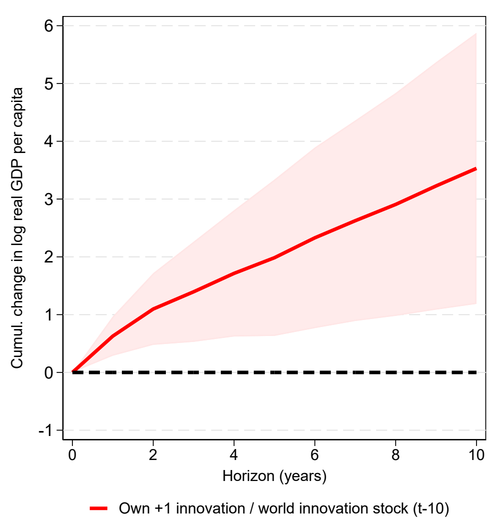
Does diffusion benefit all countries equally? In Appendix Figure IA6, we rank countries by their real GDP per capita in each year ("poorest" countries are those in the bottom GDP-per-capita quartile in year t, "richest" countries are those in the top quartile, and the middle two quartiles serve as the omitted reference), and show the relationship between leader innovation and subsequent growth using the same specification as Figure 11, now with the leader's innovation interacting with the poorest and richest country indicator variables. We observe that the response of GDP per capita to the leader's innovations

Figure 10. Own Innovations and Future Local Growth

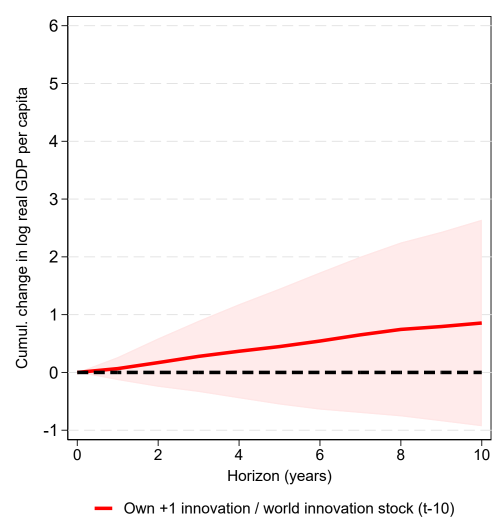
(a) Full Sample



(b) Sample 1500 to 1899

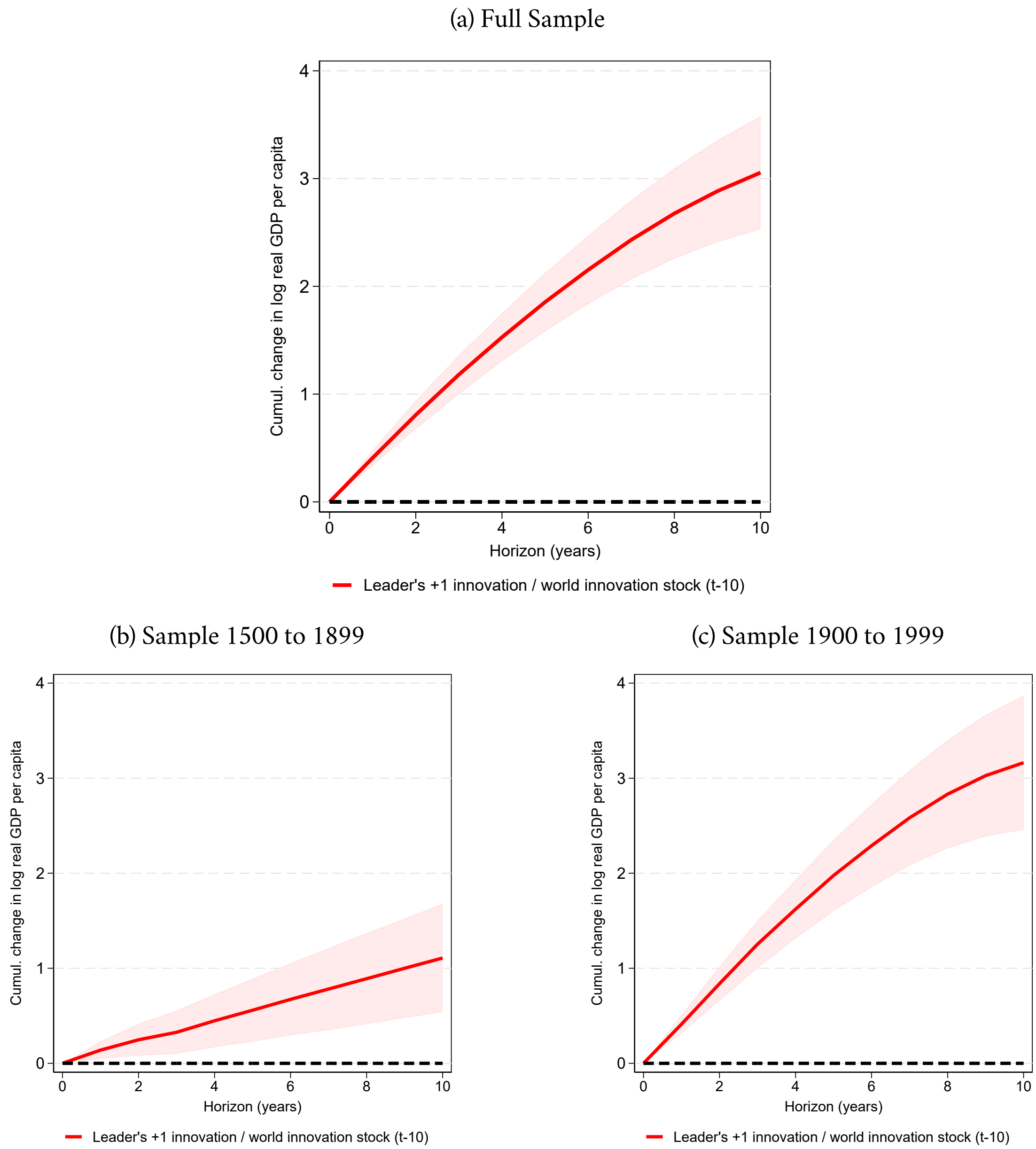


(c) Sample 1900 to 1999



Notes: The figure shows changes in a country's future real GDP per capita following increases in its own innovation activity, estimated using local projections: $\Delta_h \log(\text{real GDP per capita}_{n,t+h}) = \gamma_{n,h} + \beta_h x_{n,t} + \boldsymbol{\psi}_h' \boldsymbol{z}_{n,t} + \epsilon_{n,t,h}$, where n indexes the country and h the prediction horizon; a separate regression is estimated for each h. The outcome is the cumulative change in log real GDP per capita from year t to $t+h$. $x_{n,t}$ is the country's total innovations over the past 10 years, relative to the cumulative world innovation stock 10 years earlier. The control vector $\boldsymbol{z}_{n,t}$ contains 10 lags and the contemporaneous annual real GDP per capita growth, as well as the level of log real GDP per capita. Panel (a) shows estimates for the full sample (1500 to 1999); Panel (b) for the historical period (1500 to 1899); Panel (c) for the modern period (1900 to 1999). GDP data are from the Maddison Project Database (Bolt and van Zanden, 2024). Shaded areas mark 90% confidence intervals based on standard errors double clustered by country and year.

Figure 11. Leader's Innovations and Future Local Growth



Notes: The figure shows changes in a country's future real GDP per capita following increases in innovation activity of the global innovation leader. Estimates come from local projections: $\Delta_h \log(\text{real GDP per capita}_{n,t+h}) = \gamma_{n,h} + \beta_h x_{l,t} + \boldsymbol{\psi}_h' \boldsymbol{z}_{n,t} + \epsilon_{n,t,h}$, where n indexes the country and h the prediction horizon; a separate regression is estimated for each h. The outcome is the cumulative change in log real GDP per capita from year t to $t+h$. $x_{l,t}$ is the global innovation leader's total innovations over the past 10 years, relative to the cumulative world innovation stock 10 years earlier (the leader is defined as the country with the most total innovations over the past 10 years). The control vector includes the country's own innovations over the past 10 years relative to the cumulative world innovation stock 10 years earlier, 10 lags and the contemporaneous annual real GDP per capita growth, and the level of log real GDP per capita. The sample excludes the leader country. Panel (a) shows estimates for the full sample (1500 to 1999); Panel (b) for the historical period (1500 to 1899); Panel (c) for the modern period (1900 to 1999). GDP data are from the Maddison Project Database (Bolt and van Zanden, 2024). Shaded areas mark 90% confidence intervals based on standard errors double clustered by country and year.

is smaller for the poorest countries than for the richest ones, and the pattern is robust to controlling for log geographic distance to the leader interacted with leader innovation, so it is not simply a matter of proximity. The framework in Section 4 formalizes this gradient: countries with lower frictions in adopting foreign technology are both richer in levels and more responsive to innovations made abroad.

Appendix Section IA5.6 presents the analogous results where the outcome is future innovation activity rather than GDP growth, showing that innovation is strongly self-reinforcing in both periods, and that the leader's innovation activity has a modest positive association with innovation elsewhere.

4 A Framework with Innovation and Diffusion

We develop a framework that combines three ingredients: research that generates new techniques, technology diffusion that spreads the benefits of innovation across locations, and occupational choice, which amplifies local differences in research, adoption, and production incentives. The framework builds on the multi-location knowledge-diffusion models of Eaton and Kortum (1999) and Buera and Oberfield (2020), adding a distinction between innovation and adoption. The model exhibits semi-endogenous growth, but we deliberately abstract from scale effects that differ across locations.[5] We then focus on a small-location limit, in which the model collapses to two simple equations—one for output per capita and one for innovations per capita—that we use to derive analytical results and that are amenable to estimation.

Environment The world consists of a finite set of countries indexed by n. Each country is a collection of locations $\mathcal{L}_{nt}$, indexed by ℓ. Every location is endowed with the same measure of individuals $L_{\ell t} \equiv \bar{L}(1+\gamma)^t$, so the population of country n at time t is

$$L_{nt} \;\equiv\; \bar{L}\,|\mathcal{L}_{nt}|\,(1+\gamma)^t, \tag{2}$$

where $|\mathcal{L}_{nt}|$ denotes the number of locations in country n. Per-location population grows at the common rate γ, while the number of locations $|\mathcal{L}_{nt}|$ varies exogenously across countries to capture differences in geographic extent and population growth beyond γ. Individuals at each location choose among three activities: *research* ($L^R_{\ell t}$), which creates new techniques; *adoption* ($L^A_{\ell t}$), which commercializes

[5] This choice reflects the fact that country population is an imperfect proxy for market access. Many firms in large countries serve markets much smaller than the country, while many firms in small countries serve markets extending beyond national borders. We therefore avoid building country-size scale effects directly into incentives for research or adoption. Even without this, an indirect scale effect remains because we will assume below that ideas diffuse more quickly within countries than across countries, so those in large countries can access more ideas quickly.

existing techniques into varieties for production; and *production* ($L^Y_{\ell t}$), which manufactures output. We let $s^j_{\ell t} \equiv \frac{L^j_{\ell t}}{L_{\ell t}}$ be the share of labor in location ℓ in activity $j \in \{R, A, Y\}$; these labor shares sum to one at every location:

$$s^R_{\ell t} + s^A_{\ell t} + s^Y_{\ell t} = 1. \tag{3}$$

Within a country, fundamentals—research productivity, adoption productivity, intellectual property protection, and taste shocks—are common across locations.

Research There is a unit continuum of technological domains. Research generates new techniques, each assigned to a random domain and characterized by a productivity level q. We use the terms "idea" and "technique" interchangeably for these domain-specific productivity draws; we reserve the term "innovation" to refer to the subset of new ideas that are sufficiently important to be recorded in our dataset, as discussed below. When a new technique arrives, its productivity combines an original component o with an *insight* q' drawn from a source distribution $R_{\ell t}(q')$, according to $q = (q')^\beta \cdot o$. The parameter $\beta \in [0, 1)$ governs the strength of diffusion: it measures how much insights from existing ideas contribute to the productivity of new ideas. Ideas with original component greater than o arrive at rate

$$L^R_{\ell t} Z^R_{\ell t} \cdot o^{-\theta} .$$

We assume the flow of new ideas at location ℓ scales with the research workforce $L^R_{\ell t}$, and $Z^R_{\ell t}$ captures local research productivity, such as institutional quality and scientific infrastructure.

Let $F_{\ell t}(q)$ be the fraction of domains for which the best technique discovered in ℓ has productivity weakly less than q. As shown by Buera and Oberfield (2020), if the initial frontier is Fréchet, it remains Fréchet over time, $F_{\ell,t} = e^{-\lambda_{\ell t} q^{-\theta}}$, where $\lambda_{\ell t}$ is the local stock of effective ideas, or more simply, the local *stock of ideas*—a summary of the history of local research—and $\theta > 0$ governs the dispersion of technique productivities (we require $\theta > \varepsilon - 1$ so that the relevant moments are finite). A higher $\lambda_{\ell t}$ shifts the entire productivity distribution upward; a lower θ increases dispersion, giving frontier techniques a larger advantage over the rest. The local stock of ideas evolves according to:

$$\underbrace{\lambda_{\ell t} - \lambda_{\ell,t-1}}_{x_{\ell t}} = L^R_{\ell t} Z^R_{\ell t} \int_0^\infty x^{\beta\theta} \, dR_{\ell t}(x). \tag{4}$$

The left-hand side, $x_{\ell t} \equiv \lambda_{\ell t} - \lambda_{\ell,t-1}$, is the flow of new ideas in location ℓ. $x_{\ell t}$ is closely related to what we measure as "innovation" in the data, a relationship we will later make precise. The right-hand side shows that it depends on local research effort, research productivity, and the pool of insights available to local researchers.

We model the source distribution $R_{\ell t}$ as reflecting insights drawn from the global stock of ideas. Following Eaton and Kortum (1999), we assume ideas developed in a location diffuse stochastically to other locations, governed by a bilateral diffusion probability $\varphi^R_{\ell\tilde{\ell}} \in (0,1]$. The probability that an idea invented k periods ago in location $\tilde{\ell}$ is known this period in location ℓ is $1-(1-\varphi^R_{\ell\tilde{\ell}})^k$. Thus ideas invented in the long past are almost certainly available everywhere, whereas ideas invented recently are less likely to be available. Aggregating the independent flows of known ideas across source locations and vintages, the source distribution is Fréchet:

$$R_{\ell t}(q) = e^{-\mathcal{R}_{\ell t} q^{-\theta}}, \tag{5}$$

where we define the *insight pool*:

$$\mathcal{R}_{\ell t} \equiv \sum_{\tilde{\ell}} \sum_{k=1}^{\infty} \left[1-(1-\varphi^R_{\ell\tilde{\ell}})^k\right] x_{\tilde{\ell},t-k}.$$

The diffusion probability parameter $\varphi^R_{\ell\tilde{\ell}}$ captures frictions in the flow of ideas for research purposes. $\varphi^R_{\ell\tilde{\ell}} = 1$ for any two locations $\ell, \tilde{\ell}$ within the same country. The diffusion probability $\varphi^R_{\ell\tilde{\ell}}$ may be lower for more distant locations, reflecting greater barriers to the flow of insights. In the general model, φ^R (and similarly φ^A below) can be an arbitrary bilateral matrix. Substituting into (4), the law of motion for the stock of ideas becomes:

$$x_{\ell t} = L^R_{\ell t} Z^R_{\ell t} \Gamma(1-\beta) \cdot \underbrace{\mathcal{R}^{\beta}_{\ell t}}_{\text{insight spillovers}} = s^R_{\ell t} L_{\ell t} Z^R_{\ell t} \Gamma(1-\beta) \mathcal{R}^{\beta}_{\ell t}. \tag{6}$$

Technology adoption and output Separately from research, *adopters* (entrepreneurs) commercialize existing ideas for production. An adopter in ℓ pays cost $1/Z^A_{\ell t}$ units of time to create a new variety, drawing a random domain and selecting the best locally available idea for that domain. Ideas become locally available for adoption following a process analogous to that of research. Ideas diffuse stochastically across locations with probability $\varphi^A_{\ell\tilde{\ell}}$. An idea invented k periods ago in location $\tilde{\ell}$ is available for adoption this period in location ℓ with probability $1-(1-\varphi^A_{\ell\tilde{\ell}})^k$. In addition, we assume that ideas developed locally are available for adoption immediately within the period. The distribution of the best idea available to adopters in ℓ is therefore:

$$A_{\ell t}(q) = e^{-\mathcal{A}_{\ell t} q^{-\theta}},$$

where we define the *technology pool*:

$$\mathcal{A}_{\ell t} \;\equiv\; x_{\ell t} + \sum_{\tilde{\ell}} \sum_{k=1}^{\infty} \Big[1 - (1 - \varphi^{A}_{\ell\tilde{\ell}})^{k}\Big]\, x_{\tilde{\ell},t-k}\,. \tag{7}$$

The measure of varieties created in location ℓ is $N_{\ell t} = Z^{A}_{\ell t}\, L^{A}_{\ell t}$, and we assume that the ability to produce a variety depreciates after a single period.

Each individual has Dixit-Stiglitz preferences across all varieties, with elasticity of substitution $\varepsilon > 1$, and each producer in ℓ operates a technology with location-level productivity shifter $Z^{Y}_{\ell t}$ (capturing infrastructure and other aggregate productivity not captured by the model). Aggregating, total output in location ℓ is $N_{\ell t}^{1/(\varepsilon-1)} Q_{\ell t}\, Z^{Y}_{\ell t}\, L^{Y}_{\ell t}$, where $Q_{\ell t} \equiv \Gamma\Big(1 - \frac{\varepsilon-1}{\theta}\Big)^{1/(\varepsilon-1)} \mathcal{A}_{\ell t}^{1/\theta}$ is an index of available technology quality. Per-capita output at the location is therefore

$$y_{\ell t} = N_{\ell t}^{\frac{1}{\varepsilon-1}}\, Q_{\ell t}\, s^{Y}_{\ell t}\, Z^{Y}_{\ell t}\,. \tag{8}$$

Substituting $N_{\ell t} = Z^{A}_{\ell t}\, L^{A}_{\ell t}$, location-level output per capita is

$$y_{\ell t} = s^{Y}_{\ell t}\, Z^{Y}_{\ell t}\, \Big(Z^{A}_{\ell t}\, L^{A}_{\ell t}\Big)^{\frac{1}{\varepsilon-1}} \Gamma\Big(1 - \frac{\varepsilon-1}{\theta}\Big)^{\frac{1}{\varepsilon-1}} \mathcal{A}_{\ell t}^{1/\theta}\,. \tag{9}$$

Market structure Let $w^{R}_{\ell t}$, $w^{A}_{\ell t}$, and $w^{Y}_{\ell t}$ denote the wages in research, adoption, and production in location ℓ. Adopters engage in monopolistic competition. An adopter with productivity $q_{j\ell t}$ faces demand $Y_{\ell t} P^{\varepsilon}_{\ell t} p^{-\varepsilon}_{j\ell t}$ and unit cost $w^{Y}_{\ell t}/q_{j\ell t}$, leading to a price $p_{j\ell t} = \frac{\varepsilon}{\varepsilon-1} w^{Y}_{\ell t}/q_{j\ell t}$. The price index is $P_{\ell t} = N_{\ell t}^{1/(1-\varepsilon)} \frac{\varepsilon}{\varepsilon-1} w^{Y}_{\ell t}/Q_{\ell t}$, total expenditure on production labor satisfies $Y_{\ell t} P_{\ell t} = \frac{\varepsilon}{\varepsilon-1} w^{Y}_{\ell t}\, L^{Y}_{\ell t}$, and the expected profit per variety (before any royalty payments) is:

$$\frac{1}{N_{\ell t}} Y_{\ell t} P_{\ell t} \frac{1}{\varepsilon}.$$

Adopters may pay royalties to the researchers who invented the ideas they use. Let $T_{\ell t}$ denote total royalty payments from adopters in ℓ to researchers at time t. Free entry into adoption requires that entry costs exhaust expected profits net of royalties:

$$\frac{w^{A}_{\ell t}}{Z^{A}_{\ell t}} N_{\ell t} = \frac{1}{\varepsilon - 1}\, w^{Y}_{\ell t}\, L^{Y}_{\ell t} - T_{\ell t}\,.$$

Using the market-clearing condition $N_{\ell t} = Z^A_{\ell t} L^A_{\ell t}$, this simplifies to:

$$w^A_{\ell t} L^A_{\ell t} = \frac{1}{\varepsilon - 1} w^Y_{\ell t} L^Y_{\ell t} - T_{\ell t}. \tag{10}$$

We assume that researchers earn royalty payments only from adopters who use their ideas *within a period*. As such, the length of a period will be a calibration target. We have already assumed that ideas are available within the period only in a researcher's own location, so she earns royalties from local entrepreneurs. We further assume that the adopter keeps a share $1 - \eta_{\ell t}$ of the profit margin between the best and second-best locally available ideas for that domain, and pays the remainder $\eta_{\ell t}$ to the researcher. Ideas invented before the current period earn no royalties.

Building on Bernard et al. (2003), we obtain the following result (proved in Appendix IA7).

$$T_{\ell t} = \eta_{\ell t} \frac{1}{\theta} w^Y_{\ell t} L^Y_{\ell t} \frac{x_{\ell t}}{\mathcal{A}_{\ell t}}. \tag{11}$$

The royalty formula has an intuitive structure. A researcher in ℓ earns royalties only when her new idea is the best available to local adopters. The term $x_{\ell t}/\mathcal{A}_{\ell t}$ is the probability that a new idea from ℓ beats all alternatives in a random domain—it equals the flow of local ideas relative to the total technology pool available to adopters in ℓ. When the pool $\mathcal{A}_{\ell t}$ is large, any individual researcher's idea is less likely to be the frontier idea, reducing expected royalties.

Free entry into research implies that all royalty payments flow to researchers:

$$w^R_{\ell t} L^R_{\ell t} = T_{\ell t}. \tag{12}$$

Equations (10) and (12) show that total profits from monopolistic competition, $\frac{1}{\varepsilon-1} w^Y_{\ell t} L^Y_{\ell t}$, are split between adopters and researchers, with the division depending on royalty payments.

Labor supply Each individual chooses an activity $a \in \{R, A, Y\}$ to maximize:

$$\max_a \; \log \frac{w^a_{\ell t}}{P_{\ell t}} + \log B^a_{\ell t} + \frac{1}{\sigma} \varepsilon^a_i,$$

where $B^a_{\ell t}$ is a systematic component of the attractiveness of activity a in location ℓ (capturing cultural factors, norms, or non-pecuniary benefits), and ε^a_i is an idiosyncratic taste shock drawn IID from a standard Gumbel distribution. The parameter $\sigma > 0$ governs the elasticity of labor supply across activities: a higher σ means individuals are more responsive to wage differentials.

Under the Gumbel assumption, the shares of the population choosing activity a obey:

$$\frac{s^R_{\ell t}}{s^Y_{\ell t}} = \left(\frac{w^R_{\ell t}\, B^R_{\ell t}}{w^Y_{\ell t}\, B^Y_{\ell t}}\right)^{\sigma}, \qquad \frac{s^A_{\ell t}}{s^Y_{\ell t}} = \left(\frac{w^A_{\ell t}\, B^A_{\ell t}}{w^Y_{\ell t}\, B^Y_{\ell t}}\right)^{\sigma}. \tag{13}$$

We now combine labor demand and supply to characterize the equilibrium allocation of labor to research. Royalty payments in location ℓ can be expressed as:

$$T_{\ell t} = \phi_{\ell t}\, w^Y_{\ell t}\, L^Y_{\ell t},$$

where

$$\phi_{\ell t} \equiv \eta_{\ell t}\, \frac{1}{\theta}\, \frac{x_{\ell t}}{\mathcal{A}_{\ell t}} \tag{14}$$

is royalty payments as a multiple of the production wage bill. Total earnings in each activity are then:

$$w^R_{\ell t} L^R_{\ell t} = \phi_{\ell t}\, w^Y_{\ell t}\, L^Y_{\ell t}, \tag{15}$$

$$w^A_{\ell t} L^A_{\ell t} = \left(\frac{1}{\varepsilon - 1} - \phi_{\ell t}\right)\, w^Y_{\ell t}\, L^Y_{\ell t}. \tag{16}$$

Combining these earnings expressions with the labor supply equations (13) and the labor-clearing condition (3), we can solve for the equilibrium labor shares:

$$s^R_{\ell t} = \left[\frac{\phi_{\ell t}\, B^R_{\ell t}}{\bar{B}_{\ell t}}\right]^{\frac{\sigma}{1+\sigma}}; \qquad s^A_{\ell t} = \left[\frac{\left(\frac{1}{\varepsilon-1} - \phi_{\ell t}\right)\, B^A_{\ell t}}{\bar{B}_{\ell t}}\right]^{\frac{\sigma}{1+\sigma}}; \qquad s^Y_{\ell t} = \left[\frac{B^Y_{\ell t}}{\bar{B}_{\ell t}}\right]^{\frac{\sigma}{1+\sigma}}; \tag{17}$$

with

$$\bar{B}_{\ell t} \equiv \left\{\left[\phi_{\ell t}\, B^R_{\ell t}\right]^{\frac{\sigma}{1+\sigma}} + \left[\left(\frac{1}{\varepsilon - 1} - \phi_{\ell t}\right)\, B^A_{\ell t}\right]^{\frac{\sigma}{1+\sigma}} + \left[B^Y_{\ell t}\right]^{\frac{\sigma}{1+\sigma}}\right\}^{\frac{1+\sigma}{\sigma}}.$$

The equilibrium research share $s^R_{\ell t}$ is strictly increasing in the royalty share $\phi_{\ell t}$ and the appeal of research $B^R_{\ell t}$, and is strictly decreasing in $B^A_{\ell t}$ and $B^Y_{\ell t}$.

The object $\phi_{\ell t}$ is the endogenous link between the global distribution of knowledge and local research incentives. It depends on three components: intellectual property protection $\eta_{\ell t}$, the ratio $1/\theta$ scaling the expected royalty per new idea, and the ratio $x_{\ell t}/\mathcal{A}_{\ell t}$, which equals the flow of local ideas relative to the total technology pool available to local adopters. When the global pool $\mathcal{A}_{\ell t}$ is large relative to local innovations, a researcher's inventions are unlikely to be the frontier technique in any given domain, reducing expected royalties and discouraging research.

Substituting the equilibrium research share into the law of motion (6) and rearranging gives the implicit equation

$$x_{\ell t} = \left[\frac{\eta_{\ell t}}{\theta}\frac{\mathcal{B}^R_{\ell t}}{\mathcal{A}_{\ell t}}\right]^{\sigma} \left[Z^R_{\ell t}\, L_{\ell t}\, \Gamma(1-\beta)\, \mathcal{R}^{\beta}_{\ell t}\right]^{1+\sigma}, \tag{18}$$

where we define $\mathcal{B}^R_{\ell t} \equiv B^R_{\ell t}/\bar{B}_{\ell t}$.[6] Similarly, substituting the adoption and production shares into the expression for output per capita (9) gives

$$y_{\ell t} = \Gamma\Big(1 - \tfrac{\varepsilon-1}{\theta}\Big)^{\frac{1}{\varepsilon-1}} \left[\mathcal{B}^{AY}_{\ell t}\right]^{\frac{\sigma}{1+\sigma}} \left(Z^A_{\ell t}\, L_{\ell t}\right)^{\frac{1}{\varepsilon-1}} Z^Y_{\ell t}\mathcal{A}^{1/\theta}_{\ell t}, \tag{19}$$

where we define $\mathcal{B}^{AY}_{\ell t} \equiv \left(\frac{\left(\frac{1}{\varepsilon-1}-\phi_{\ell t}\right)B^A_{\ell t}}{\bar{B}_{\ell t}}\right)^{\frac{1}{\varepsilon-1}} \frac{B^Y_{\ell t}}{\bar{B}_{\ell t}}$.

A small-location limit We now specialize the model by considering a small-location limit in which $\bar{L}$ is small. This limit preserves the economics of our baseline model but is easier to work with and analyze, and it delivers two equations that we take to the data.

Formally, we define variables $z^R_{\ell t}$, $z^A_{\ell t}$, and $z^Y_{\ell t}$ as normalized levels of productivity for each activity. We normalize $Z^R_{\ell t} = z^R_{\ell t}/\bar{L}^{\frac{\sigma}{1+\sigma}}$ and $Z^A_{\ell t} = z^A_{\ell t}/\bar{L}$ (along with $Z^Y_{\ell t} = z^Y_{\ell t}$, for symmetry). We study the limiting economy as $\bar{L}$ becomes small, holding fixed $z^R_{\ell t}$ and $z^A_{\ell t}$, as well as overall population. That is, the measure $d\ell$ is normalized so that, as $\bar{L} \to 0$, each country contains more locations while its aggregate population path L_{nt} is held fixed. The limit captures an economy in which local research incentives still determine entry into research, but any single location's flow of new ideas is small relative to the set of ideas available to its adopters. The normalizations ensure that research levels and per-capita output have well-defined limiting behavior.

Recall that population in a location follows $L_{\ell t} = \bar{L}(1+\gamma)^t$. In the limit as $\bar{L}$ approaches zero, (18) and (19) become

$$\frac{x_{\ell t}}{\bar{L}} \to \left[\frac{\eta_{\ell t}}{\theta}\frac{\mathcal{B}^R_{\ell t}}{\mathcal{A}_{\ell t}}\right]^{\sigma} \left[z^R_{\ell t}\,(1+\gamma)^t\, \Gamma(1-\beta)\, \mathcal{R}^{\beta}_{\ell t}\right]^{1+\sigma}, \tag{20}$$

$$y_{\ell t} \to \Gamma\Big(1 - \tfrac{\varepsilon-1}{\theta}\Big)^{\frac{1}{\varepsilon-1}} \left[\mathcal{B}^{AY}_{\ell t}\right]^{\frac{\sigma}{1+\sigma}} \left(z^A_{\ell t}\,(1+\gamma)^t\right)^{\frac{1}{\varepsilon-1}} z^Y_{\ell t}\mathcal{A}^{1/\theta}_{\ell t}. \tag{21}$$

In this limit the royalty share $\phi_{\ell t}$ becomes small: a single location's new ideas are negligible relative to

[6] This equation is implicit because $\mathcal{B}^R_{\ell t}$ depends on $\phi_{\ell t}$, and hence on $x_{\ell t}$. In the small-location limit below, the corresponding limiting equation becomes explicit.

the pool of ideas its adopters can draw on, so from (14), $\phi_{\ell t} = \frac{\eta_{\ell t}}{\theta} \frac{x_{\ell t}}{\mathcal{A}_{\ell t}} \to 0$. The spillover pools become

$$\mathcal{R}_{\ell t} \to \int \sum_{k=1}^{\infty} \left[1 - (1 - \varphi^R_{\ell \tilde{\ell}})^k\right] x_{\tilde{\ell}, t-k} d\tilde{\ell} \tag{22}$$

$$\mathcal{A}_{\ell t} \to \int \sum_{k=1}^{\infty} \left[1 - (1 - \varphi^A_{\ell \tilde{\ell}})^k\right] x_{\tilde{\ell}, t-k} d\tilde{\ell}, \tag{23}$$

and preference terms become

$$\mathcal{B}^R_{\ell t} \to \frac{B^R_{\ell t}}{\left\{\left[\frac{1}{\varepsilon-1} B^A_{\ell t}\right]^{\frac{\sigma}{1+\sigma}} + \left[B^Y_{\ell t}\right]^{\frac{\sigma}{1+\sigma}}\right\}^{\frac{1+\sigma}{\sigma}}} \tag{24}$$

$$\mathcal{B}^{AY}_{\ell t} \to \frac{\left(\frac{1}{\varepsilon-1} B^A_{\ell t}\right)^{\frac{1}{\varepsilon-1}} B^Y_{\ell t}}{\left\{\left[\frac{1}{\varepsilon-1} B^A_{\ell t}\right]^{\frac{\sigma}{1+\sigma}} + \left[B^Y_{\ell t}\right]^{\frac{\sigma}{1+\sigma}}\right\}^{\frac{1+\sigma}{\sigma} \frac{\varepsilon}{\varepsilon-1}}}. \tag{25}$$

The research share of workers shrinks to zero, but the normalization of research productivity ensures that the limiting flow of new ideas remains positive and elastic with respect to preferences for and incentives to conduct research.

Aggregation to countries The empirical analysis is conducted at the country level, so we now aggregate the location-level objects to countries. Within each country, locations share common fundamentals and ideas fully diffuse across locations after one period. Abusing notation, define country n's flow of new ideas as $x_{nt} = \int_{\ell \in \mathcal{L}_{nt}} x_{\ell t} d\ell$. Then the two key equations are

$$x_{nt} = \tilde{\alpha}^R_t L_{nt} \left[\frac{\eta_{nt} \mathcal{B}^R_{nt}}{\mathcal{A}_{nt}}\right]^{\sigma} \left[z^R_{nt} \mathcal{R}^{\beta}_{nt}\right]^{1+\sigma} \tag{26}$$

$$y_{nt} = \tilde{\alpha}^y_t \left[\mathcal{B}^{AY}_{nt}\right]^{\frac{\sigma}{1+\sigma}} \left(z^A_{nt}\right)^{\frac{1}{\varepsilon-1}} z^Y_{nt} \mathcal{A}^{1/\theta}_{nt}, \tag{27}$$

where we defined $\tilde{\alpha}^R_t \equiv \theta^{-\sigma} \Gamma(1-\beta)^{1+\sigma} (1+\gamma)^{\sigma t}$ and $\tilde{\alpha}^y_t \equiv \Gamma\left(1 - \frac{\varepsilon-1}{\theta}\right)^{\frac{1}{\varepsilon-1}} (1+\gamma)^{\frac{t}{\varepsilon-1}}$, and the spillover pools are

$$\mathcal{R}_{nt} = \sum_{\tilde{n}} \sum_{k=1}^{\infty} \left[1 - (1 - \varphi^R_{n\tilde{n}})^k\right] x_{\tilde{n}, t-k} \tag{28}$$

$$\mathcal{A}_{nt} = \sum_{\tilde{n}} \sum_{k=1}^{\infty} \left[1 - (1 - \varphi^A_{n\tilde{n}})^k\right] x_{\tilde{n}, t-k}. \tag{29}$$

All locations $\ell \in \mathcal{L}_{nt}$ are symmetric, so variation in population is absorbed by changes in the measure of locations over time. Country-level innovation scales linearly with the country's population. Population matters at the country level only by determining how many independent research efforts a country contains, not by inflating the marginal product of each individual researcher. The size of a country matters indirectly, as ideas diffuse more quickly within a country than across countries, affecting the magnitudes of $\mathcal{R}_{nt}$ and $\mathcal{A}_{nt}$.

The forces driving the flow of new ideas at each location are:

1. *Research productivity* $z^R_{\ell t}$: institutional factors affecting the productivity of researchers.
2. *Research attractiveness* $\mathcal{B}^R_{\ell t}$: cultural or normative factors that draw people into research (operating through s^R).
3. *Spillovers to Research* $\mathcal{R}_{\ell t}$: the pool of insights from other locations, governing how much a location benefits from the global stock of ideas.
4. *IP protection* $\eta_{\ell t}$: the share of adopter profit flowing back to researchers (operating through ϕ and then s^R).
5. *Discouragement*: the ratio $x_{\ell t}/\mathcal{A}_{\ell t}$ within ϕ, which reduces research incentives when other countries are far ahead, because new local ideas are less likely to be the best available ideas.

Foreign ideas thus have two opposing effects on local research incentives. By increasing $\mathcal{R}_{\ell t}$, they improve the insights that local researchers can draw on. However, by increasing $\mathcal{A}_{\ell t}$, they dilute local researchers' expected royalties by making foreign ideas more likely to be the best techniques available to local adopters. The evidence presented in Figure IA8 suggests that, in the data, the research productivity spillover dominates, as foreign research raises the probability of local innovation.

A balanced growth path Consider an economy in which the measure of locations in each country is fixed over time, so that population of each country grows at rate γ. If all fundamentals are constant over time, then the economy features a balanced growth path (BGP) in which stocks of ideas and per capita output grow at constant rates. Along this balanced growth path, stocks of ideas and spillover pools grow at a common, constant rate:

$$\frac{\lambda_{n,t}}{\lambda_{n,t-1}} = \frac{\mathcal{R}_{n,t}}{\mathcal{R}_{n,t-1}} = \frac{\mathcal{A}_{n,t}}{\mathcal{A}_{n,t-1}} = (1+\gamma)^{\frac{1}{1-\beta}},$$

while output per capita grows at rate

$$\frac{y_{n,t}}{y_{n,t-1}} = (1+\gamma)^{\frac{1}{\varepsilon-1}+\frac{1}{\theta}\frac{1}{1-\beta}}.$$

Modeling measurement of innovation Here we describe the mapping from ideas in the model to measured innovations in the data. The model has a unit continuum of domains. We interpret this as an approximation to an environment with a large finite number K of domains. For each domain k and time t, let q^*_{kt} denote the productivity of the frontier idea: the highest-productivity idea for that domain discovered anywhere in the world. An innovation is recorded when a newly discovered idea has productivity exceeding the frontier in its domain by a multiple b_t. That is, an idea in domain k discovered in t is measured as an innovation if its productivity exceeds $b_t q^*_{k,t-1}$. The threshold b_t is allowed to vary over time; for example, the threshold for recording an innovation may be less stringent for ideas generated recently than for ideas generated in the distant past.

In Appendix IA7.2, we show that if K is large and the threshold is sufficiently high, the number of measured innovations in country n approaches a Poisson distribution with mean $x_{nt} r_t / \Lambda_{t-1}$, where $r_t \equiv K b_t^{-\theta}$ and $\Lambda_t = \sum_n \lambda_{nt}$ is the global stock of ideas.

The model predictions in the remainder of this section are written in a way that does not require taking a stand on the evolution of r_t over time, as it cancels from the relevant expressions. The evolution of r_t becomes relevant when we estimate the model in Section 5.

4.1 The Mechanics of the Model

In this section, we consider several simple examples that illustrate the mechanics of the model. While each example is stylized, they are designed to show the connection between the forces in the model and our empirical findings.

Concentration of output and innovation In the data, we found that innovations per capita were more concentrated across countries than output per capita. Here we show that technology adoption naturally compresses variation in output per capita relative to innovation. This happens because the ideas used in production are mixtures of ideas generated across countries, and this mixing dampens variation.

Consider a symmetric world with N continents, each containing J countries. For distinct countries $n \neq \tilde{n}$ on the same continent, $\varphi^A_{n\tilde{n}} = \varphi^R_{n\tilde{n}} = \varphi_w$; across continents, $\varphi^A_{n\tilde{n}} = \varphi^R_{n\tilde{n}} = \varphi_a$, with $\varphi_w > \varphi_a$.

In addition, suppose that the only variation in fundamentals across countries is in research productivities $\{Z_n^R\}$. Then along a BGP, there is more variation in innovations per capita than in output per capita if $\theta > 1$. The condition $\theta > 1$ ensures that output does not respond more than one-for-one with ideas. As a measure of concentration, we focus on the coefficient of variation because it is scale invariant and allows us to include observations with zero innovation.

Proposition 1 (Concentration of innovation)**.** *In the symmetric geography described above, suppose that countries are equally sized, that the only variation in fundamentals across countries is in research productivity, and that $\theta > 1$. Then the coefficient of variation of innovations per capita is larger than the coefficient of variation of output per capita.*

Proof. Along a BGP, $x_{n,t-k} = \delta^k x_{n,t}$, where we define $\delta \equiv (1+\gamma)^{-\frac{1}{1-\beta}} < 1$, so that we can express the technology pool as

$$\mathcal{A}_{nt} = \sum_{\tilde{n}} \sum_{k=1}^{\infty} [1-(1-\varphi_{n\tilde{n}}^A)^k] x_{\tilde{n},t-k} = \sum_{\tilde{n}} \sum_{k=1}^{\infty} [1-(1-\varphi_{n\tilde{n}}^A)^k] \delta^k x_{\tilde{n},t} = D^A \sum_{\tilde{n}} p_{n\tilde{n}} x_{\tilde{n},t}$$

where we defined

$$p_{n\tilde{n}} = \frac{\sum_{k=1}^{\infty} [1-(1-\varphi_{n\tilde{n}}^A)^k] \delta^k}{D^A}, \qquad D^A \equiv \sum_{n'} \sum_{k=1}^{\infty} [1-(1-\varphi_{nn'}^A)^k] \delta^k .$$

Let P denote the matrix with entries $p_{n\tilde{n}}$. By construction, each row of P sums to one. In the symmetric geography considered here, each row contains the same own-country, within-continent, and across-continent weights, so each column also sums to one. Thus P is doubly stochastic. It follows that P is an averaging operator: $\mathrm{Var}(Px_t) \le \mathrm{Var}(x_t)$ and $E[Px_t] = E[x_t]$. Hence $\mathrm{CV}(Px_t) \le \mathrm{CV}(x_t)$. Since $\theta > 1$, $\mathrm{CV}\left((Px_t)^{1/\theta}\right) \le \mathrm{CV}(Px_t)$ because, for nonnegative random variables and $a \in (0,1)$, the coefficient of variation of X^a is weakly smaller than that of X. Lastly, since D^A is common, we have

$$\mathrm{CV}(y_t) = \mathrm{CV}\left(\mathcal{A}_t^{1/\theta}\right) \le \mathrm{CV}(\mathcal{A}_t) = \mathrm{CV}(Px_t) \le \mathrm{CV}(x_t)$$

Therefore, as long as $\theta > 1$, the CV of output is lower than the CV of innovation. Because countries are equally sized in this example, the coefficient of variation of x_t is the same as the coefficient of variation of innovations per capita. $\square$

This simple example abstracts from variation in non-research determinants of productivity, such as z^A and z^Y. Enough variation in these other determinants would, in principle, overturn the result.

Conversely, if we observe greater variation in innovation than in output, this suggests that variation in these other determinants is not large enough to overturn the compressive force of technology diffusion.

Persistence of innovation leadership across countries and continents A comparison of Figures 1 and 3 shows that leadership in innovation among continents is more persistent than leadership among countries. Here we show that this naturally follows when within-continent frictions are weaker than across-continent frictions, and research spillovers are positive ($\beta(1+\sigma) > \sigma$).

We consider again a symmetric world with N continents, each containing J countries, and with common diffusion probabilities φ_w within continents and φ_a across continents for both research and adoption, where $\varphi_w > \varphi_a$. Then country-level innovation is less persistent than continent-level innovation.

Proposition 2 (Persistence in Innovation). *Suppose that $\beta(1+\sigma) - \sigma > 0$ and $\varphi_w > \varphi_a$. Log-linearizing around a symmetric balanced growth path, the log deviation of a country's innovation from the mean on its continent is less persistent than the log deviation of a continent's innovation from the global mean.*

Proof. Log-linearizing around a symmetric balanced growth path gives

$$\hat{x}_{nt} = \hat{u}^R_{nt} + \sum_{k=1}^{\infty} \left\{ \omega^k_o \hat{x}_{n,t-k} + \sum_{\tilde{n}\in c(n),\, \tilde{n}\neq n} \omega^k_w \hat{x}_{\tilde{n},t-k} + \sum_{\tilde{n}\notin c(n)} \omega^k_a \hat{x}_{\tilde{n},t-k} \right\},$$

where $c(n)$ denotes the continent containing n, the within-continent sum runs over the other $J-1$ countries on n's continent, the across-continent sum over the remaining $(N-1)J$ countries, and:

$$\begin{aligned}
\hat{u}^R_{nt} &\equiv \sigma\hat{\mathcal{B}}^R_{nt} + \sigma\hat{\eta}_{nt} + (1+\sigma)\hat{z}^R_{nt} \\
\omega^k_o &\equiv \frac{\beta(1+\sigma)-\sigma}{D^R}\delta^k \\
\omega^k_w &\equiv \frac{\beta(1+\sigma)-\sigma}{D^R}\delta^k[1-(1-\varphi_w)^k] \\
\omega^k_a &\equiv \frac{\beta(1+\sigma)-\sigma}{D^R}\delta^k[1-(1-\varphi_a)^k] \\
D^R &\equiv \sum_{k=1}^{\infty}\delta^k \left\{ 1 + \sum_{\tilde{n}\in c(n),\, \tilde{n}\neq n} [1-(1-\varphi_w)^k] + \sum_{\tilde{n}\notin c(n)} [1-(1-\varphi_a)^k] \right\}.
\end{aligned}$$

and $\delta \equiv (1+\gamma)^{-\frac{1}{1-\beta}} < 1$. Define $\tilde{x}_{nt} \equiv \hat{x}_{nt} - \bar{x}_{c(n),t}$ as a country's deviation from its continental mean—where $\bar{x}_{c,t}$ averages $\hat{x}_{nt}$ over the countries in continent c—and $\tilde{x}_{c,t} \equiv \bar{x}_{c,t} - \bar{x}_t$ as a continent's deviation from the global mean. Summing over countries within a continent, as well as over continents, and subtracting, gives that the country-level and continent-level deviations follow stochastic processes:

$$\tilde{x}_{n,t} = \sum_{k=1}^{\infty} (\omega_o^k - \omega_w^k)\, \tilde{x}_{n,t-k} + \tilde{u}_{n,t}^R$$

$$\tilde{x}_{c,t} = \sum_{k=1}^{\infty} \left[(\omega_o^k - \omega_w^k) + J(\omega_w^k - \omega_a^k) \right] \tilde{x}_{c,t-k} + \tilde{u}_{c,t}^R\,,$$

where $\tilde{u}_{n,t}^R \equiv \hat{u}_{nt}^R - \bar{u}_{c(n),t}^R$ and $\tilde{u}_{c,t}^R = \bar{u}_{c,t}^R - \bar{u}_t^R$.

If $\beta(1+\sigma) - \sigma > 0$ and $\varphi_w > \varphi_a$, then $\omega_w^k > \omega_a^k$ for every k. The autoregressive kernel for continent-level deviations therefore exceeds the kernel for within-continent country deviations by $J(\omega_w^k - \omega_a^k)$ at each lag k. Hence deviations of continents from the global mean are more persistent than deviations of countries from their continental means. $\square$

This result arises from stronger diffusion within continents than across them. When research spillovers are positive and within-continent spillovers are larger ($\varphi_w > \varphi_a$), a leading country will push up its neighbors' innovation, making future innovation leadership more likely to remain on the same continent than to move to a different continent.[7] Importantly, this differential persistence would not arise from correlated shocks alone: even if shocks to innovation incentives were more correlated within continents, that would not generate more persistence at the continent level than at the country level.

The Growth Rate of Innovation and the Great Divergence One prominent feature of the data is that the growth rate of innovation has been higher during the past two centuries than before the industrial revolution. Here we study the consequences of faster growth in innovation.

Consider a balanced growth path with equally sized countries in which each country's research productivity grows at the common rate g^R, so that $z_{nt}^R = z_n^R (1+g^R)^t$, and all other fundamentals are constant. Such an economy will have a balanced growth path in which innovation grows at gross rate $1 + g_x = [(1+\gamma)(1+g^R)]^{\frac{1}{1-\beta}}$. We contrast the variation in output along a BGP with high g_x with that in an economy with low g_x.

In doing this, we consider a stylized example in which frictions take the form $\varphi_{n,\tilde{n}}^A = \varphi_n^A$ for $n \neq \tilde{n}$. That is, the diffusion probability depends only on the identity of the receiving country.

[7] This example can be generalized in two ways. First, if we allow for frictions limiting diffusion of ideas for research to differ from the frictions limiting diffusion for adoption, so that $\mathcal{A}_{nt} \neq \mathcal{R}_{nt}$, then the condition for stronger persistence within continents is more involved than simple $\beta(1+\sigma) \geq \sigma$. Nevertheless, the condition is the same as the one that guarantees that higher innovation in one country raises innovation elsewhere, which seems to be the empirically relevant case. Second, we can relax the symmetry within and across continents, building on the notion of network cohesion of Cavalcanti, Giannitsarou, and Johnson (2017). The result follows if there is uniformly stronger network cohesion within continents than across continents. See also Golub and Jackson (2012).

We further focus on the empirically relevant case in which $\mathrm{Cov}\left(u_n^Y, \left(1 - \frac{\exp(u_n^R)}{\sum_{\tilde{n}} \exp(u_{\tilde{n}}^R)}\right) \frac{1-\varphi_n^A}{\varphi_n^A}\right) < 0$, where $\exp(u_n^Y) \equiv (z_n^A)^{\frac{1}{\varepsilon-1}} z_n^Y (\mathcal{B}_n^{AY})^{\frac{\sigma}{1+\sigma}}$ and $\exp(u_n^R) \equiv (z_n^R)^{1+\sigma} (\mathcal{B}_n^R \eta_n)^\sigma$ are the non-research determinants of output productivity and research productivity, respectively. We view this case as natural in that countries that are less open (low φ_n^A) are also likely to be less productive in other dimensions.

Proposition 3. *Suppose that* $\mathrm{Cov}\left(u_n^Y, \left(1 - \frac{\exp(u_n^R)}{\sum_{\tilde{n}} \exp(u_{\tilde{n}}^R)}\right) \frac{1-\varphi_n^A}{\varphi_n^A}\right) < 0$. *Then, to a first-order approximation around* $g_x \searrow 0$, *a BGP with higher growth of innovation will feature:*

1. *higher variance of log output per capita*
2. *a larger share of the variance of log output per capita that is accounted for by* $\log \mathcal{A}_n$.

Proof. See Appendix IA7.3 ◻

The proposition states that in response to an increase in the growth of innovation, there will be a divergence in output, with a larger role for differences in access to ideas. The key is that it takes time for ideas to diffuse across countries. If innovations are growing faster, the same frictions prevent the dissemination of a larger share of ideas.[8]

The Responses of GDP per Capita to Foreign Innovation Figures 10 and 11 show that the response of output to foreign innovations became stronger over time relative to responses to domestic innovation. Proposition 4 shows that this is consistent with gradual weakening of adoption frictions.

Proposition 4. *The ratio of the response of a country's GDP per capita to a foreign innovation relative to its response to a domestic innovation is larger when its resistance to foreign technologies is weaker.*

Proof in Appendix IA7.4.

We also study heterogeneous responses to foreign innovation. Appendix Figure IA6 shows that countries in the bottom GDP-per-capita quartile respond substantially less to the leader's innovations than countries in the top quartile. Proposition 5 shows that this will be the case if a poor country is poor because of adoption frictions. Note that if a poor country is poor for other reasons (Z^A or Z^Y), there would be no reason for a smaller response to foreign innovation.

[8]This mechanism is related to, but distinct from, the small-world mechanism in Lindner and Strulik (2020). In their model, knowledge diffuses through a network with smaller spillovers for more distant countries, and countries taking off at different times depending on their proximity to the early leaders. Their model features divergence in relative outputs along the transition path, but eventual equality in relative outputs. In our model, divergence in relative outputs is a BGP phenomenon. The key difference is that we model diffusion frictions as delaying the transmission of ideas, following Eaton and Kortum (1999), rather than muting their influence. When the frontier moves faster, the cost of delay rises.

Proposition 5. *Consider two countries that are identical except that country 1 is more resistant to adopting foreign technologies than country 2. Then country 2 will both be richer and have a stronger response to foreign innovation.*

Proof in Appendix IA7.5.

5 Quantitative Exploration

In this section we use the model to examine the key quantitative patterns in the data: the concentration of innovation and output, the evolution of innovation and output, and the role of diffusion frictions in shaping the distribution of output across countries.

5.1 Estimation Approach

The model delivers two key estimating equations. Let I_{nt} denote measured innovations in country n during period t. Since $\mathbb{E}[I_{nt}] = r_t x_{nt}/\Lambda_{t-1}$, we can express the two equations as

$$\log \mathbb{E}[I_{nt}] = \log L_{nt} + \beta(1+\sigma)\log \mathcal{R}_{nt} - \sigma \log \mathcal{A}_{nt} + u^R_{nt} \quad (30)$$

$$\log y_{nt} = \frac{1}{\theta}\log \mathcal{A}_{nt} + u^y_{nt}\,. \quad (31)$$

Here

$$u^R_{nt} \equiv \log\left(\frac{r_t}{\Lambda_{t-1}}\tilde{\alpha}^R_t \eta^\sigma_{nt} (\mathcal{B}^R_{nt})^\sigma (z^R_{nt})^{1+\sigma}\right)$$

$$u^y_{nt} \equiv \log\left(\tilde{\alpha}^y_t (\mathcal{B}^{AY}_{nt})^{\frac{\sigma}{1+\sigma}} (z^A_{nt})^{\frac{1}{\varepsilon-1}} z^Y_{nt}\right)$$

collect the non-diffusion determinants of measured innovation and output per capita, respectively. We allow these residual components to contain country-specific persistent and transitory shocks, as well as unrestricted common time effects:

$$u^R_{nt} = \alpha^R_t + \epsilon^R_{nt} + \upsilon^R_{nt},$$

$$u^y_{nt} = \alpha^y_t + \epsilon^y_{nt} + \upsilon^y_{nt}.$$

The time effects α_t^R and α_t^y absorb common movements in innovation and output. The persistent component ϵ_{nt}^y follows an AR(2) process and ϵ_{nt}^R an AR(1) process, while υ_{nt}^R and υ_{nt}^y are white-noise shocks.

We parameterize diffusion probabilities as country-specific barriers, with an additional penalty for cross-continent diffusion. That is, let $\xi^A \geq 1$ and $\xi^R \geq 1$ be the cross-continent penalties and let φ_n^A and φ_n^R be country-specific diffusion parameters. Then we set $\varphi_{n\tilde{n}}^A = \varphi_n^A$ and $\varphi_{n\tilde{n}}^R = \varphi_n^R$ if n and $\tilde{n}$ are on the same continent and $\varphi_{n\tilde{n}}^A = \varphi_n^A/\xi^A$ and $\varphi_{n\tilde{n}}^R = \varphi_n^R/\xi^R$ if n and $\tilde{n}$ are on different continents (with $\varphi_{nn}^A = \varphi_{nn}^R = 1$). Thus there are two country-specific parameters to be estimated, φ_n^A and φ_n^R.

We maintain the simplifying assumption that diffusion probabilities are constant over time. This assumption is clearly restrictive. There are well-known examples in which this assumption has been violated, particularly growth miracles that were often associated with reductions of barriers to foreign ideas. Nevertheless, we view it as a useful starting point to understand how innovation and diffusion shaped the distribution of outputs over the long run.

Identification approach The model provides two reasons a country may be poor. It may have low non-diffusion productivity, captured by u_{nt}^y, or it may face high barriers to adopting foreign ideas, captured by low φ_n^A, or more likely, a combination of the two. To distinguish the two, we make use of the fact that these forces have different implications for how output responds to recent global innovation. A low-productivity country with good access to foreign ideas may have the same output level as a high-productivity country with poor access. However, their dynamics differ following a surge in foreign innovation. Countries with high diffusion probabilities respond more quickly and more strongly, while countries with low diffusion probabilities respond only with delay.

This is also why the assumption of time-invariant diffusion probabilities is useful for identification. With fixed φ_n^A and φ_n^R, repeated observations over time discipline each country's sensitivity to recent innovation flows, allowing the model to separate persistent productivity differences from persistent diffusion barriers.

A potential worry is that time effects α_t^R and α_t^y might be close to collinear with recent global innovation flows. A key distinction is that the time effects induce a common response, but global innovation flows induce differential responses. An increase in the time effect raises output in all countries. An increase in recent global innovation raises output first in countries that are open to foreign ideas; its impact on relatively closed countries is delayed, often substantially.

Past flows of innovation We link measured innovations to the flow of ideas using $x_{nt} = \frac{\Lambda_{t-1}}{r_t}\mathbb{E}[I_{nt}]$. The evolution of the research pool, the adoption pool, and the global stock of ideas is then characterized by

$$\mathcal{R}_{nt} = \sum_{\tilde{n}} \sum_{k=1}^{\infty} \left[1-(1-\varphi^R_{n\tilde{n}})^k\right] \frac{\Lambda_{t-k-1}}{r_{t-k}} \mathbb{E}[I_{\tilde{n},t-k}] \tag{32}$$

$$\mathcal{A}_{nt} = \sum_{\tilde{n}} \sum_{k=1}^{\infty} \left[1-(1-\varphi^A_{n\tilde{n}})^k\right] \frac{\Lambda_{t-k-1}}{r_{t-k}} \mathbb{E}[I_{\tilde{n},t-k}] \tag{33}$$

$$\Lambda_t = \sum_n x_{nt} + \Lambda_{t-1} = \sum_n \frac{\Lambda_{t-1}}{r_t}\mathbb{E}[I_{nt}] + \Lambda_{t-1} \tag{34}$$

To operationalize these equations, we take three further steps. First, when constructing the diffusion pools, we approximate expected innovations with realized innovations, $\mathbb{E}[I_{nt}] \approx I_{nt}$, which will be a good approximation if the number of innovations is large. Second, we allow for the innovation threshold to change over time, $r_t = r_0\nu^t$, with parameters r_0 and ν. Third, we initialize the stocks of ideas using the count of innovation before 1500.

Estimation strategy We set a period length of a decade and fix $\theta = 2$ and $\beta = 0.5$. We then estimate the remainder of the parameters using MLE.

The output equation is cast in state-space form. We evaluate the likelihood with a linear Kalman filter, which integrates out the latent productivity state $\epsilon^y_{n,t}$, concentrates out the time fixed effects α^y_t, and accommodates staggered country entry and internal gaps by skipping missing measurements rather than imputing them.

We model innovation counts as Poisson. This gives a count-data likelihood suited to the many country-decades with zero recorded innovations. We evaluate the innovation likelihood using a Poisson state-space filter, implemented with a Laplace approximation to the non-Gaussian measurement equation. This integrates out the latent innovation component ϵ^R_{nt} and concentrates out the time effects α^R_t.

We assume that the reciprocals of the diffusion probabilities, $(\varphi^R_n)^{-1}$ and $(\varphi^A_n)^{-1}$, are drawn from a joint distribution with Pareto marginals, with tail parameters ζ^R and ζ^A, and dependence governed by a Gumbel copula.

Estimation sample We estimate the model on a decade-level panel from 1500 to 1999, giving 50 10-year periods from the 1500s through the 1990s. An important bottleneck is that we do not have long histories of GDP data for many countries. We incorporate a country in the estimation sample if it has at least 10 of these 50 decades with at least one year of non-missing, strictly positive GDP per capita in the Maddison Project Database (Bolt and van Zanden, 2024). This "10-decade" rule keeps countries with enough long-run output coverage to discipline the sensitivity of output and innovation to recent global flows of innovation. This rule yields 63 countries; tightening it to 15 or 20 decades gives nested subsamples, and the estimates are stable across them.[9] The estimation uses the full 63-country sample, but each country's diffusion pool reflects the entire observed world. For readability the figures and tables display 26 example countries—8 in the Americas, 4 in Asia, and 14 in Europe—chosen for their long output histories and their large shares of world innovation and GDP, which together they overwhelmingly account for; the estimation itself uses all 63. The full display list appears in Table IA4. Innovation counts come from the Wikipedia dataset of Section 2.

5.2 Estimation Results

Appendix Table IA3 reports the parameter estimates and Table IA4 shows the diffusion probabilities for each country.

Output-side estimates Our estimates indicate that distance reduces adoption probabilities. The cross-continent penalty is $\xi^A = 2.05$, so a technology from another continent diffuses at roughly half the speed of a technology from the same continent. The country-specific diffusion probabilities φ_n^A span more than an order of magnitude (shape index $\zeta^A = 0.53$). Given a period length of a decade, φ_n^A has the interpretation as the probability that an unavailable idea becomes available in a country in a decade for adoption. At the top sit the frontier economies—Switzerland (0.971), Belgium (0.942), the United States (0.913), Denmark (0.885), and the United Kingdom (0.857)—which draw rapidly on the world technology pool.

Innovation-side estimates The insight diffusion probabilities are distinct from the technology diffusion probabilities, but two features stand out. First, they are strongly but imperfectly aligned with the adoption speeds: across countries the rank correlation between φ_n^A and φ_n^R is Kendall's $\tau_K = 0.65$ (Gumbel copula parameter 2.8), so a country that is slow to adopt foreign technologies is typically also

[9]Within the panel, data enter as observed. A country enters the estimation when its GDP per capita is first observed, and the two blocks are treated symmetrically: a country-decade contributes to the output likelihood only in decades when its GDP per capita is observed, and a country's pre-entry decades—those before its first observed GDP—are likewise excluded from the innovation likelihood, so pre-entry country-decades count for neither block. The Kalman filter handles staggered entry and internal gaps, so missing country-decades are skipped rather than imputed. From entry onward, innovation counts are set to zero in country-years with no recorded innovation. Population, which is needed in every decade to express counts per capita, is log-linearly interpolated across internal gaps and held flat beyond the observed range.

slow to draw foreign insights—though the alignment is far from exact. Second, insights generally diffuse *more slowly* than technology, so that $\varphi_n^R < \varphi_n^A$ for most countries, and insights are more localized than the technologies they eventually become. This can be seen in Figure 12. The United States is the clearest example: it is close to the technology frontier in adoption ($\varphi_n^A = 0.913$), but draws much more slowly on foreign insights ($\varphi_n^R = 0.054$). In the model, its innovation therefore reflects domestic research productivity and domestic insight accumulation much more than rapid access to foreign research insights. The labor-supply elasticity is $\sigma = 0.35$ (imprecisely estimated), which with $\beta = 0.5$ gives a net pool elasticity $\beta(1+\sigma) - \sigma = 0.32$. Research productivity is highly persistent, $\rho_R = 0.99$, close to a unit root—once a country pulls ahead in u_{nt}^R the lead erodes only slowly, which is why innovation leadership turns over on the scale of centuries rather than decades. Allowing a second AR lag does not improve the fit, so the reported specification uses a single persistent component.

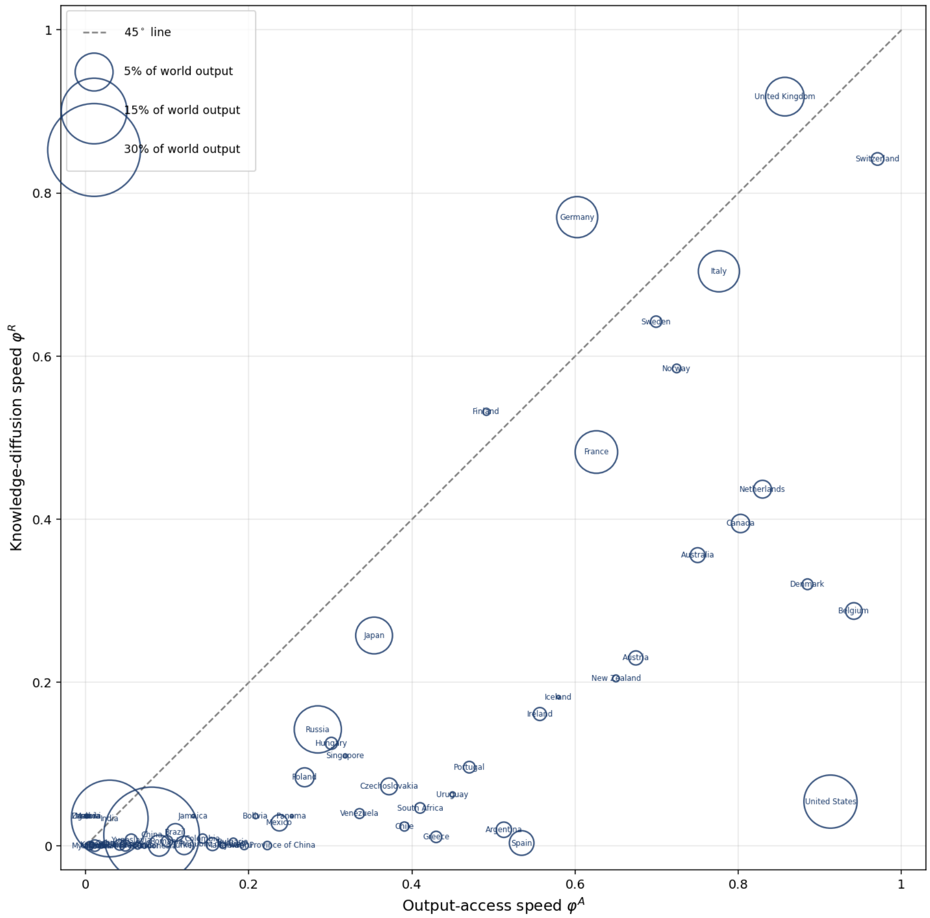
Implied research productivity The innovation block yields a panel of research productivity u_{nt}^R—the part of innovation not explained by population scale or by access to the diffusion pools. Appendix Figure IA18 plots its persistent component ϵ_{nt}^R—filtered from the Poisson likelihood, net of the common decade effect—for 26 example countries on a common scale, against each country's sample mean. The United States climbs steadily across the 19th and 20th centuries to the top of the distribution; the United Kingdom and the early European frontier sit high from the outset and drift gently down; Japan rises sharply in the 20th century; and China and India sit far below the frontier throughout. That this dispersion is overwhelmingly persistent—not a matter of differential access—is the counterpart of the near-unit-root ρ_R.

What drives the geography of innovation and output? We decompose the cross-country variance of log innovations per capita and log output per capita into two structural channels. The first is differential access to global idea pools: the insight pool $\mathcal{R}$ for innovation and the technology pool $\mathcal{A}$ for output. The second is residual country-specific forces, captured by u^R and u^y. Figure 13 presents the results.[10] For output, the access channel grows from about a fifth of cross-country dispersion before 1800 to nearly nine-tenths by 1900: differential access to the global technology pool accounts for most of the modern cross-country output gap. For innovation, the same channel is small until the 19th century—country-specific shocks dominate—and then rises steeply, accounting for about half of the variation by the mid-20th century.

The timing difference reflects the estimated separation between adoption and research diffusion. Adoption diffusion is generally faster than knowledge diffusion ($\varphi^A > \varphi^R$), so foreign technologies

[10]The values, split into the separate persistent and transitory pieces, are available in Table IA5.

Figure 12. Spatial Diffusion Speeds φ^A and φ^R



Notes: The figure plots, for each country, the per-decade diffusion probabilities φ_n^A (horizontal axis, adoption) and φ_n^R (vertical axis, knowledge), in levels—each the probability that an unavailable foreign idea becomes accessible within a decade. Each open circle is a country; its area is proportional to that country's average share of world output, with the legend circles corresponding to 5%, 15%, and 30% of world output. The dashed line is the 45-degree line ($\varphi_n^A = \varphi_n^R$); points below it have $\varphi_n^R < \varphi_n^A$, i.e., insights diffuse more slowly than technologies. Full set of φ_n^A and φ_n^R available in Table IA4.

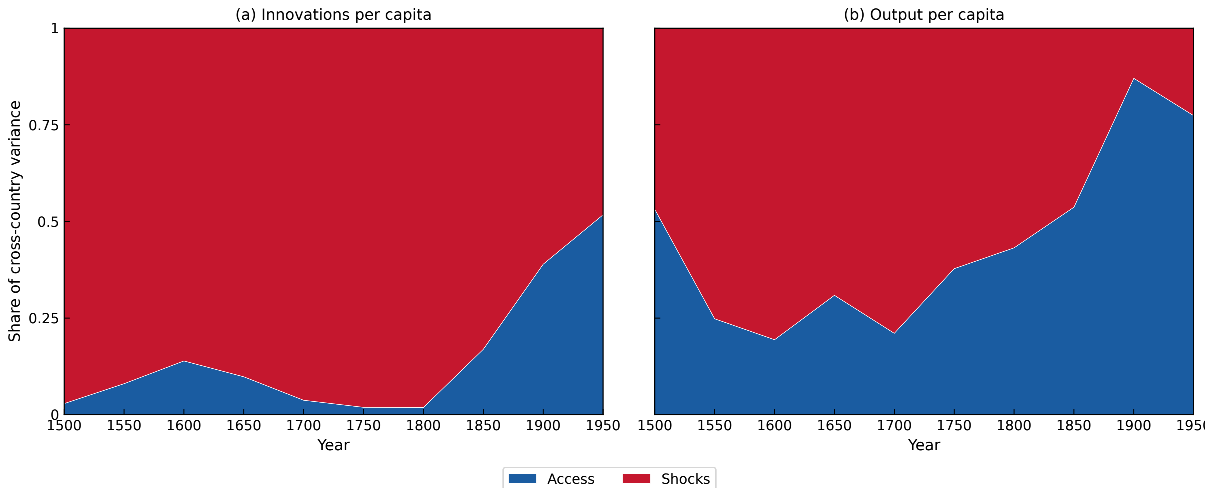
show up in output earlier than foreign insights show up in research. As a result, differential access becomes central to the geography of output before it becomes central to the geography of innovation.

Figure 14 illustrates the role of differential access in the evolution of the GDP-per-capita distribution. Since output per capita satisfies

$$\log y_{nt} = \frac{1}{\theta} \log \mathcal{A}_{nt} + u_{nt}^y,$$

we separate the access term, $\frac{1}{\theta} \log \mathcal{A}_{nt}$, from the residual component. The figure plots the cross-country standard deviation of log GDP per capita and of the access term. The standard deviation of log output per capita rises from 0.51 around 1700 to 1.06 by 1990. Most of this increase runs through access: by 1990, $\mathrm{sd}(\frac{1}{\theta} \log \mathcal{A}) = 0.85$, close to the dispersion of log output itself. In the model, the modern output

Figure 13. Structural Variance Decomposition over Time



Notes: Each panel shows the share of the cross-country variance of log innovations per capita (left) and log output per capita (right) attributable to the structural components of the estimable equations, by 50-year period: access (differential access to the global technology pool through the diffusion speeds φ^A and φ^R, blue) and shocks (the filtered AR research- and output-productivity states together with transitory disturbances, red). Shares are covariance decompositions, $\mathrm{Cov}(\text{component}, \text{total})/\mathrm{Var}(\text{total})$, that sum to one within each panel; small negative shares are set to zero in the area plot. Table IA5 reports the same shares split into the separate persistent and transitory pieces.

divergence is therefore largely a divergence in the stock of ideas countries can access.

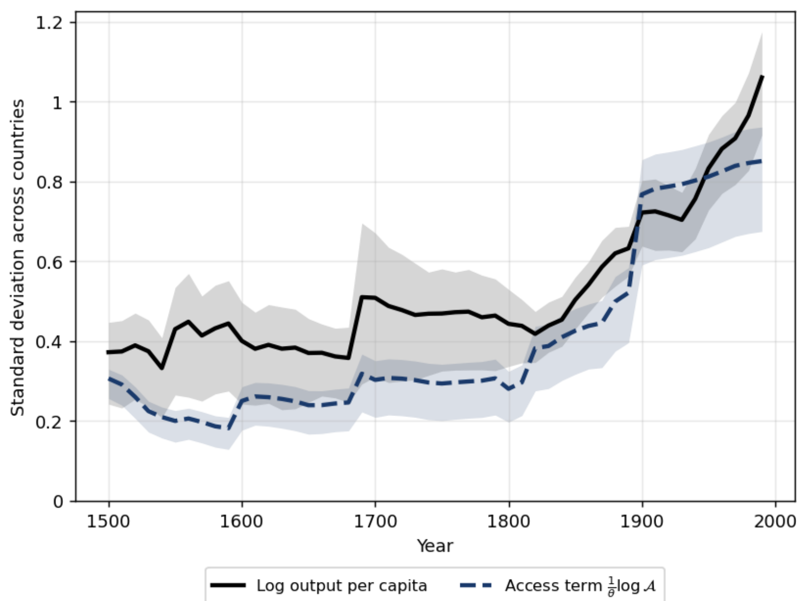
These patterns are consistent with Proposition 3, which showed that an acceleration in the pace of innovation would raise variation in log output and increase the share of variation accounted for by access.

Counterfactual diffusion speeds We next use the estimated model to ask how the 1990 distribution of GDP per capita changes when we alter access to foreign ideas for adoption and research.

Figure 15 reports counterfactual changes in end-of-period log output per capita when we change adoption speeds φ^A. The effects are large and highly unequal. Setting $\varphi^A = 1$ for all countries raises output per capita by 0.95 log points on average—a factor of about 2.6—but the gains are concentrated among isolated economies. Zambia and Uganda gain more than three log points, while the frontier changes little: the United States gains 0.01 log points and Switzerland 0.15, because they already draw on the world technology pool almost freely.

The opposite experiment, capping adoption speeds at $\varphi_n^A \leq 0.05$, falls hardest on well-connected advanced economies. Belgium, Denmark, and New Zealand each lose about a full log point, while the United States is nearly unchanged. Strikingly, 15 of the 63 countries actually gain under the cap. This reflects an equilibrium response: slowing frontier access reallocates labor toward research, raising the

Figure 14. Cross-Country Dispersion of Output and Access



Notes: The figure plots the cross-country standard deviation, across the countries observed each decade from 1500 to 1990, of log output per capita (black) and of the access term $\frac{1}{\theta}\log\mathcal{A}$ (blue). Output per capita is decomposed as $\log y_{nt} = \frac{1}{\theta}\log\mathcal{A}_{nt} + u^{y}_{nt}$, so the access term measures the knowledge a country can reach. Shaded bands are 90% confidence intervals (output via country bootstrap; access via model parameter draws).

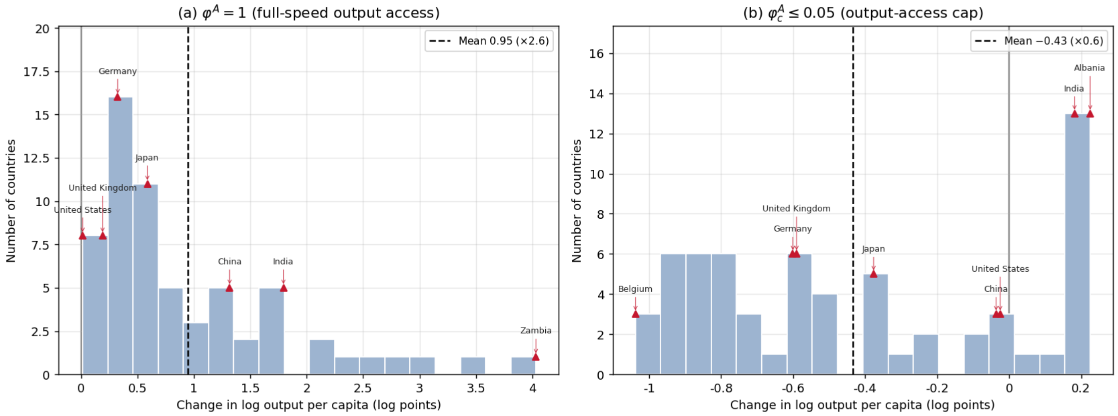
world stock of ideas and expanding the pool eventually available to low-access countries. Thus adoption speeds mainly reshape the cross-country distribution of output.

Figure 16 shows that insight diffusion probabilities φ^R operate differently. Changing φ^R mostly shifts the level of world output rather than reshaping its distribution. Setting $\varphi^R = 1$ raises nearly every country's output by a similar amount: the mean gain is 0.90 log points, with a cross-country standard deviation of only 0.24. Capping $\varphi^R_n \leq 0.05$ lowers output by a nearly uniform 0.11 log points, with a standard deviation of 0.04. Since the distribution moves almost as a block, country rankings and output concentration change little.

The contrast is clear. Adoption frictions govern cross-country convergence because they determine which countries can use the world technology pool. The largest counterfactual gains when output access is opened up accrue to poor, isolated countries. Research-diffusion frictions primarily affect the level of world output because they shape the production of ideas that eventually diffuse broadly.

These patterns follow directly from the model's reduced form. Output per capita scales with the technology pool a country can reach, $y_n \propto \mathcal{A}_n^{1/\theta}$ (Equation 9), and the adoption speeds φ^A are precisely what make that pool unequal across countries, since each country reaches the world stock only through the diffusion-weighted sum in Equation 7, in which a foreign vintage of age k enters at rate $1-(1-\varphi^A_{n\tilde{n}})^k$.

Figure 15. Counterfactual Output under Output-Access Speeds (φ^A)



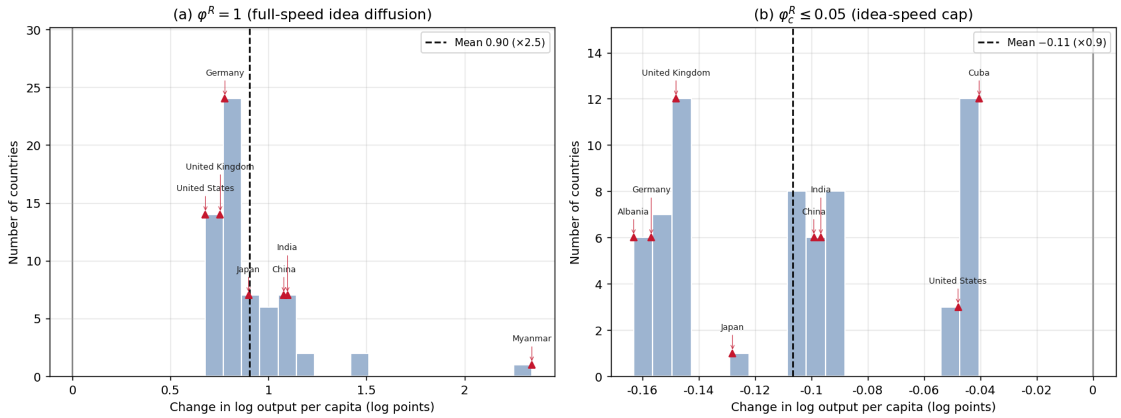
Notes: Each panel shows the histogram, across the countries observed in the final decade (1990), of the change in end-of-period log output per capita relative to its actual value, $\Delta \log y$ (in log points), under a counterfactual to the estimated output-access speeds. Panel (a) sets $\varphi^A = 1$ (full-speed output access); panel (b) imposes the cap $\varphi^A_n \leq 0.05$ (output-access cap). The grey vertical line marks no change; the dashed black line marks the cross-country mean, with the implied multiplicative factor reported in the legend. Red triangles flag selected countries (full country names), including the largest gain and loss. The distributions are wide and skewed: output access mainly redistributes.

Raising φ^A equalizes access to a common stock and thus *compresses* the cross-country distribution of $\frac{1}{\theta} \log \mathcal{A}_n$, while capping φ^A pulls the well-connected frontier back toward the laggards—so φ^A moves dispersion. Knowledge speeds φ^R, by contrast, act on the insight pool $\mathcal{R}$ that governs innovation (Equation 6) and reach output only through the world idea stock that subsequently diffuses to all; raising φ^R therefore raises every country's technology pool roughly in proportion—a level shift in $\log y$ that leaves relative positions, and hence concentration, intact. The contrast is the counterfactual counterpart of the model's central separation between the technology pool $\mathcal{A}$ (governed by φ^A) and the insight pool $\mathcal{R}$ (governed by φ^R).

Understanding the rise and fall of giants: the United Kingdom and the United States Figure 17 traces the two dominant innovators of the past half-millennium and asks how the model views their rise and fall. Panel (a) shows the raw data. The United Kingdom's share of world innovation rose to roughly one half by 1800, near the height of the Industrial Revolution, and then fell below one tenth by 1990. The United States rose from near zero to roughly seven tenths, overtaking the United Kingdom around 1890. The same reversal appears in per-capita terms: British innovations per capita rose sharply during the Industrial Revolution but grew only modestly afterward, while American innovations per capita rose steadily and by the late 20th century were roughly twice as high.

Panel (b) decomposes each country's innovation lead. It plots log excess innovation relative to the average innovating country (solid), the access component of that lead (dashed), and a fixed-population

Figure 16. Counterfactual Output under Knowledge Diffusion Speeds (φ^R)



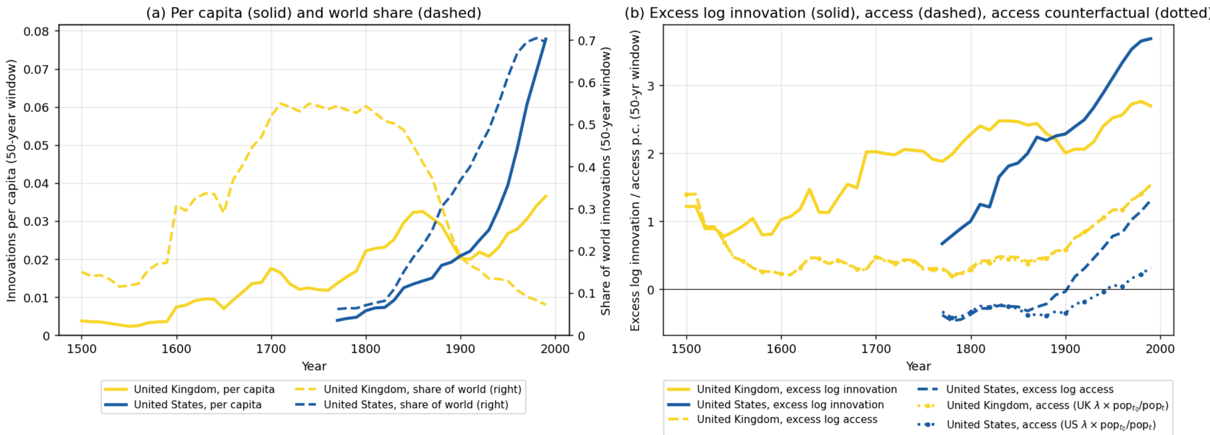
Notes: Each panel shows the histogram, across the countries observed in the final decade (1990), of the change in end-of-period log output per capita relative to its actual value, $\Delta \log y$ (in log points), under a counterfactual to the estimated knowledge diffusion speeds. Panel (a) sets $\varphi^R = 1$ (full-speed idea diffusion); panel (b) imposes the cap $\varphi_n^R \leq 0.05$ (idea-speed cap). The grey vertical line marks no change; the dashed black line marks the cross-country mean, with the implied multiplicative factor reported in the legend. Red triangles flag selected countries (full country names), including the largest gain and loss. The distributions are tight spikes—far from zero under $\varphi^R = 1$, just below it under $\varphi_n^R \leq 0.05$: knowledge diffusion shifts the level.

counterfactual for access (dotted). The access component is $m^R \equiv (1+\sigma)\beta \log \mathcal{R}_{nt} - \sigma \log \mathcal{A}_{nt}$— the part of innovation explained by the insight and technology pools. The gap between total innovation (solid) and access (dashed) reflects research productivity and transitory shocks u_{nt}^R. In the counterfactual (dotted) lines, the access series are recomputed assuming that $\tilde{\lambda}_{n,t} = \lambda_{n,t} L_{n,t_0} / L_{n,t}$, for $n = UK, US$, respectively, where t_0 is given by each country's first year in the sample.

For the first four hundred years for the UK and the first century for the US, in the model's account, both ascents are overwhelmingly stories of *research productivity*, not changes in access. At the height of British leadership around 1800, the United Kingdom's excess innovation reaches roughly $+2.3$ log points (solid), while the access term contributes only about $+0.3$. The model therefore attributes most of Britain's lead to research productivity rather than to unusually favorable access to the global idea pools. The rise of the United States is a mixed story: research productivity is the primary driving force through the 19th century—over which its excess access stays near zero or even slightly negative—with access becoming a central contributor only in the 20th century, when it climbs to roughly a third of the U.S. innovation lead.

The fixed-population counterfactual (dotted line) shows that, for the United States, the rise in access was driven primarily by the expansion of its domestic insight pool as population grew. For both countries, access also rose when innovative but relatively closed countries entered the global stock of ideas, especially in the second half of the 19th century. This channel was more important for the United

Figure 17. The Rise and Fall of Giants: United Kingdom and United States Innovation



Notes: Panel (a) plots, for the United Kingdom and the United States by decade, innovations per capita (solid, left axis)—the count of innovations over the trailing 50 years relative to population—and each country's share of world innovations (dashed, right axis), summed over the model's 63 countries. Panel (b) shows each country's (log) excess *innovation* relative to the world average (solid lines), the (log) excess *access* component (dashed lines), and a counterfactual excess access that holds the country's population fixed at its initial value (dotted lines); yellow denotes the United Kingdom and blue the United States. Access is defined as $m^R \equiv (1+\sigma)\beta \log \mathcal{R}_{nt} - \sigma \log \mathcal{A}_{nt}$—the part of innovation attributable to the global insight and technology pools $\mathcal{R}$ and $\mathcal{A}$ through the estimated diffusion speeds φ^A and φ^R; the gap between the innovation (solid) and access (dashed) lines reflects research productivity u^R and transitory shocks; units are log points. Excess series are deviations from the contemporaneous mean over innovating countries, smoothed with a 50-year trailing window. In the counterfactual (dotted) lines, the access series are recomputed assuming that $\tilde{\lambda}_{n,t} = \lambda_{n,t} L_{n,t_0}/L_{n,t}$, for $n = UK, US$, respectively, where t_0 is given by each country's first year in the sample. Each country's series begins in its appropriate first decade (the United Kingdom in 1500, the United States in the 1770s), and all series use the 50-year trailing window. Innovation counts are the Wikipedia/LLM data; population is from the Maddison Project Database (Bolt and van Zanden, 2024).

Kingdom, which was more open to foreign insights than the United States, $\varphi^R_{UK} > \varphi^R_{US}$.

6 Additional Analyses on Mechanisms

Here we provide suggestive evidence of mechanisms behind the concentration of innovation.

6.1 The Nature of Research Spillovers

Our main analysis shows that innovation by the global leader predicts future innovation in other countries. This relationship could reflect technological spillovers: new ideas create insights that others build on. Alternatively, it could reflect a broader diffusion of innovative capacity, culture, or institutions.

We distinguish these mechanisms by separating innovations into technological domains. Within-domain spillovers point to a technological-insight channel: ideas in a domain help generate further ideas in that same domain. Cross-domain spillovers would instead suggest a more general force raising innovation across the board.

Appendix IA5.7 shows that the spillovers are concentrated within domains. The local-projection response of local innovation to leader innovation is accounted for almost entirely by same-domain innovation. This evidence points to research spillovers operating primarily through technological insights rather than through a broad diffusion of innovation culture.

A related question is whether the rise and fall of overall innovation leaders is driven by changes in broad based local innovative activity or by fixed local comparative advantages in innovation in particular domains combined with the rise and fall of individual domains. To investigate this, we decompose innovation activity into a global domain-specific component and a local country-specific component in Appendix Section IA5.7. The local component accounts for a substantially larger share of the variance in innovation growth. Countries do not rise and fall as innovators because they happen to specialize in globally ascendant technologies—they rise and fall because of local changes in innovative activity.

6.2 Private versus Public Innovation

In Appendix Section IA5.8, we present the patterns for private versus public innovations. We define private innovations as those attributed to companies and individual inventors, which respond more directly to market incentives and local economic conditions; and public innovations as those by universities, government entities, or pertaining to military and defense or basic research. The broad spatial patterns are similar for both types. Both are geographically concentrated, with private innovations perhaps a bit more concentrated.

The literature provides extensive evidence that public research generates spillovers to the private sector. A natural question is whether the productivity of local research in our setting is driven primarily by public research. We examine this in Appendix IA5.8 by studying the temporal covariance between public and private innovations.

The evidence points against public research as the main leading force. Private innovation is more closely associated with future public innovation than with past public innovation, while public innovation is similarly associated with past and future private innovation. This timing pattern suggests that, if anything, private innovation tends to lead public innovation, rather than the reverse. One interpretation is that private innovation helps shape or inspire subsequent public research.

6.3 Other Creative Endeavors

We also examine the geographic concentration of other creative activity. We collect data on advances in mathematics and on significant works of literature, two creative activities that are relatively well measured and have broad coverage over long periods of time.

Appendix IA5.9 shows that innovation, mathematics, and literature are all substantially more concentrated than economic output or population. This is true whether concentration is measured by the number of countries accounting for half of total activity (Panel (a) of Figure IA17) or by the coefficient of variation in activity per capita (Panel (b) of Figure IA17). Mathematics was the most concentrated of the three through the 19th century, followed by innovation and then literature. Over time, however, the concentration of mathematics declined while the concentration of innovation rose, so that technological innovation became the most concentrated activity in recent decades. While there is some overlap in the countries at the forefront of each type of creative activity, there are also many differences. For example, France and Germany play a larger role in mathematics than in technological innovation.

The concentration of all three suggests that agglomeration forces are important across creative activities. The especially high concentration of mathematics is also informative: because mathematics is less directly tied to intellectual property protection, its concentration suggests that variation in patent rights or other formal appropriability mechanisms is unlikely to be the sole source of the extreme concentration of innovation.

7 Conclusion

We analyze patterns of innovation and production around the world over the very long run. The data show that innovation is extraordinarily concentrated geographically, far more than population or economic output, and that this concentration persists even among high-income countries. The centers of innovation shift over time, and many advanced economies have achieved substantial growth in recent times despite scarce or stagnant domestic innovation. The combination of concentrated innovation and widespread growth suggests that diffusion of technological adoption is key to economic prosperity.

To investigate the economic mechanisms behind these facts, we develop a framework combining innovation, knowledge flows, and technology adoption. Innovation concentrates while production remains relatively dispersed, since diffusion allows all countries to potentially benefit from innovations made elsewhere. Taking the model to five centuries of data, we find that frictions in knowledge diffusion exceed frictions in technology adoption, that differential access to the world technology pool accounts for most of the modern divergence in output, and that the forces in the model shed light on the shift of innovation leadership from Europe to the United States.

References

Acemoglu, Daron, Philippe Aghion, and Fabrizio Zilibotti. 2006. "Distance to frontier, selection, and economic growth." *Journal of the European Economic Association*, 4(1): 37–74.

Aghion, Philippe, and Peter Howitt. 1992. "A model of growth through creative destruction." *Econometrica*, 60(2): 323–351.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas. 2017. "The rise of American ingenuity: Innovation and inventors of the golden age." Working Paper.

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi. 2018. "Dancing with the stars: Innovation through interactions." Working Paper.

Alvarez, Fernando E, Francisco J Buera, Robert E Lucas, et al. 2013. "Idea flows, economic growth, and trade." National Bureau of Economic Research.

Andrews, Michael J., and Alexander Whalley. 2022. "150 years of the geography of innovation." *Regional Science and Urban Economics*, 94: 103627.

Arkolakis, Costas, Natalia Ramondo, Andrés Rodríguez-Clare, and Stephen Yeaple. 2018. "Innovation and production in the global economy." *American Economic Review*, 108(8): 2128–2173.

Asirvatham, Hemanth. 2024. "The dynamo is not the computer." Working Paper.

Asirvatham, Hemanth, Elliott Mokski, and Andrei Shleifer. 2026. "GPT as a measurement tool." Working Paper.

Audretsch, David B, and Maryann P Feldman. 1996. "R&D spillovers and the geography of innovation and production." *American Economic Review*, 86(3): 630–640.

Benhabib, Jess, Jesse Perla, and Christopher Tonetti. 2014. "Catch-up and fall-back through innovation and imitation." *Journal of Economic Growth*, 19(1): 1–35.

Bergeaud, Antonin, Ruveyda Nur Gozen, and John Van Reenen. 2026. "Mapping technological trajectories: Evidence from two centuries of patent data." Working Paper.

Bernard, Andrew B, Jonathan Eaton, J Bradford Jensen, and Samuel Kortum. 2003. "Plants and productivity in international trade." *American Economic Review*, 93(4): 1268–1290.

Bolt, Jutta, and Jan Luiten van Zanden. 2024. "Maddison Project Database 2023." *Groningen Growth and Development Centre, University of Groningen*, https://doi.org/10.34894/INZBF2.

Buera, Francisco J, and Ezra Oberfield. 2020. "The global diffusion of ideas." *Econometrica*, 88(1): 83–114.

Cai, Jie, Nan Li, and Ana Maria Santacreu. 2022. "Knowledge diffusion, trade, and innovation across countries and sectors." *American Economic Journal: Macroeconomics*, 14(1): 104–145.

Carlino, Gerald, and William R Kerr. 2015. "Agglomeration and innovation." In *Handbook of Regional and Urban Economics*. Vol. 5, 349–404. Elsevier.

Cavalcanti, Tiago VV, Chryssi Giannitsarou, and Charles R Johnson. 2017. "Network cohesion." *Economic Theory*, 64(1): 1–21.

Chikis, Craig A, Benny Kleinman, and Marta Prato. 2025. "The geography of innovative firms." Working Paper.

Comin, Diego, and Bart Hobijn. 2004. "Cross-country technology adoption: Making the theories face the facts." *Journal of Monetary Economics*, 51(1): 39–83.

Comin, Diego, and Bart Hobijn. 2010. "An exploration of technology diffusion." *American Economic Review*, 100(5): 2031–2059.

Comin, Diego, and Martí Mestieri. 2018. "If technology has arrived everywhere, why has income diverged?" *American Economic Journal: Macroeconomics*, 10(3): 137–178.

Comin, Diego, William Easterly, and Erick Gong. 2010. "Was the wealth of nations determined in 1000 BC?" *American Economic Journal: Macroeconomics*, 2(3): 65–97.

Eaton, Jonathan, and Samuel Kortum. 1999. "International technology diffusion: Theory and measurement." *International Economic Review*, 40(3): 537–570.

European Patent Office. 2024. "PATSTAT Global: EPO Worldwide Patent Statistical Database." https://www.epo.org/en/searching-for-patents/business/patstat Accessed: 2025.

Gerschenkron, Alexander. 1962. *Economic Backwardness in Historical Perspective: A Book of Essays.* Cambridge, MA:Harvard University Press.

Giroud, Xavier, Ernest Liu, and Holger M Mueller. 2026. "Innovation spillovers across U.S. tech clusters." *Journal of Financial Economics*, 179: 104264.

Golub, Benjamin, and Matthew O Jackson. 2012. "How homophily affects the speed of learning and best-response dynamics." *Quarterly Journal of Economics*, 127(3): 1287–1338.

Gordon, Robert J. 2016. *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War.* Princeton, NJ:Princeton University Press.

Grossman, Gene M, and Elhanan Helpman. 1991. "Quality ladders and product cycles." *Quarterly Journal of Economics*, 106(2): 557–586.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic localization of knowledge spillovers as evidenced by patent citations." *Quarterly Journal of Economics*, 108(3): 577–598.

Jones, Charles I. 1995. "R&D-based models of economic growth." *Journal of Political Economy*, 103(4): 759–784.

Jordà, Òscar, Moritz Schularick, and Alan M Taylor. 2017. "Macrofinancial history and the new business cycle facts." *NBER Macroeconomics Annual*, 31(1): 213–263.

Jovanovic, Boyan, and Rafael Rob. 1989. "The growth and diffusion of knowledge." *Review of Economic Studies*, 56(4): 569–582.

Kalyani, Aakash, Nicholas Bloom, Marcela Carvalho, Tarek Hassan, Josh Lerner, and Ahmed Tahoun. 2025. "The diffusion of new technologies." *Quarterly Journal of Economics*, 140(2): 1299–1365.

Klenow, Peter J, and Andres Rodriguez-Clare. 2005. "Externalities and growth." *Handbook of Economic Growth*, 1: 817–861.

König, Michael D, Jan Lorenz, and Fabrizio Zilibotti. 2016. "Innovation vs. imitation and the evolution of productivity distributions." *Theoretical Economics*, 11(3): 1053–1102.

Kortum, Samuel S. 1997. "Research, patenting, and technological change." *Econometrica*, 1389–1419.

Krugman, Paul. 1991. "Increasing returns and economic geography." *Journal of Political Economy*, 99(3): 483–499.

Kuvshinov, Dmitry, and Kaspar Zimmermann. 2022. "The big bang: Stock market capitalization in the long run." *Journal of Financial Economics*, 145(2): 527–552.

Landes, David S. 1969. *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present.* Cambridge:Cambridge University Press.

Laouenan, Morgane, Palaash Bhargava, Jean-Benoît Eyméoud, Olivier Gergaud, Guillaume Plique, and Etienne Wasmer. 2022. "A cross-verified database of notable people, 3500BC-2018AD." *Scientific Data*, 9(290): 1–19.

Lind, Nelson, and Natalia Ramondo. 2023. "Global innovation and knowledge diffusion." *American Economic Review: Insights*, 5(4): 494–510.

Lind, Nelson, and Natalia Ramondo. 2024. "Global knowledge and trade flows: Theory and measurement." *Journal of International Economics*, 151. Modern advances in International Trade, Robert Staiger and Cecile Gaubert, Eds.

Lindner, Ines, and Holger Strulik. 2020. "Innovation and inequality in a small world." *International Economic Review*, 61(2): 683–719.

Liu, Ernest, and Song Ma. 2024. "Innovation networks and R&D allocation." Working Paper.

Mokyr, Joel. 2002. *The Gifts of Athena: Historical Origins of the Knowledge Economy.* Princeton, NJ:Princeton University Press.

Moretti, Enrico. 2021. "The effect of high-tech clusters on the productivity of top inventors." *American Economic Review*, 111(10): 3328–3375.

Needham, Joseph. 1954. *Science and Civilisation in China: Volume 1, Introductory Orientations.* Cambridge:Cambridge University Press.

Parente, Stephen L, and Edward C Prescott. 1994. "Barriers to technology adoption and development." *Journal of Political Economy*, 102(2): 298–321.

Perla, Jesse, Christopher Tonetti, and Michael E. Waugh. 2021. "Equilibrium technology diffusion, trade, and growth." *American Economic Review*, 111(1): 73–128.

Romer, Paul M. 1990. "Endogenous technological change." *Journal of Political Economy*, 98(5, Part 2): S71–S102.

Sakakibara, Mariko, and Lee Branstetter. 2001. "Do stronger patents induce more innovation? Evidence from the 1988 Japanese patent law reforms." *RAND Journal of Economics*, 32(1): 77–100.

Sampson, Thomas. 2023. "Technology gaps, trade, and income." *American Economic Review*, 113(2): 472–513.

Serafinelli, Michel, and Guido Tabellini. 2022. “Creativity over time and space.” *Journal of Economic Growth*, 27: 1–43.

Trouvain, Florian. 2024. “Technology adoption, innovation, and inequality in a global world.” Working Paper.

Usher, Abbott Payson. 1929. *A History of Mechanical Inventions.* New York:McGraw-Hill Book Company.

Wikimedia Foundation. 2025. “Wikimedia Downloads: English Wikipedia Dumps.” https://dumps.wikimedia.org/enwiki/latest/ Accessed: 2025.

Yu, Amy Zhao, Shahar Ronen, Kevin Hu, Tiffany Lu, and César A. Hidalgo. 2016. “Pantheon 1.0, a manually verified dataset of globally famous biographies.” *Scientific Data*, 3(150075): 1–16.

Internet Appendix

IA1 Processing Wikipedia Articles for Technological Innovations

This appendix describes in detail how we obtain and process Wikipedia articles on technological inventions used in Section 3.

IA1.1 Wikipedia Articles on Technological Innovations

We obtain English Wikipedia articles and their topic tags from the Wikimedia English Wikipedia Dumps (Wikimedia Foundation, 2025). To construct the initial candidate set, we retain all articles assigned at least one STEM topic tag. Because this broad classification includes many articles that are related to people, events, academic journals, creatures, general scientific discoveries, and other topics that do not constitute technological innovations, we apply a single LLM-based filtering step.

For each candidate article, we provide Gemini-3.0-Flash with its title and short description. The LLM classifies the article as either `KEEP` or `REMOVE`. An article is retained if it describes a technological innovation—that is, a novel device, method, process, system, or engineered artifact that applies scientific or engineering principles to solve a problem, improve efficiency, or introduce new functionality. The filtering prompt is reproduced below.

```
You are filtering a list of Wikipedia articles to identify technological
innovations---novel devices, methods, processes, systems, or engineered
artifacts that apply scientific or engineering principles to solve a
problem, improve efficiency, or introduce new functionality.

Given an article title, lead extract, and short description, respond with
only KEEP or REMOVE.

Remove the article if it is primarily about any of the following:

- A person or biography (inventor, scientist, engineer, military figure,
  etc.)
- A historical event, war, battle, or military organization/unit
- A company, organization, or institution (unless the name also refers
  to the technology itself, e.g. "Dropbox")
```

```
- A place, region, country, city, or geographic feature (including named
  bodies of water, mountains, etc.)
- A natural phenomenon or environment (weather, climate, ecosystems,
  geology)
- An astronomical object or phenomenon (stars, planets, moons, eclipses,
  constellations)
- A scientific concept, theory, law, or mathematical abstraction with
  no direct physical instantiation
- A building, architectural style, or structure that is not primarily
  functional engineering (e.g. cathedrals, palaces, monuments)
- A dam, power plant, bridge, road, canal, or other specific named
  infrastructure project (i.e. a single named instance, not the
  technology class)
- A biological species, organism, disease, or natural substance
- A cultural or social practice, movement, or art form
- A media work (book, film, TV show, video game title, album, etc.)
- A specific model, version, or product variant that does not represent
  a significant technological advancement in its own right (e.g.
  "iPhone 13" or "Boeing 747" should be removed, but "iPhone" or
  "airplane" should be kept)

When in doubt, prefer KEEP if the article describes a general class of
technology rather than a specific named instance or event.

Respond only with JSON matching the provided schema.
```

We retain articles classified as KEEP for the subsequent extraction of information about each innovation.

IA1.2 Information about Technological Innovations

We use the following prompt to summarize basic information about the technological inventions with Gemini-3.0-Flash, including the time and location of the invention, as well as the type of inventor (i.e., private company, individual inventor, government, or university/non-profit).

```
Instructions:
```

You are an expert in the history of technologies. Your task is to answer questions about when, where, who invented the technology, and which category of application it best fits into, based on the given information about the technology and your domain knowledge.

Provide the following information along with a brief reason (less than 20 words) for each:

1) When the technology first became workable. If the exact year is unavailable or uncertain, provide the closest approximation. Be as accurate as possible. Include "BCE" in your response if the date is before 0 CE/very ancient.

2) Where it was first created (country where the first workable version was created).

3) Who invented it (choose one of these categories by the type of inventor):
- "Private Company": For-profit company/companies or people working for these companies developed the technology (e.g. Apple, Google);
- "University/Non-Profit": Non-profit institution or organization developed it, such as universities, non-profit institutions, open-source projects (e.g. UC Berkeley, SRI, Mozilla);
- "Government Project": Developed via government-funded or military project (e.g. NASA, the Apollo program, the Manhattan Project);
- "Individual Inventor": Developed by an individual or a small team of inventors unaffiliated with large institutions or companies (e.g. Nikola Tesla, Philo Farnsworth);

Note: Some inventions involve collaboration. When multiple parties are involved, choose the category that best reflects the majority of developers or the most responsible entity. Aim for the most accurate classification based on available evidence. When inventors are not explicitly mentioned, use your best judgment to classify the technology based on the context and purpose of the invention.
For example:
- A collaboration between private companies is still classified as "Private Company".

```
- A small team of unaffiliated inventors is still "Individual Inventor".
- If an individual works under a government-funded initiative, classify it
as "Government Project".
- If it was published by an industry association, it should be either
"Private Company" or "University/Non-Profit" depending on the inventor and
purpose of the technology. For example, SAE (Society of Automotive
Engineers) is mostly ran by engineers from private companies, so it should
be classified as "Private Company". W3C (World Wide Web Consortium) is ran
by university researchers, so it should be classified as
"University/Non-Profit".
- If it was designed for commercial use or production efficiency, it's
likely "Private Company".
- If it served military, national defense, or public interest, it's likely
"Government Project".
- If it was developed for academic research or knowledge advancement, it's
likely "University/Non-Profit".
- If it was created by individuals not working for large institutions, it's
likely "Individual Inventor".

4) The `domain_of_application` (choose exactly one from the list below):
{domain_values}

Importantly: Use only the listed categories above. Do not invent new
categories.

You will receive one JSON object containing the technology name and
Wikipedia page text. Return exactly one JSON object matching the provided
schema. Do not return markdown, comments, citations, or extra keys. Preserve
`page_id` and `canonical_title` exactly as supplied.

Set `answer_source` to {answer_source_values}. Use `wikipedia` if the key
answers are sourced from the supplied Wikipedia text. Use `domain_knowledge`
if any key answer is approximated from your domain knowledge because the
Wikipedia text is missing, ambiguous, or incomplete.

Return fields:
{{
```

```
  "page_id": input page_id,
  "canonical_title": input canonical_title,
  "technology": "Title",
  "when": year or "Not Available",
  "when_reason": "Brief reason for chosen year",
  "where": country name or "Not Available",
  "where_reason": "Brief reason for chosen country",
  "who": one of the 4 listed inventor categories,
  "who_reason": "Brief reason for chosen category",
  "domain_of_application": one of the 15 listed categories,
  "domain_reason": "Brief reason for chosen domain",
  "answer_source": "wikipedia" or "domain_knowledge"
}}

Example Output #1:
{{
  "page_id": 123,
  "canonical_title": "IPhone",
  "technology": "iPhone",
  "when": "2007",
  "when_reason": "Apple released the first iPhone in 2007.",
  "where": "USA",
  "where_reason": "Apple developed the iPhone in the United States.",
  "who": "Private Company",
  "who_reason": "Developed by Apple, which is a for profit company.",
  "domain_of_application": "Information",
  "domain_reason": "Primarily used for communication and computing.",
  "answer_source": "wikipedia"
}}

Example Output #2:
{{
  "page_id": 456,
  "canonical_title": "CRISPR_gene_editing",
  "technology": "CRISPR",
  "when": "2012",
  "when_reason": "First practical demonstration of gene editing was in
  2012.",
```

```
  "where": "USA",
  "where_reason": "Initial research conducted at UC Berkeley.",
  "who": "University/Non-Profit",
  "who_reason": "Developed at UC Berkeley by academic researchers.",
  "domain_of_application": "Health Care and Social Assistance",
  "domain_reason": "Primarily applied to medicine and biotechnology.",
  "answer_source": "wikipedia"
}}
```

IA2 Processing Wikipedia Articles for Mathematical Discoveries

This appendix describes in detail how we obtain and process Wikipedia articles on mathematical discoveries used in Section 6.

IA2.1 Wikipedia Articles on Mathematical Discoveries

We obtain English Wikipedia articles and their topic tags from the Wikimedia English Wikipedia Dumps (Wikimedia Foundation, 2025). To construct the initial candidate set, we retain all articles assigned the mathematics STEM topic tag. Because this broad classification includes many articles that are related to mathematicians, historical events, institutions, books, notation, mathematical games, and other topics that do not constitute mathematical discoveries, we apply an LLM-based filtering step.

For each candidate article, we provide Gemini-3.0-Flash with its title, lead extract, and short description. The LLM classifies the article as either `KEEP` or `REMOVE`. An article is retained if it describes a mathematical discovery—a named theorem, lemma, identity, inequality, conjecture or hypothesis, method or algorithm, probability distribution or statistical test, or a newly introduced mathematical structure, object, or concept—that constitutes a specific, datable contribution first established by identifiable people. The filtering prompt is reproduced below.

```
You are filtering a list of Wikipedia articles to identify mathematical
discoveries---named theorems, lemmas, identities, inequalities, conjectures
or hypotheses, methods or algorithms, probability distributions or
statistical tests, or newly introduced mathematical structures, objects, or
concepts, that constitute specific, datable contributions first established
by identifiable people.
```

```
Given an article title, lead extract, and short description, respond with
only KEEP or REMOVE.

Remove the article if it is primarily about any of the following:

- A person, biography, or group/school of mathematicians (mathematician,
statistician, logician, collectives such as Bourbaki)
- A historical event, controversy, priority dispute, or
history-of-mathematics survey (e.g. the Leibniz--Newton calculus
controversy, "History of geometry")
- An institution, society, university department, journal, prize, award,
competition, or conference (e.g. Fields Medal, Annals of Mathematics,
International Mathematical Olympiad)
- A book, textbook, publication, or piece of software
- A physical instrument or calculating device (e.g. abacus, slide rule)
- A notation, symbol, or unit of measurement
- A specific individual number (e.g. "285 (number)", "1729 (number)")
- A mathematical game, puzzle, or recreational or educational topic (e.g.
Sudoku, mathematics education)---but note that named paradoxes (e.g.
Russell's paradox) are genuine results and should be kept
- A list, glossary, outline, timeline, index, or disambiguation page

When in doubt, prefer KEEP if the article could plausibly describe a
theorem, result, method, concept, structure, or object---even broad areas of
mathematics should be kept.

Respond only with JSON matching the provided schema.
```

We retain articles classified as KEEP for the subsequent extraction of information about each result.

Some retained articles describe conjectures, hypotheses, or open problems that have not been proved (e.g., the Riemann hypothesis). After extracting information about each result (Appendix IA2.2), we therefore apply a final LLM-based filtering step to retain only established results. We provide Gemini-3.0-Flash with each article's title and short description, together with the information extracted in Appendix IA2.2, and the LLM classifies the result as either KEEP or REMOVE. A result is retained if it has been proved, rigorously introduced, or otherwise established; it is removed if it remains open, unproven, or has been disproved.

IA2.2 Information about Mathematical Discoveries

We use the following prompt to summarize basic information about the mathematical discoveries with Gemini-3.0-Flash, including the time and location where each result was first established, the type of contributor (i.e., private company, university/non-profit, government, or individual), and the subfield of mathematics.

```
Instructions:
You are an expert in the history of mathematics. Your task is to answer
questions about when, where, who first established a mathematical result,
and which subfield of mathematics it best fits into, based on the given
information about the result and your domain knowledge.

Provide the following information along with a brief reason (less than 20
words) for each:

1) When the result was first established (proved or introduced). If the
exact year is unavailable or uncertain, provide the closest approximation.
Be as accurate as possible. Include "BCE" in your response if the date is
before 0 CE/very ancient. For a conjecture or open problem proved later, use
the year it was first proved, not the year it was conjectured. If it has
never been proved, answer "Not Available".

2) Where it was first established (country where it was first proved or
introduced).

3) Who established it (choose one of these categories by the type of
contributor):
- "Private Company": an industrial or corporate research lab (e.g. Bell
Labs, IBM Research, Microsoft Research).
- "University/Non-Profit": a university, institute, or non-profit research
effort (most academic mathematics).
- "Government Project": a government-funded, national-laboratory, or
classified effort (e.g. a national lab, NSA).
- "Individual": an individual unaffiliated with an institution (e.g.,
university, institute, company, or government body).
```

Note: classify by the contributor's institutional affiliation. When multiple parties or institutions are involved, choose the one that best reflects the majority of contributors or the most responsible entity. Aim for the most accurate classification based on available evidence. When contributors are not explicitly mentioned, use your best judgment to classify the result based on the context and purpose of the work.
For example:
- A collaboration among researchers at universities is still "University/Non-Profit".
- If a mathematician is affiliated with a university or research institute, use "University/Non-Profit". If the work was established by an individual not affiliated with any institution, use "Individual".
- A result produced at a corporate research lab is "Private Company". For example, information theory and coding results from Bell Labs are "Private Company", while a result from a university mathematics department is "University/Non-Profit".
- If an individual works under a government-run or classified program (e.g. the Monte Carlo method, developed at Los Alamos), classify it as "Government Project".

4) The subfield of mathematics (choose exactly one from the list below):
- "Algebra"
- "Number theory"
- "Analysis"
- "Geometry and topology"
- "Combinatorics"
- "Logic and foundations"
- "Probability and statistics"
- "Applied and computational"

Importantly: Use only the listed categories above. Do not invent new categories. If a result spans several subfields, choose the single area it is most central to.

```
You will receive one JSON object containing the result name and Wikipedia
page text. Return exactly one JSON object matching the provided schema. Do
not return markdown, comments, citations, or extra keys. Preserve `page_id`
and `canonical_title` exactly as supplied.

Set `answer_source` to `wikipedia` or `domain_knowledge`. Use `wikipedia` if
the key answers are sourced from the supplied Wikipedia text. Use
`domain_knowledge` if any key answer is approximated from your domain
knowledge because the Wikipedia text is missing, ambiguous, or incomplete.

Return fields:
{{
  "page_id": input page_id,
  "canonical_title": input canonical_title,
  "result": "Title",
  "when": year or "Not Available",
  "when_reason": "Brief reason for chosen year",
  "where": country name or "Not Available",
  "where_reason": "Brief reason for chosen country",
  "who": one of the 4 listed contributor categories,
  "who_reason": "Brief reason for chosen category",
  "subfield": one of the 8 listed subfields,
  "subfield_reason": "Brief reason for chosen subfield",
  "answer_source": "wikipedia" or "domain_knowledge"
}}

Example Output #1:
{{
  "page_id": 123,
  "canonical_title": "Gödel's_incompleteness_theorems",
  "result": "Gödel's incompleteness theorems",
  "when": "1931",
  "when_reason": "Published by Gödel in 1931.",
  "where": "Austria",
  "where_reason": "Gödel worked in Vienna.",
  "who": "University/Non-Profit",
  "who_reason": "Gödel worked at the University of Vienna.",
  "subfield": "Logic and foundations",
```

```
  "subfield_reason": "Concerns the foundations of mathematics.",
  "answer_source": "domain_knowledge"
}}

Example Output #2:
{{
  "page_id": 456,
  "canonical_title": "Fermat's_Last_Theorem",
  "result": "Fermat's Last Theorem",
  "when": "1994",
  "when_reason": "First proved by Wiles in 1994, not the 1637 conjecture.",
  "where": "United States",
  "where_reason": "Proved at Princeton.",
  "who": "University/Non-Profit",
  "who_reason": "Wiles worked at Princeton University.",
  "subfield": "Number theory",
  "subfield_reason": "A statement about integers.",
  "answer_source": "domain_knowledge"
}}

Example Output #3:
{{
  "page_id": 789,
  "canonical_title": "Riemann_hypothesis",
  "result": "Riemann hypothesis",
  "when": "Not Available",
  "when_reason": "Never proved; remains an open conjecture.",
  "where": "Not Available",
  "where_reason": "No proof exists.",
  "who": "Not Available",
  "who_reason": "No one has proved it yet.",
  "subfield": "Number theory",
  "subfield_reason": "Concerns the distribution of primes.",
  "answer_source": "domain_knowledge"
}}

Example Output #4:
{{
```

```
  "page_id": 1011,
  "canonical_title": "Fermat's_spiral",
  "result": "Fermat's spiral",
  "when": "1636",
  "when_reason": "Described by Fermat around 1636.",
  "where": "France",
  "where_reason": "Fermat worked in Toulouse.",
  "who": "Individual",
  "who_reason": "Fermat was a magistrate who did mathematics privately, not
  at an institution.",
  "subfield": "Geometry and topology",
  "subfield_reason": "A plane curve.",
  "answer_source": "domain_knowledge"
}}
```


IA3 Processing Wikipedia Articles for Artwork

This appendix describes in detail how we obtain and process Wikipedia article titles on artwork used in Section 6.

IA3.1 Wikipedia Articles on Artwork

We extract English Wikipedia article titles from the Wikimedia English Wikipedia Dumps (Wikimedia Foundation, 2025) and filter them in two steps to identify articles related to artwork.

Step 1 Unlike technological inventions, which span a wide and heterogeneous range of topics that are difficult to delineate with simple rules, artwork on Wikipedia is systematically organized into well-defined categories corresponding to artistic genres and media (e.g., "paintings," "novels," "films," "sculptures"). We exploit this structure by performing keyword-based filtering on the Wikipedia categories associated with each title to obtain a broad set of artwork-related titles. We retain titles whose categories contain at least one of the following keywords: painting, sculpture, photography, film, music, literature, performing arts, architecture, installation, digital art, ceramics, opera, ballet, theatre, fresco, mural, portrait, landscape painting, still life, oil painting, watercolor, ink painting, miniature painting, illumination, icon painting, altarpiece, triptych, diptych, statue, bust, relief, carving, bronze sculpture, marble sculpture, wood sculpture, terracotta, monument, photograph, photographic print, daguerreotype, photomontage, decorative arts, visual arts, fine arts, craft, folk art, applied arts, design, iconography, artifact, masterpiece, among others.

Step 2 We validate each Wikipedia article's title from Step 1 using its Wikipedia categories and summary to obtain the final set of artwork. For each title, we use the Wikipedia API library to retrieve the corresponding English Wikipedia article's summary. We then supply the title, categories, and summary to the DeepSeek-Chat model, which is prompted to determine whether the article indeed describes a single, concrete artwork. The LLM's responses indicate whether the Wikipedia article should be retained.

```
You are a strict JSON-only classifier.

Task:
Determine whether the given Wikipedia entry refers to a single, concrete
artwork.

Definition:
A specific creative work in painting, sculpture, photography, film, music,
literature, performing arts, architecture, installation, digital art,
ceramics, opera, ballet, theatre, or other artistic media.
Examples: ''Mona Lisa'', ''The Starry Night'', ''Hamlet'', ''The Rite of
Spring'', ''Forrest Gump'', ''Bohemian Rhapsody'', ''Swan Lake''.

Exclude:
- People, concepts, events, movements, genres, or general terms (e.g.
Romanticism, Jazz, Painting in China).
- Organizations, companies, locations, periods.
- Generic categories that are not a specific, notable artwork.

Return ONLY a JSON object like:
{{''is_valid'': true, ''explanation'': ''short reason''}}

Title: {title}

Categories:
\''\''\''{categories}\''\''\''

Summary:
\''\''\''{summary}\''\''\''
```

IA3.2 Artwork Information

We use the following prompt to summarize basic information about the artwork with Gemini-2.0-Flash, including the genre, creator, year and location of creation.

```
You are an expert cataloguer of artworks. Your task is to extract structured
facts about the artwork's title, artist, creation date and location, and
genre, based on the given information about the artwork and your domain
knowledge.

Provide the following information:

1) Title of the artistic piece

2) Genre of the artwork: Choose EXACTLY ONE genre from this fixed set (do
not invent new labels):
[''Painting'', ''Sculpture'', ''Photography'', ''Film'', ''Music'',
''Literature'', ''Performing Arts'', ''Architecture'', ''Installation'',
''Digital Art'', ''Ceramics'', ''Opera'', ''Ballet'', ''Theatre'',
''Other'']
If the work spans multiple forms, choose the best primary genre.

3) Creator(s) of the artwork: Include the primary credited creator(s). If
multiple, join names with ''; ''

4) Creator's country of origin: If multiple locations are involved, choose
the most significant one. If unclear, return ''Not Available''. If multiple
creators, choose the country of origin of the most primary creator. If equal
prominence, choose the country associated with where that creator conducted
most of their career-defining work. If still unclear, return ''Mixed''.

5) Year of creation: If the exact year is unavailable or uncertain, provide
the closest approximation. Be as accurate as possible. Include ''BCE'' in
your response if the date is before 0 CE/very ancient.

6) Country of creation: The country where the work was principally
created/produced. If multiple locations, pick the primary
production/creation location. If unclear, ''Not Available''.
```

```
Title: {title}

Wikipedia Page:
\''\''\''{wiki_text[:5000]}\''\''\''

Output strictly as JSON (no backticks, no preamble, no trailing text),
schema:
{{
  ''piece_name'': ''str'',
  ''genre'': ''str'',
  ''creator'': ''str'',
  ''creator_country_of_origin'': ''str'',
  ''year_of_creation'': ''str'',
  ''country_of_creation'': ''str''
}}
```

IA4 Processing University Lists

This appendix describes how we construct the country-year measure of universities per capita used in the analysis.

IA4.1 University Lists by Country

We construct country-specific lists of universities using web-grounded LLM queries. For each country in our main country-year panel, we first ask Gemini-2.5-Flash, with Google Search enabled, to estimate the total number of universities in that country. We define universities as nationally recognized higher-education institutions that grant undergraduate, bachelor-level degrees. We include public and private universities, institutes of technology, polytechnics, and universities of applied sciences when they have university status or award bachelor-level degrees. We exclude graduate-only or research-only institutes, stand-alone government research centers, national laboratories, administrative agencies, training centers, community or junior colleges without university status, and branch campuses. We deduplicate institutions by institutional lineage, so that renamings and mergers are counted once.

The target count is used to guide the construction of a comprehensive university list for each country. We instruct the LLM to rely on authoritative sources, including education ministry registries, national accreditation or quality-assurance agencies, WHED/UNESCO, official university registries, and official university websites. If an exact official count is unavailable, the LLM is instructed to return the best-supported lower-bound count based on authoritative lists.

```
You must use web search (Google Search tool is available to you) to answer.
You are estimating the total historical count of universities that grant
undergraduate
(bachelor-level, ISCED 6) degrees in a country.

Scope & definition:
- Include: institutions officially recognized as universities, or national
equivalents
  such as universities of applied sciences and polytechnics where they award
  bachelor degrees.
- Exclude: graduate-only or research-only institutes, stand-alone national
labs,
  training centers, seminaries that do not award bachelor degrees,
  community/junior
  colleges that lack university status or do not award bachelor degrees, and
  branch campuses.
- Mergers/renamings: count by institutional lineage once.

Evidence policy:
- Prefer official sources: education ministry registries, national
accreditation/QA agencies,
  UNESCO/UIS, WHED, official university registries, and university websites.
- If an exact total is unavailable, provide the best-supported lower-bound
integer.
- Do not hallucinate. If uncertain, return your best supported single
integer.

Country: {country}

Return only JSON:
{{
  "country": "{country}",
  "institutions_count": 1234
}}
```

We then ask Gemini-2.5-Flash to return the list of universities for each country. The model is instructed to return native-language university names, deduplicate branch campuses, and aim for the target count obtained in the previous step.

```
You must use web search (Google Search tool is available to you) to answer.

Task:
Return the list of universities only in {country}, meaning higher-education
granting institutions recognized nationally.

- Include universities, including institutes of technology that are
universities in practice.
- Exclude independent research institutes that do not award university
degrees,
  primary/secondary schools, training centers, and administrative bodies.
- Deduplicate branch campuses; keep the main university entity once.
- Return only native-language names, without English translations.
- Your target is to return approximately {target_count} unique universities.

Output strictly as JSON:
{{
  "country": "{country}",
  "universities": [
    "native university name 1",
    "native university name 2"
  ]
}}
```

For countries where the initial list was incomplete, we use additional iterative LLM calls that provide the already collected names and ask the model to return additional universities not yet in the list. For Russia and the United States, where general LLM search performed less well because of the large number of institutions, we supplement the list using special country-specific sources.

IA4.2 University Information

After obtaining country-level university lists, we collect institution-level information for each university. We ask Gemini-2.5-Flash, with Google Search enabled, to retrieve the English name, city, founding year, official website, and source URLs for each institution. The founding year is the key variable for constructing the historical stock of universities.

```
You must use web search (Google Search tool is available to you).

Given a university, retrieve standardized details.

Country context: {country}
University native name: {name_native}

Return strictly as JSON:
{{
  "name_native": "{name_native}",
  "name_en": "English name or Not Available",
  "city": "main city or Not Available",
  "founding_year": "YYYY or Not Available",
  "official_website": "https://... or Not Available",
  "source_urls": ["https://...", "https://..."]
}}

Rules:
- "founding_year" should be a four-digit year if reliably stated; otherwise
return
  "Not Available".
- "official_website" should be the institution's official homepage.
- "source_urls" should contain one to three authoritative URLs used, such as
ministry
  registries, accreditation agency lists, WHED/UNESCO, official university
  websites,
  or reputable encyclopedia pages.
- Return JSON only.
```

For institutions with missing founding years, we further search Wikipedia. First, for each country, we identify an appropriate Wikipedia list page, preferably in an official or national language of the country, falling back to English Wikipedia when necessary. We then use the MediaWiki search API to search for each university name, keep the closest title match, retrieve and clean the relevant Wikipedia page text, and ask Gemini-2.0-Flash-Lite to extract the founding year only from the provided text.

```
You are given a cleaned Wikipedia-like page text for a university.
```

```
Do NOT browse or search. Extract ONLY from the provided text.

Context:
- Country: {country}
- University native name: {name_native}
- Title from search: {title}

Text:
'''{page_clean_text[:20000]}'''

Return strict JSON with exactly these keys:
{{
  "city": "str",
  "founding_year": "YYYY",
  "official_website": "https://..."
}}

Rules:
- "city": main city if clearly stated. If multiple cities are mentioned,
choose
  the primary campus or home city.
- "founding_year": four-digit year only if clearly stated.
- "official_website": full URL only if explicitly present in the text.
- If the information is not clearly available, return "Not Available".
- No explanations or extra text.
```

Finally, we merge the web-grounded LLM output and the Wikipedia-based extraction. When the original founding year is missing, we fill it using the Wikipedia-extracted founding year if available. We then drop universities without a usable founding year and collapse the data to the country-year level.

IA5 Additional Results

IA5.1 Innovation Counts over Time

Figure IA1 plots the total number of innovations worldwide over the past 50 years (left axis, solid line) and innovations per million people (right axis, dashed line) from 0 to 1999. The total count rises sharply from the 18th century onward, consistent with the acceleration of technological progress during the Industrial Revolution and after. The per capita series increases as well, though it levels off in the second half of the 20th century as global population growth catches up.

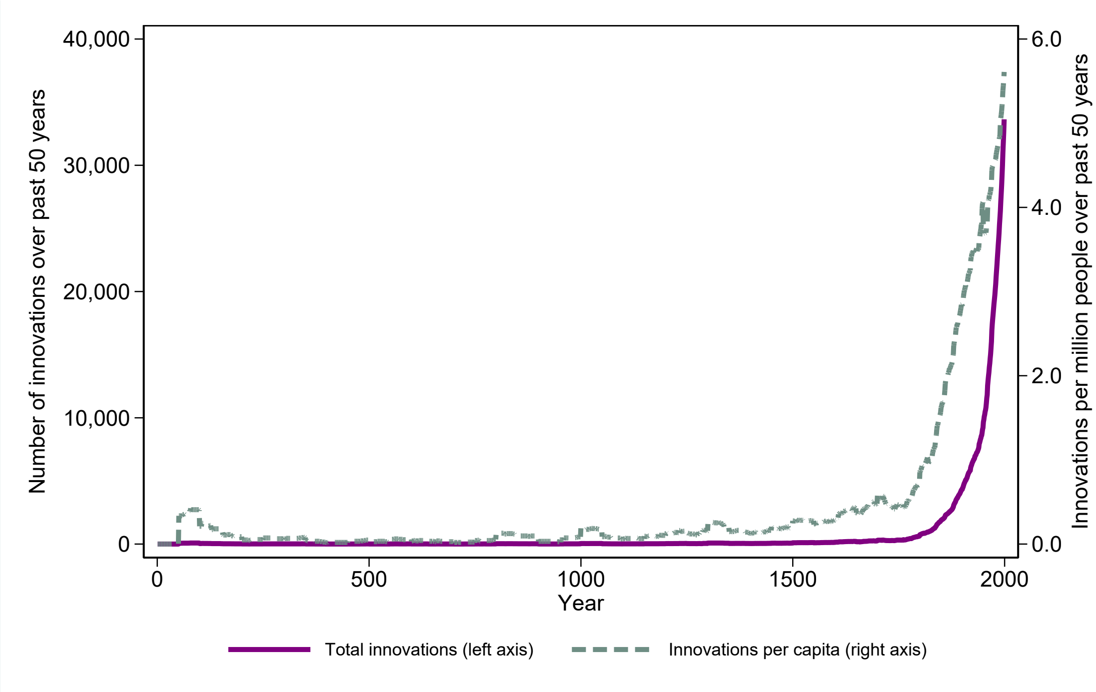
IA5.2 Cross-Checking Wikipedia Innovations against Patent Data

Figure IA2 presents a cross-sectional scatter plot of country-level innovation percentiles (based on Wikipedia innovations per capita) against patent percentiles (based on patent applications per capita from PATSTAT), pooling observations from 1970 to 1999. The strong positive relationship confirms that the Wikipedia-based measure aligns well with patent data. English-speaking countries—defined as the United States, the United Kingdom, Australia, Canada, Ireland, and New Zealand—do not appear systematically above the regression line, suggesting that English Wikipedia does not disproportionately inflate innovation counts for these countries.

Figure IA3 provides a complementary check by tracking the log ratio of Wikipedia innovations to patent applications over time for nine major innovating countries from 1960 to 1999. If English Wikipedia systematically over-represented English-speaking countries, the United States and the United Kingdom should stand together at the top. The data show otherwise. The United States ratio is elevated—roughly 3 to 7 Wikipedia innovations per 1,000 patents through most of the period, and trending mildly downward. The United Kingdom, however, sits in the same range as France, Germany, and Italy, hovering close to 1 innovation per 1,000 patents. Italy through the 1960s and the Netherlands by the late 1990s actually exceed the United States. Spain sits well below the rest, as does Japan—though Japan's low ratio partly reflects its distinctive patent system, which historically required separate filings for each narrow claim and continues to produce many more patent applications per underlying invention than the United States or European systems.[11] The cross-country variation in this ratio thus reflects differences in how research output translates into broad, milestone inventions notable enough to merit a Wikipedia entry—and, in Japan's case, differences in the patent system itself—rather than the language in which inventions are catalogued. Taken with the scatter plot in Figure IA2, the evidence suggests that the geographic patterns documented in the main text are not artifacts of English-language coverage on

[11] Until the 1988 reform, Japanese patent law effectively required one invention per application, so a single multi-claim United States or European patent often corresponded to 10 or more Japanese filings; even after the reform, Japanese firms file more patents per invention than firms elsewhere (Sakakibara and Branstetter, 2001).

Figure IA1. Number of Innovations over Time



Notes: The figure plots the global number of innovations over the past 50 years (left axis, solid line) and innovations per million people over the past 50 years (right axis, dashed line), from 0 to 1999. Innovations are counted using the Wikipedia-based dataset described in Section 2.

Wikipedia.

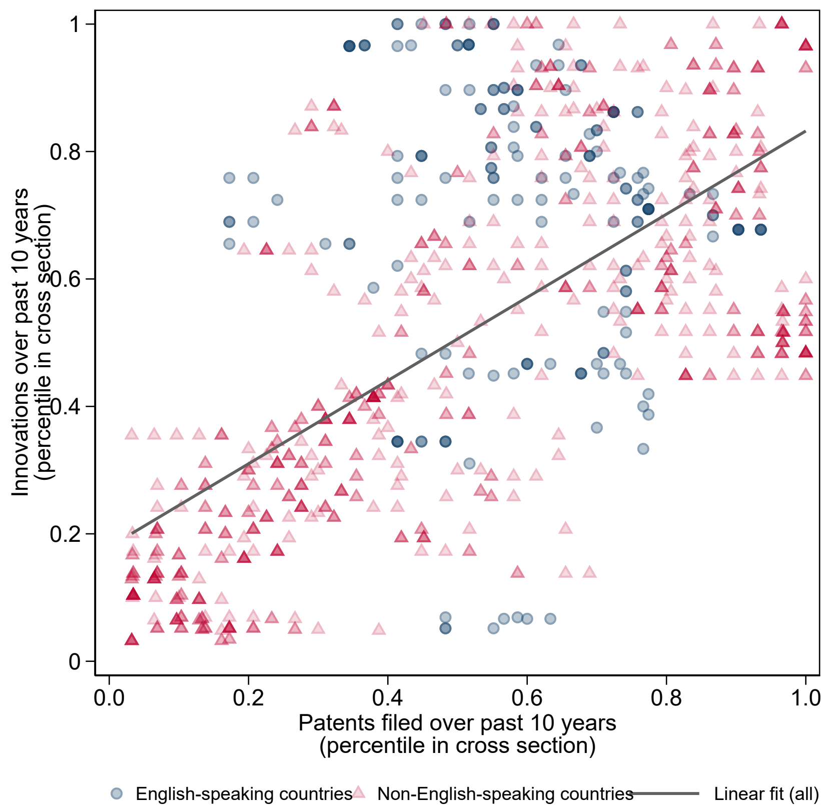
IA5.3 Concentration of Innovation: Herfindahl Indices

This appendix collects the Herfindahl-Hirschman concentration indices discussed in Section 3.1, at the continent level (Figure IA4) and the country level (Figure IA5).

IA5.4 Country Characteristics and Innovation Intensity

We explore which observable country characteristics are associated with higher innovation intensity by regressing innovations per capita on social, economic, and financial development indicators for 17 advanced economies over the period 1870 to 1999 (Table IA1). All variables are standardized, so coefficient magnitudes are directly comparable. Population size has little explanatory power for innovation intensity, and real GDP per capita is a significant correlate across countries but not within them once country fixed effects are included. The variable that stands out is stock market capitalization relative to GDP, which is a consistently positive correlate of innovation intensity both across and within countries. Bank lending, by contrast, shows no significant association, and universities per capita enters

Figure IA2. Wikipedia Innovations vs. Patent Applications across Countries



Notes: The figure plots the cross-sectional percentile of Wikipedia innovations per capita (over the past 10 years) against the percentile of patent applications per capita (over the past 10 years), pooling country-year observations from 1970 to 1999. Countries with fewer than 100 patent applications over the past 10 years are excluded. English-speaking countries (United States, United Kingdom, Australia, Canada, Ireland, and New Zealand) are shown in dark circles; non-English-speaking countries are shown in lighter triangles. The line shows the linear fit using all observations. Patent data are from PATSTAT (European Patent Office, 2024).

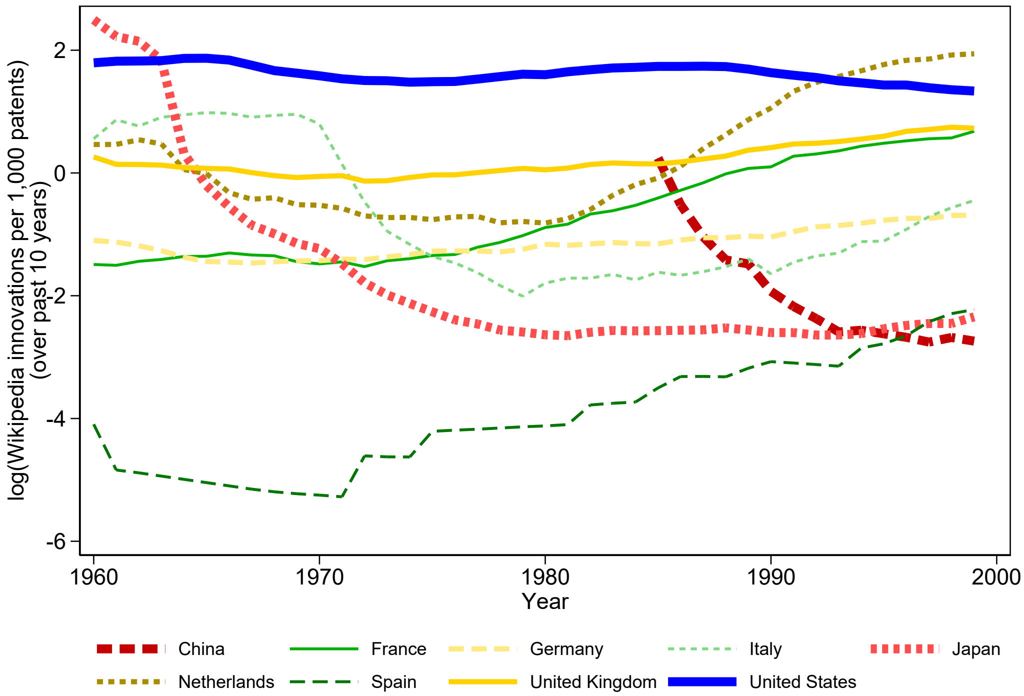
positively but is not statistically significant.

IA5.5 Spillovers from the Leader for Poor vs. Rich Countries

We re-estimate the leader specification of Section 3.2 with the sample split by within-year GDP-per-capita quartile. The regression adds two interactions with the leader's innovation shock, an indicator for the bottom quartile ("Poor") and an indicator for the top quartile ("Rich"); the omitted reference category is the middle two quartiles. We trace the implied impulse responses separately for the bottom and top quartiles, and plot the poor–rich difference together with its confidence interval.

Figure IA6 reports the GDP-per-capita response. The cumulative GDP-per-capita response to the leader's innovations is substantially smaller for bottom-quartile countries than for top-quartile

Figure IA3. Wikipedia Innovations per 1,000 Patents over Time



Notes: The figure plots the log of Wikipedia innovations per 1,000 patent applications—both measured cumulatively over the past 10 years—for nine major innovating countries from 1960 to 1999. Country lines are distinguished by color and line style, as indicated in the legend. Patent data are from PATSTAT (European Patent Office, 2024).

ones, with the gap widening over horizons. The right panel confirms that the poor–rich difference is negative and statistically distinguishable from zero across most horizons. Panel (b) repeats the exercise with an additional control for log geographic distance to the leader, interacted with the shock; the gradient is essentially unchanged, indicating that physical proximity to the leader does not account for the heterogeneity.

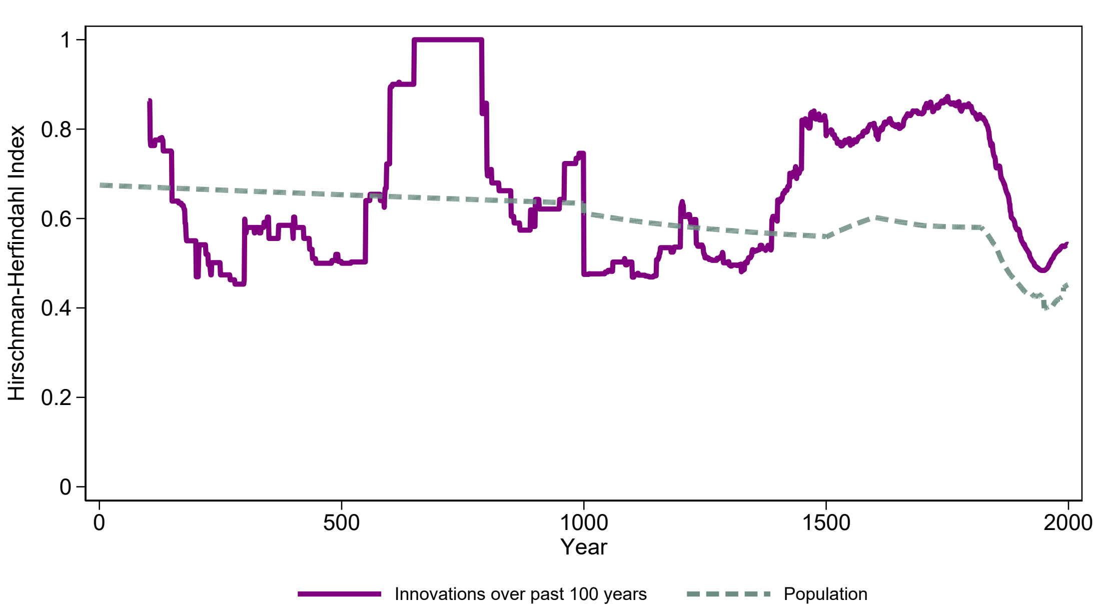
IA5.6 Innovation Persistence and Spillovers to Innovation Activity

Figures IA7 and IA8 present local projections analogous to those in Section 3.2, but with future innovation activity as the outcome rather than GDP growth.

IA5.7 Innovation Within and Across Domains

We first ask whether research spillovers diffuse more readily within than across technology domains, and then decompose innovation activity into a global domain-specific component and a local country-

Figure IA4. HHI of Innovation at the Continent Level



Notes: Herfindahl-Hirschman concentration indices for innovation and population at the continent level, from 0 to 1999, with countries aggregated to continents. The sample is the same 63 countries as the country-level figure: those with at least 10 decades of non-missing GDP per capita from 1500 to 1999 (the quantitative estimation sample of Section 5). We use country-years with non-missing population, so that the innovation and population HHIs are computed on the same country-years and are directly comparable. Innovations are cumulative counts over the past 100 years.

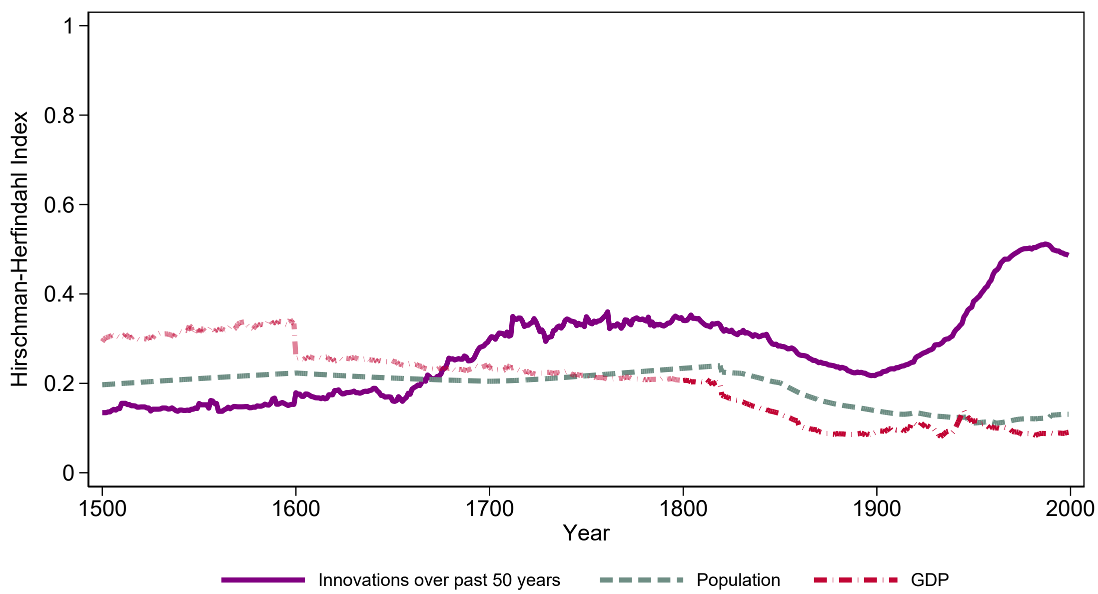
specific one.

Diffusion within and across domains To assess whether technology diffuses more readily within the same domain than across domains, we estimate local projections on a panel in which each country-year is split into six technology domains $d \in \mathcal{D}$ (agriculture, mining, energy & construction; manufacturing; defense; trade & transportation; telecom & information processing; basic science & medicine), and each country–domain–year observation is paired with the global innovation leader's innovation activity in every domain $d' \in \mathcal{D}$. As before, we define every year's *global innovation leader* as the country with the highest 10-year total (all-domain) innovation count. Specifically, we consider local projections of the form

$$x_{n,d,\,t\to t+h} = \gamma_{n,d,h}+\gamma_{n,t,h}+\mathbf{1}\{d=d'\}\,\beta_h\,x_{l,d',t}+\mathbf{1}\{d\neq d'\}\,\delta_h\,x_{l,d',t}+\mu_h\,\mathbf{1}\{d=d'\}+\boldsymbol{\psi}_h'\boldsymbol{z}_{n,d,t}+\epsilon_{n,d,d',t,h},$$

where n indexes the country, d the domain of the country's own innovation activity, d' the domain of the leader's innovation activity, and h the prediction horizon; a separate regression is estimated for each h. The outcome $x_{n,d,\,t\to t+h}$ is the country's total innovations in domain d from year t to $t+h$, relative to the cumulative world innovation stock. $x_{l,d',t}$ is the global innovation leader's total innovations in

Figure IA5. HHI of Innovation at the Country Level



Notes: Herfindahl-Hirschman concentration indices for innovation, population, and GDP at the country level, from 1500 to 1999. The sample is the 63 countries with at least 10 decades of non-missing GDP per capita from 1500 to 1999 (the quantitative estimation sample of Section 5). We use country-years with non-missing population, so that the innovation and population HHIs are computed on the same country-years and are directly comparable. Innovations are cumulative counts over the past 50 years; the GDP HHI is shown with a lighter color before 1800, where Maddison Project coverage is sparser.

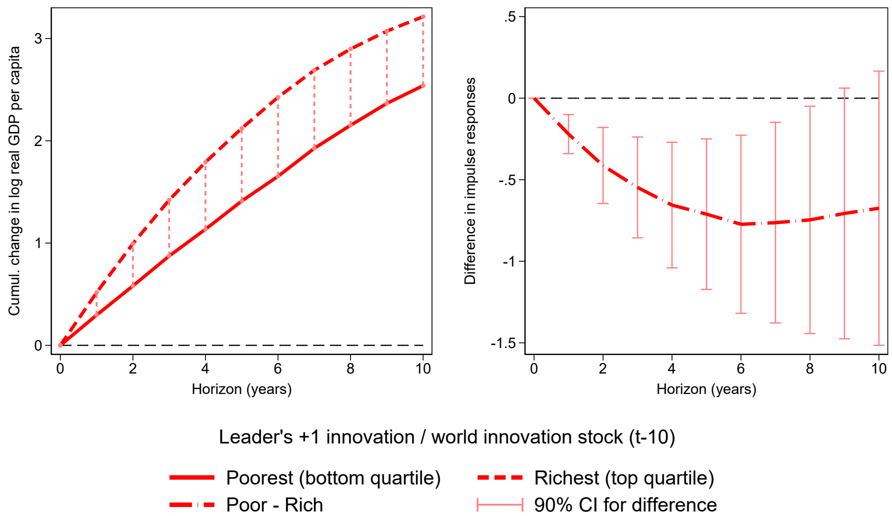
domain d' over the past 10 years, relative to the cumulative world innovation stock 10 years earlier. Thus, β_h captures within-domain spillovers, and δ_h captures cross-domain spillovers. The control vector $\boldsymbol{z}_{n,d,t}$ contains $x_{n,d,t}$ and $x_{n,t}$, the country's own innovations in domain d and its own total innovations over the past 10 years, relative to the cumulative world innovation stock 10 years earlier. We include country$\times$domain ($\gamma_{n,d,h}$) and country$\times$year ($\gamma_{n,t,h}$) fixed effects, cluster standard errors by country and year, and exclude the leader itself from the sample. Figure IA9 plots the impulse responses estimated by $\hat{\beta}_h$ and $\hat{\delta}_h$.

Decomposing local and global factors The main domain of global innovation activity changes over time. Figure IA10 illustrates the evolution from defense-focused innovation before 1600, through the scientific revolution, the industrial revolution starting in the 18th century, the automotive revolution in the early 20th century, and the spurt in telecommunications and information processing in recent decades.

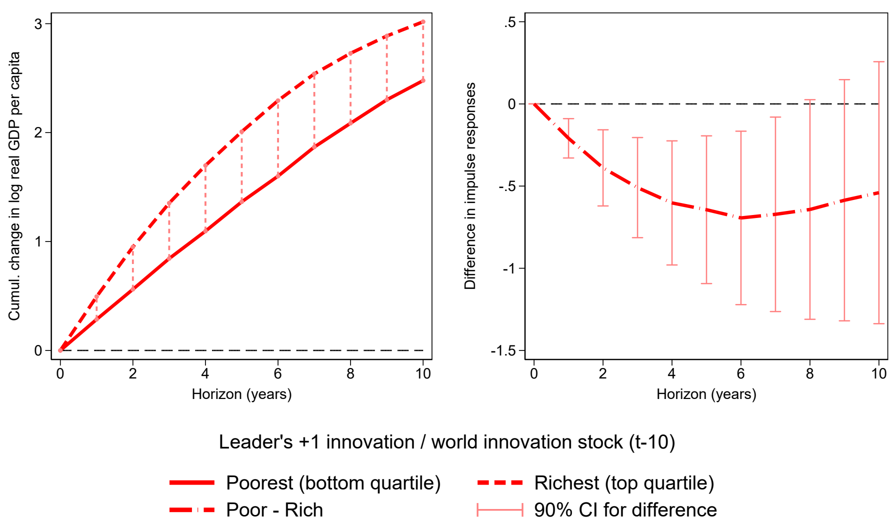
Is innovation concentration driven by a global domain-specific component, or by local country-specific factors? We estimate the following fixed effects regression of the logarithm of country n's

Figure IA6. Leader's Innovations and Future Local Growth: Poor vs. Rich Countries

(a) Baseline



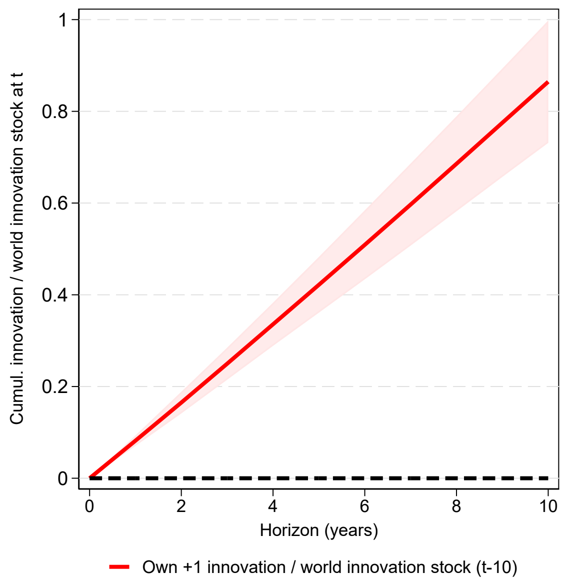
(b) Controlling for Log Distance to Leader



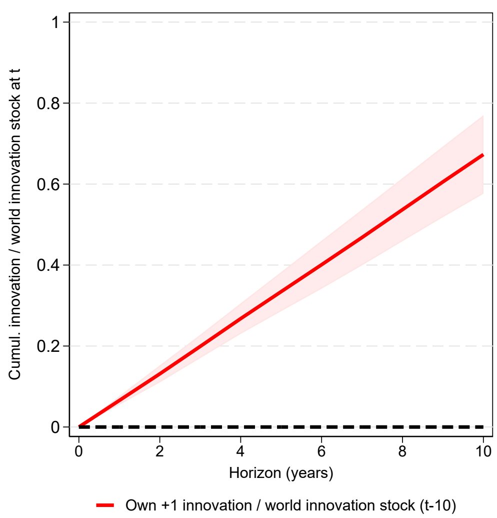
Notes: The figure shows changes in a country's future real GDP per capita following increases in the global innovation leader's activity, separately for "poorest" and "richest" countries. In each year t, country n is classified as "Poor" ("Rich") if its real GDP per capita falls in the bottom (top) quartile of the cross-country distribution that year; the middle two quartiles are the omitted reference. Estimates come from local projections: $\Delta_h \log(\text{real GDP per capita}_{n,t+h}) = \gamma_{n,h} + \beta_h\, x_{l,t} + \beta_{P,h}\, x_{l,t} \cdot \text{Poor}_{n,t} + \beta_{R,h}\, x_{l,t} \cdot \text{Rich}_{n,t} + \boldsymbol{\delta}_h' \boldsymbol{w}_{n,t} + \boldsymbol{\psi}_h' \boldsymbol{z}_{n,t} + \epsilon_{n,t,h}$. The left panel of each subfigure plots the implied impulse responses for poor and rich countries; the right panel plots the poor–rich difference (bottom minus top). The outcome is the cumulative change in log real GDP per capita from year t to $t+h$. $x_{l,t}$ is the global innovation leader's total innovations over the past 10 years relative to the cumulative world innovation stock 10 years earlier. The control vector $\boldsymbol{z}_{n,t}$ contains 10 lags and the contemporaneous annual real GDP per capita growth, the level of log real GDP per capita, and the country's own innovations over the past 10 years relative to the cumulative world innovation stock 10 years earlier, $\boldsymbol{w}_{n,t}$ contains $\text{Poor}_{n,t}$ and $\text{Rich}_{n,t}$ themselves. Panel (b) additionally includes log distance to the leader and its interaction with $x_{l,t}$, with the implied impulse responses evaluated at the within-sample mean of log distance (the poor–rich difference is invariant to the evaluation point). The sample covers 1500 to 1999 and excludes the leader country. GDP data are from the Maddison Project Database (Bolt and van Zanden, 2024). Capped spikes mark 90% confidence intervals based on standard errors double clustered by country and year.

Figure IA7. Own Innovations and Future Innovation Activity

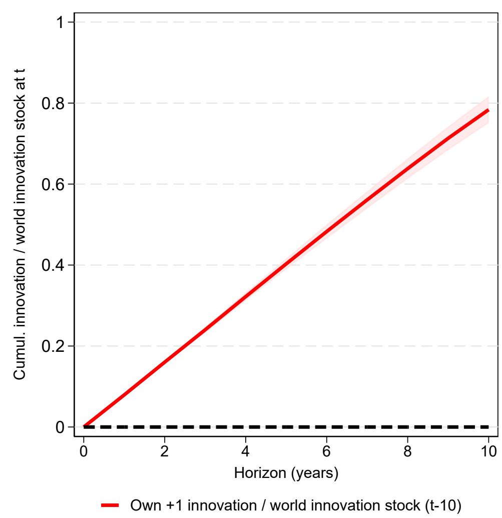
(a) Full Sample



(b) Sample 1500 to 1899



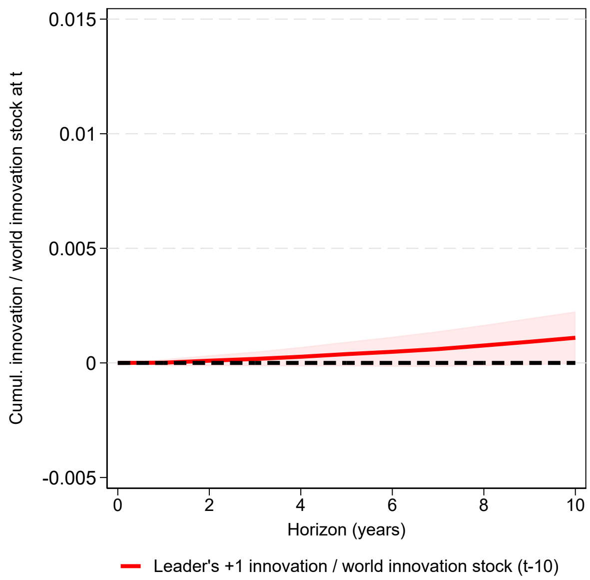
(c) Sample 1900 to 1999



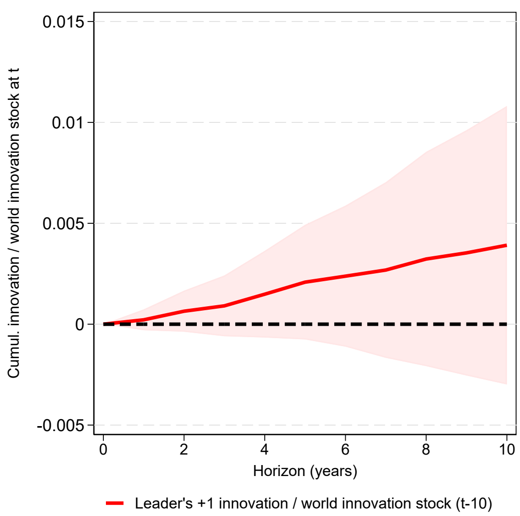
Notes: The figure shows changes in a country's future innovation activity following increases in its own past innovation activity, estimated using local projections: $x_{n,\,t\rightarrow t+h} = \gamma_{n,h} + \beta_h\, x_{n,t} + \epsilon_{n,t,h}$, where n indexes the country and h the prediction horizon; a separate regression is estimated for each h. The outcome $x_{n,\,t\rightarrow t+h}$ is the country's total innovations from year t to $t+h$, relative to the cumulative world innovation stock. $x_{n,t}$ is the country's total innovations over the past 10 years, relative to the cumulative world innovation stock 10 years earlier. Panel (a) shows estimates for the full sample (1500 to 1999); Panel (b) for the historical period (1500 to 1899); Panel (c) for the modern period (1900 to 1999). Shaded areas mark 90% confidence intervals based on standard errors double clustered by country and year.

Figure IA8. Leader's Innovations and Future Local Innovation Activity

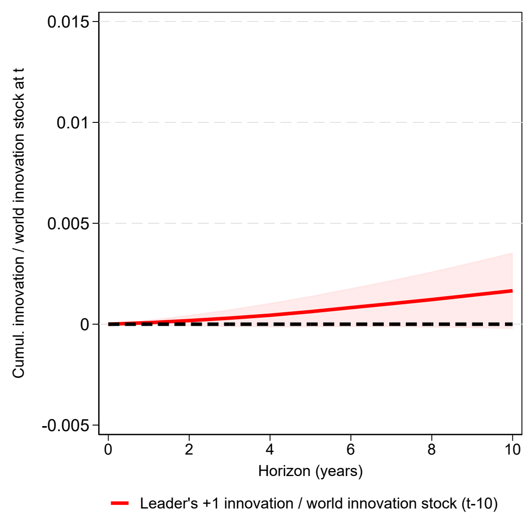
(a) Full Sample



(b) Sample 1500 to 1899



(c) Sample 1900 to 1999



Notes: The figure shows changes in a country's future innovation activity following increases in innovation activity of the global innovation leader. Estimates come from local projections: $x_{n,\,t\rightarrow t+h} = \gamma_{n,h} + \beta_h\, x_{l,t} + \boldsymbol{\psi}_h' \boldsymbol{z}_{n,t} + \epsilon_{n,t,h}$, where n indexes the country and h the prediction horizon; a separate regression is estimated for each h. The outcome $x_{n,\,t\rightarrow t+h}$ is the country's total innovations from year t to $t+h$, relative to the cumulative world innovation stock. $x_{l,t}$ is the global innovation leader's total innovations over the past 10 years, relative to the cumulative world innovation stock 10 years earlier (the leader is defined as the country with the most total innovations over the past 10 years). The control vector includes the country's own innovations over the past 10 years relative to the cumulative world innovation stock 10 years earlier. The sample excludes the leader country. Panel (a) shows estimates for the full sample (1500 to 1999); Panel (b) for the historical period (1500 to 1899); Panel (c) for the modern period (1900 to 1999). Shaded areas mark 90% confidence intervals based on standard errors double clustered by country and year.

Table IA1 – Correlates of Total Innovation per Population

	Innovation per capita			
	(1)	(2)	(3)	(4)
Log population	-0.208	4.086**	-0.198	4.003
	(0.266)	(1.801)	(0.257)	(2.943)
Log real GDP per capita	0.498***	0.017	1.332***	-0.344
	(0.133)	(0.371)	(0.449)	(0.638)
Universities per capita	0.338	0.470	0.259	0.577
	(0.269)	(0.459)	(0.222)	(0.477)
Loans/GDP	-0.092	-0.276	0.065	-0.059
	(0.131)	(0.171)	(0.205)	(0.137)
Stock market capitalization/GDP	0.491**	0.378**	0.599*	0.525**
	(0.220)	(0.149)	(0.334)	(0.234)
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
R^2 (within)	0.411	0.567	0.434	0.315
Obs	861	861	856	856

Notes: This table presents estimates from regressions of innovations per capita on social, economic, and financial development indicators: $y_{n,t} = \alpha + \boldsymbol{\beta}'\boldsymbol{x}_{n,t} + \gamma_n + \delta_t + \epsilon_{n,t}$, where $y_{n,t}$ is innovations per capita in country n at year t, $\boldsymbol{x}_{n,t}$ is a vector of country characteristics, and γ_n and δ_t are country and year fixed effects included in some specifications. All variables are standardized (mean zero, unit standard deviation) so that coefficient magnitudes are directly comparable. Given limited data availability on financial development indicators, the sample comprises 17 advanced economies between 1870 and 1999: Australia, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and the United States. Standard errors are double clustered by country and year, shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

trailing 50-year innovations in domain d at year t, denoted innovations$_{n,d,t}$:

$$\log(\text{innovations}_{n,d,t}) = \underbrace{\gamma_{d,t}}_{\text{global domain component}} + \underbrace{\gamma_{n,t}}_{\text{local component}} + \epsilon_{n,d,t}, \tag{IA1}$$

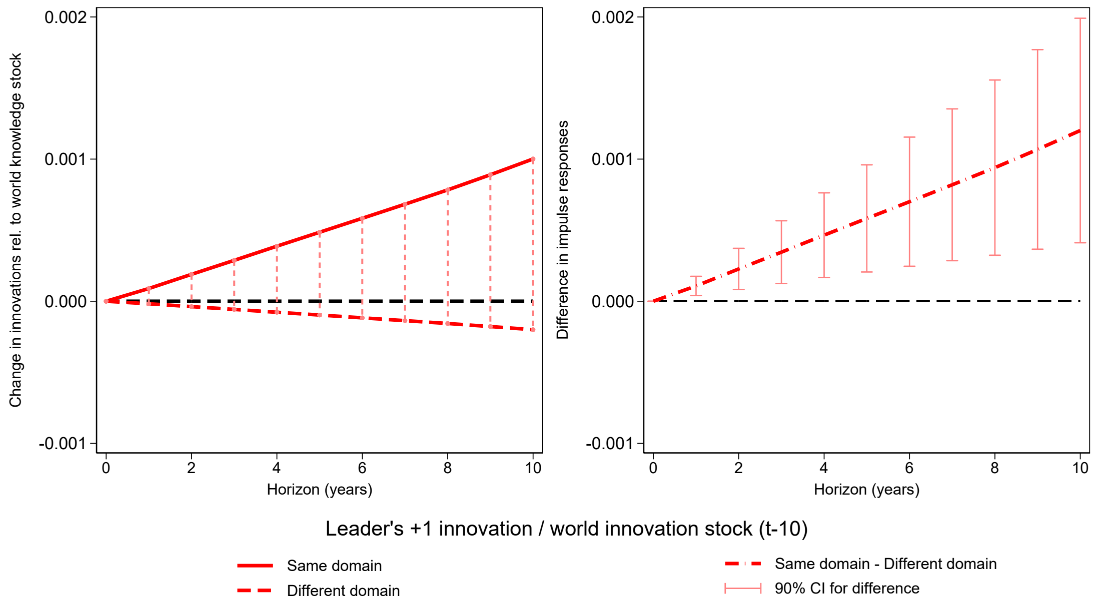
with observations for which innovations$_{n,d,t} = 0$ dropped from the regression, and an R^2 of 0.88. The implied growth decomposition is:

$$\Delta \log(\text{innovations}_{n,d,t}) = \Delta\gamma_{d,t} + \Delta\gamma_{n,t} + \Delta\epsilon_{n,d,t}. \tag{IA2}$$

The variance decomposition (Figure IA11) shows that the local country component accounts for a substantially larger share of the variance in innovation growth than the global domain component. The pattern is very similar when computed for each domain separately.

We also construct counterfactual innovation series by muting either the global or local component. Figure IA12 shows time series for the global component for each domain. Figure IA13 presents three

Figure IA9. Spillovers from Leader Innovations, by Domain



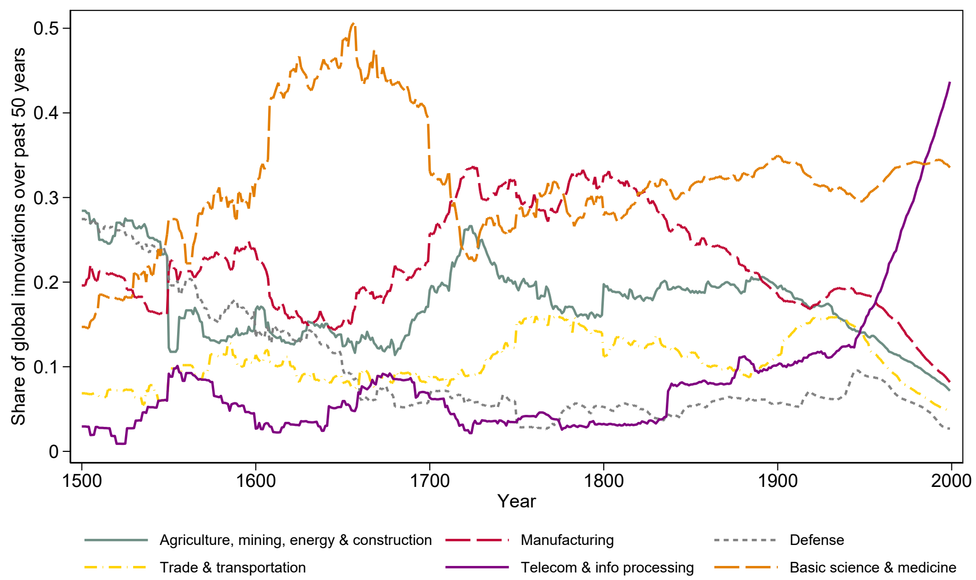
Notes: The figure shows changes in a country's future innovation activity in a domain following increases in the global innovation leader's innovation activity in the same versus a different domain. Estimates come from local projections: $x_{n,d,\,t\to t+h} = \gamma_{n,d,h} + \gamma_{n,t,h} + \mathbf{1}\{d = d'\}\,\beta_h\,x_{l,d',t} + \mathbf{1}\{d \neq d'\}\,\delta_h\,x_{l,d',t} + \mu_h\,\mathbf{1}\{d = d'\} + \boldsymbol{\psi}_h' \boldsymbol{z}_{n,d,t} + \epsilon_{n,d,d',t,h}$, where n indexes the country, d the domain of the country's own innovation activity, d' the domain of the leader's innovation activity, and h the prediction horizon; a separate regression is estimated for each h. The outcome $x_{n,d,\,t\to t+h}$ is the country's total innovations in domain d from year t to $t+h$, relative to the cumulative world innovation stock. $x_{l,d',t}$ is the global innovation leader's total innovations in domain d' over the past 10 years, relative to the cumulative world innovation stock 10 years earlier (the leader is defined as the country with the most total innovations over the past 10 years). The control vector includes $x_{n,d,t}$ and $x_{n,t}$, the country's own innovations in domain d and its own total innovations over the past 10 years, relative to the cumulative world innovation stock 10 years earlier. The specification includes country$\times$domain and country$\times$year fixed effects. The sample excludes the leader country. The solid (dashed) line in the left panel shows $\hat{\beta}_h$ ($\hat{\delta}_h$) for $h = 1, \ldots, 10$, i.e., the response to the leader's innovations in the same (a different) domain. The right panel shows the difference $\hat{\beta}_h - \hat{\delta}_h$, with 90% confidence intervals based on standard errors double clustered by country and year.

panels: counterfactual innovation in major countries when the global domain component is muted, when the local country component is muted, and the actual path of innovation for comparison. Muting the global component leaves country-level trajectories largely unchanged, whereas muting the local component dramatically alters them. Global domain-specific shocks matter quantitatively only for small countries.

IA5.8 Private and Public Innovations

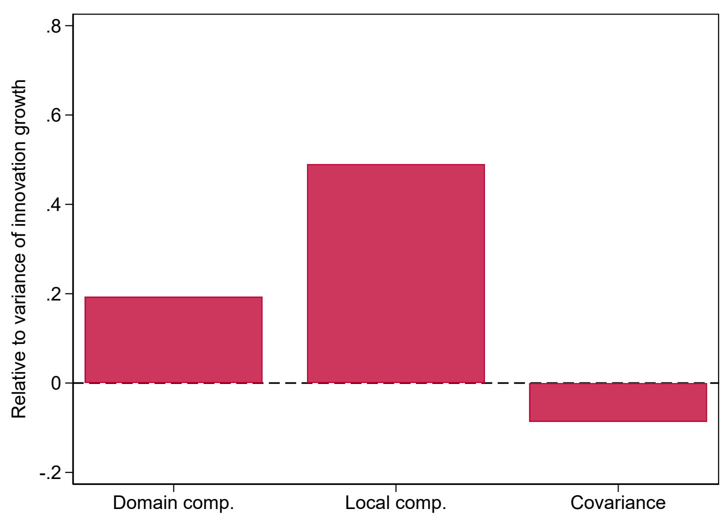
Innovation has both private and public sources. Private innovations are those attributed to companies and individual inventors, which respond more directly to market incentives and local economic

Figure IA10. Decomposition of Global Innovation by Domain



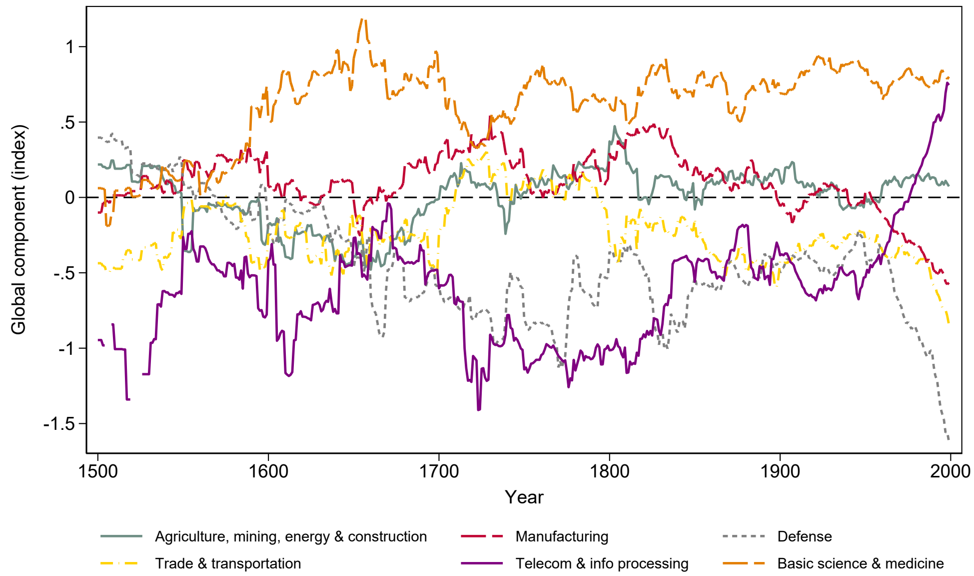
Notes: The figure shows the share of each domain of application in total global innovations over time. Innovations are classified into domains following the structure of two-digit NAICS codes, supplemented with categories for military and defense and basic research.

Figure IA11. Variance Decomposition of Innovation Growth



Notes: The figure presents the variance decomposition of innovation growth $\Delta \log(\text{innovations}_{n,d,t})$ into the global domain component $\Delta\gamma_{d,t}$, the local country component $\Delta\gamma_{n,t}$, their covariance, and the residual. The local country component accounts for a substantially larger share of the variance than the global domain component.

Figure IA12. Global Components by Domain



Notes: The figure shows the estimated global domain component $\gamma_{d,t}$ from the fixed effects decomposition for each domain of application. The global component captures domain-specific trends in innovation that are common across all countries.

conditions; public innovations are those by universities, government entities, or tied to the military, defense, or basic research. We ask two questions. Is the geographic concentration documented in the main text common to both types? And, given the extensive evidence that public research spills over to the private sector, is local innovation driven primarily by public research? The first concerns the spatial distribution of the two types; the second, their dynamic relationship.

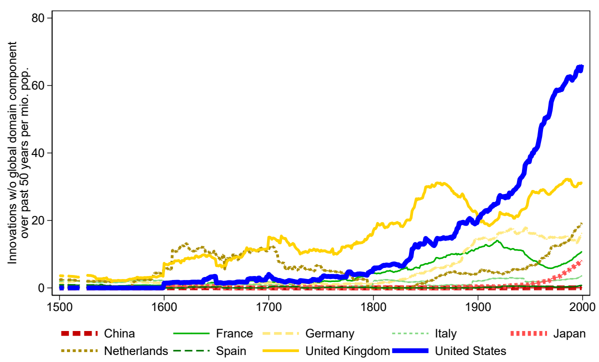
Spatial distribution of private and public innovations Figures IA14 and IA15 plot innovation shares for the two types separately, across continents (0 to 1999) and across countries (1500 to 1999). The patterns closely track those for total innovation: private and public innovations share the same shifting centers of gravity and the same high concentration. The geography of innovation is thus not an artifact of any single sector—it holds whether innovation is driven by the market or by the state.

Dynamic relation between private and public innovations Do private and public innovations move together, and does one tend to lead the other? We estimate lead-lag correlograms over horizons $h \in \{-100, -50, 0, 50, 100\}$ years, regressing a country's public (private) innovations over the 50 years before $t + h$ on its private (public) innovations over the 50 years before t:

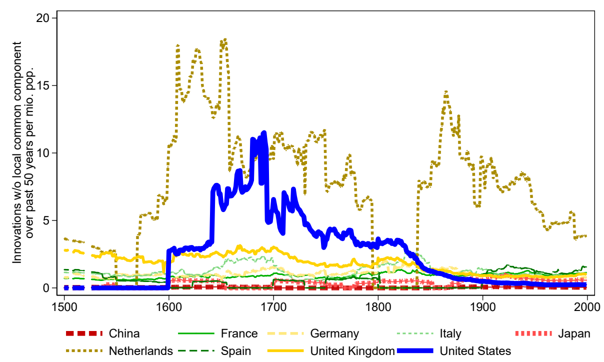
$$y_{n,t+h} = \alpha_n + \alpha_t + \beta_h \, x_{n,t} + \epsilon_{n,t}. \tag{IA3}$$

Figure IA13. Global vs. Local Components for Aggregate Innovation

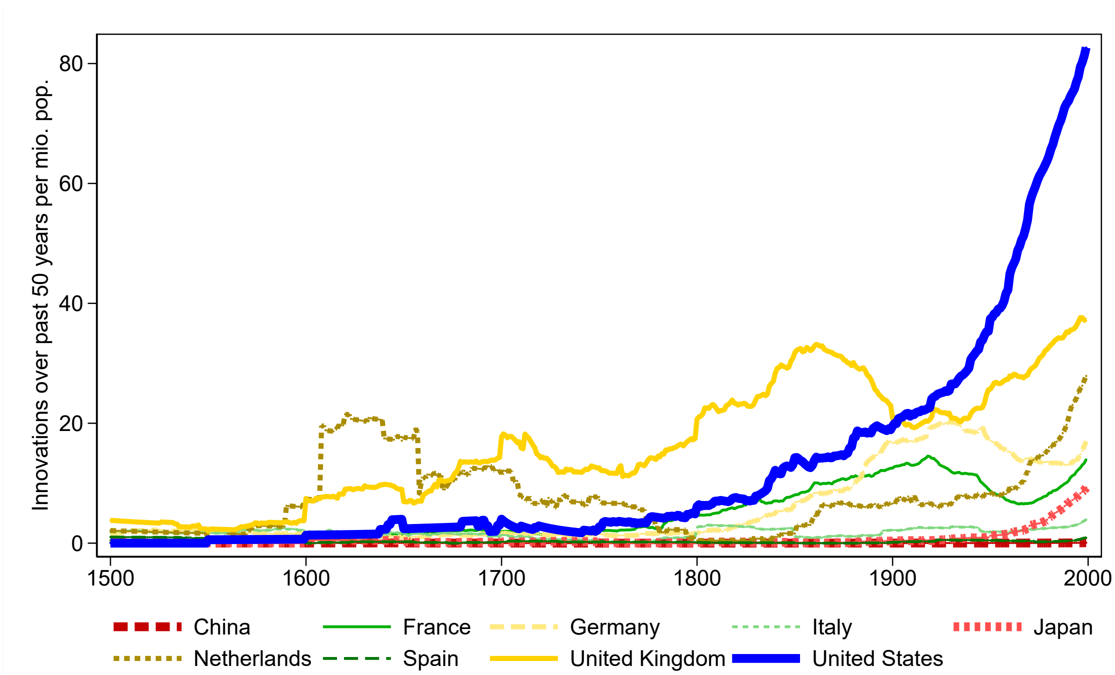
(a) Muting Global Component



(b) Muting Local Component

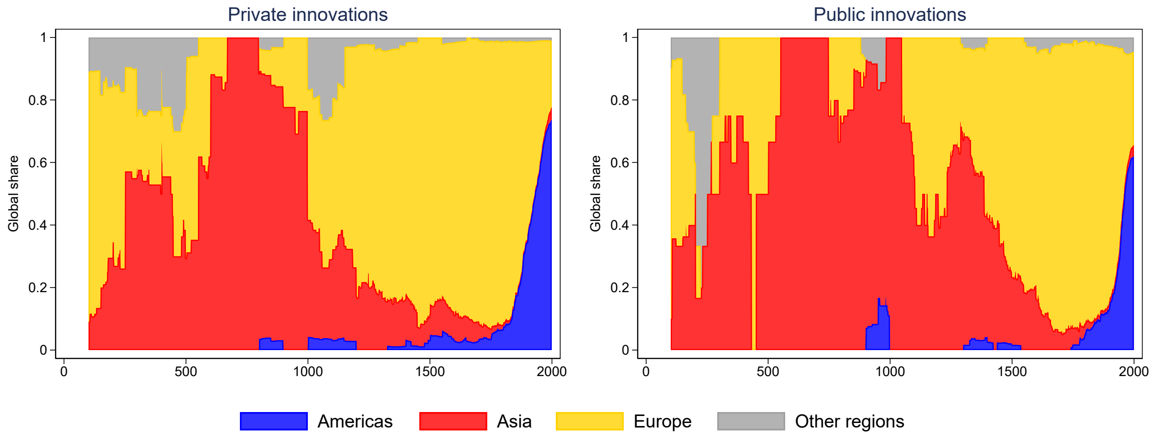


(c) Actual Path of Innovation



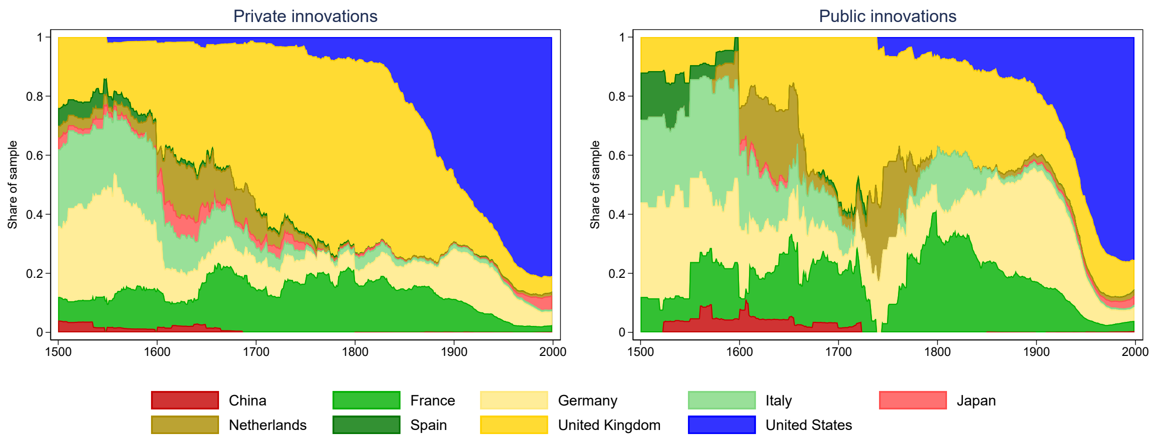
Notes: The figure shows counterfactual innovation activity in major countries since 1600. Each counterfactual is constructed independently from the estimated two-way fixed effects decomposition $\log(\text{innovations}_{n,d,t}) = \gamma_{d,t} + \gamma_{n,t} + \epsilon_{n,d,t}$. "Muting global" (Panel a) sets $\gamma_{d,t} = 0$ for all domains and years, computing counterfactual innovations as $\exp(\log \text{innovations}_{n,d,t} - \hat{\gamma}_{d,t})$; "muting local" (Panel b) sets $\gamma_{n,t} = 0$, computing $\exp(\log \text{innovations}_{n,d,t} - \hat{\gamma}_{n,t})$. The two counterfactuals are not cumulative—each removes one component from the baseline. Panel (c) shows the actual path for comparison. The contrast illustrates that innovation trajectories are driven primarily by local factors. Gaps in the counterfactual series in early periods arise where countries have zero innovations in some domains, so the fixed effects cannot be estimated and the counterfactual cannot be computed.

Figure IA14. Innovation Concentration in Private vs. Public Innovations at the Continent Level



Notes: The figure shows each continent's stacked share of global innovations, separately for private innovations (left panel) and public innovations (right panel), between 0 and 1999. Private innovations are those attributed to companies and individual inventors. Public innovations are those by universities, government entities, or classified as military or basic research. "Other regions" (shown in gray) aggregates all continents other than the Americas, Asia, and Europe. Innovations are measured as the cumulative count over the past 100 years.

Figure IA15. Innovation Concentration in Private vs. Public Innovations at the Country Level



Notes: The figure shows the stacked share of innovations accounted for by nine major countries (each expressed as a share of the nine-country total), separately for private innovations (left panel) and public innovations (right panel), between 1500 and 1999. Private innovations are those attributed to companies and individual inventors. Public innovations are those by universities, government entities, or classified as military or basic research. Innovations are measured as the cumulative count over the past 50 years.

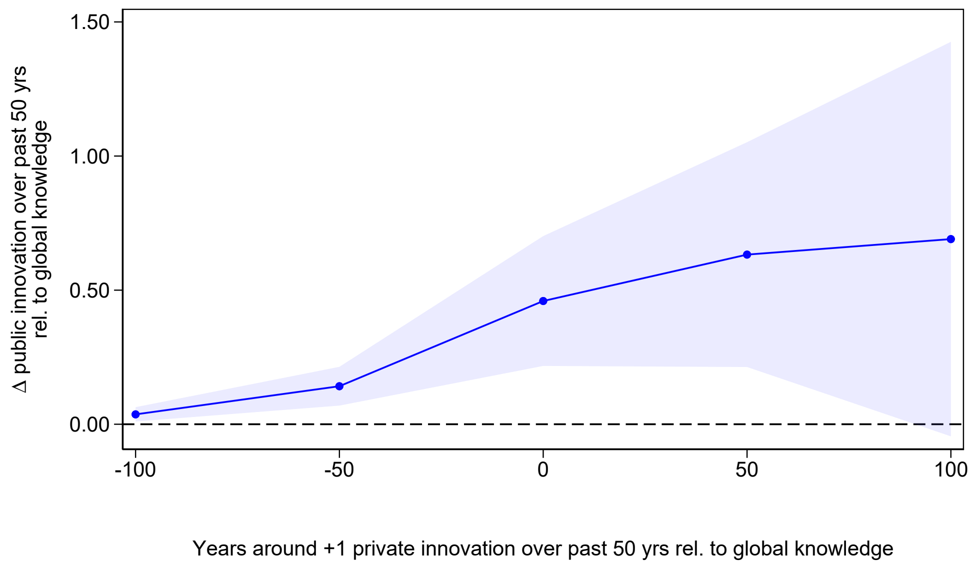
Country and year fixed effects absorb each country's average innovation intensity and common global trends, and innovation is normalized by the global knowledge stock, as before. Figure IA16 reports the estimates. Both types are positively related over long windows: a country rich in one tends to be rich in the other. The timing, however, is asymmetric. Private innovation is more closely associated with *future* public innovation than with past public innovation, whereas public innovation is associated similarly with past and future private innovation. If anything, then, private innovation tends to lead public innovation, rather than the reverse—evidence against public research being the main force behind local innovation. One interpretation is that private innovation helps shape or inspire subsequent public research, though the long-horizon estimates are imprecise.

IA5.9 Concentration of Other Creative Endeavors

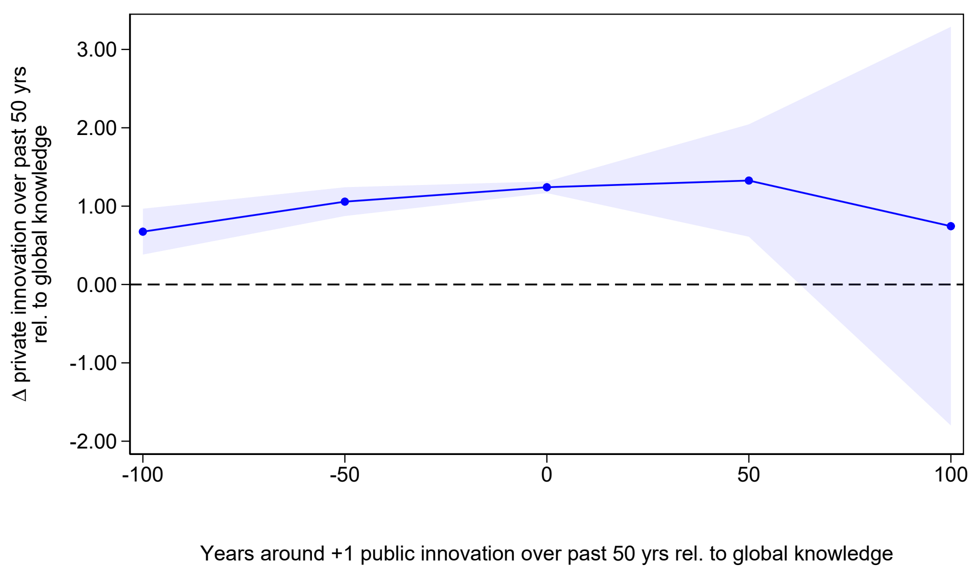
This appendix compares the geographic concentration of technological innovation with that of two other types of creative activities discussed in Section 6: mathematical discoveries and literature. Figure IA17 presents two measures of concentration: Panel (a) reports the number of countries needed to account for half of total activity, and Panel (b) the cross-country coefficient of variation of activity per capita. All three activities are far more concentrated than economic output or population. Mathematics was the most concentrated through the 19th century, but its concentration declined over time while that of innovation rose, so that technological innovation became the most concentrated activity in recent decades.

Figure IA16. Dynamic Relation between Private and Public Innovations

(a) Public Innovations around additional Private Innovations



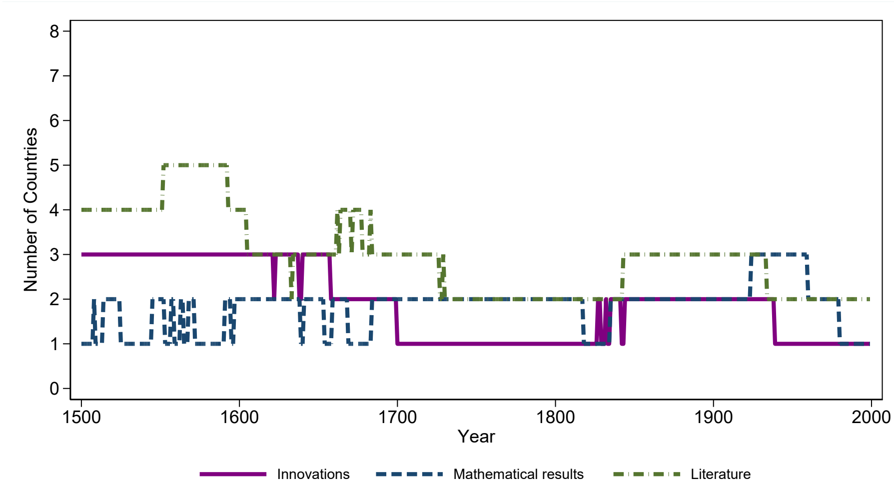
(b) Private Innovations around additional Public Innovations



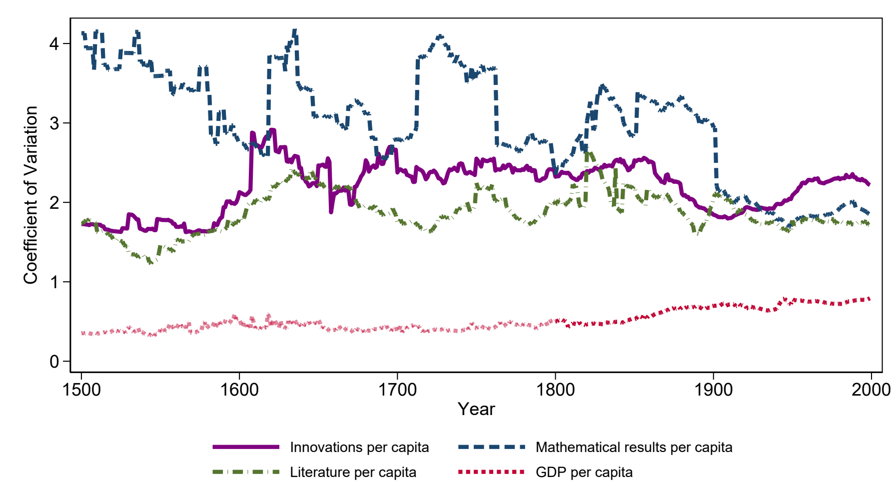
Notes: The figures show the lead-lag correlogram between private- and public-sector innovation. For each horizon $h \in \{-100, -50, 0, 50, 100\}$ years, the top panel shows $\hat{\beta}_h$ from $y_{n,t+h} = \alpha_n + \alpha_t + \beta_h x_{n,t} + \epsilon_{n,t}$, where y (x) is a country's public-sector (private-sector) innovation count over the trailing 50 years, scaled by the global stock of innovations. The bottom panel shows $\hat{\gamma}_h$ from $x_{n,t+h} = \alpha_n + \alpha_t + \gamma_h y_{n,t} + \epsilon_{n,t}$. The shaded band is a 95% confidence interval from standard errors clustered by country. Sample includes all countries and years 1500 to 1999.

Figure IA17. Concentration of Creative Activities across Countries

(a) Number of countries for half of world creative activity



(b) Dispersion across countries: coefficient of variation



Notes: Panel (a) plots the number of countries needed to account for 50% of global activity for technological innovations, mathematical discoveries, and literature, each year from 1500 to 1999; within each year, countries are ranked in descending order of activity and the smallest set whose cumulative share first reaches 50% is counted. Panel (b) plots the cross-country coefficient of variation (standard deviation divided by the mean) of per-capita output for the same three activities and for real GDP. In Panel (b), the sample is the 63 countries with at least 10 decades of non-missing GDP per capita from 1500 to 1999 (the quantitative estimation sample of Section 5), and the three creative measures are computed over a common set within this sample (a country enters a given year only if its population is positive and all three are observed that year); real GDP per capita is from the Maddison Project Database (Bolt and van Zanden, 2024), computed over its own set of countries with available data within this sample and shown in a lighter shade before 1800 where the historical data are sparser. Each measure is the cumulative count of activity over the past 50 years, and each series begins only once that type of activity first appears in the data. A higher value in Panel (b) indicates greater dispersion across countries.

IA6 Model Estimation

This appendix collects the supporting details for the quantitative exploration in Section 5: the estimable equations and identification, the parameter estimates, the per-country frictions, the structural variance decomposition, and the implied research-productivity panel. Both blocks are estimated by maximum likelihood, simultaneously—the output log-likelihood (on the panel of log GDP per capita) and the innovation log-likelihood (on the panel of innovation counts) are summed and maximized over the full parameter vector at once—with parametric-bootstrap confidence intervals. The estimation uses all 63 countries that meet the 10-decade coverage rule of Section 5.1 over 1500 to 1999; the tables and figures display the 26 example countries.

IA6.1 Estimation Results

Table IA2 – Structural Parameters: Fixed Parameters

	Parameter	Estimate	95% CI
Fixed parameters			
Fréchet dispersion	θ	2.00	—
Insight-spillover elasticity	β	0.50	—
Population-growth wedge	γ	0.01	—
Dynamic-agglomeration exponent	ζ	0	—

Notes: This table reports the fixed parameters of the quantitative model. The fixed parameters are the Fréchet dispersion θ (so output per capita has elasticity $1/\theta = 0.5$ in the technology pool), the insight-spillover elasticity β, the population-growth wedge γ, and the dynamic-agglomeration exponent ζ (set to zero).

Table IA3 – Structural Parameters: Estimated and Calibrated Parameters

	Parameter	Estimate	95% CI
Output block — ML on log GDP per capita			
Across-continent technology discount	ξ^A	2.05	[1.00, 5.80]
Speed dispersion shape, adoption	ζ^A	0.53	[0.47, 0.74]
AR(2) productivity, lag 1	ρ_1	1.31	[1.16, 1.41]
AR(2) productivity, lag 2	ρ_2	-0.37	[-0.48, -0.22]
Persistent shock s.d.	σ_η^y	0.122	[0.102, 0.135]
Transitory shock s.d.	σ_ε^y	0.041	[0.024, 0.061]
Innovation block — ML on innovation counts			
Labor-supply elasticity†	σ	0.35	[0.24, 2.76]
Speed dispersion shape, knowledge	ζ^R	0.20	[0.15, 0.53]
Across-continent insight discount†	ξ^R	1.31	[1.00, 226.21]
AR(1) research productivity	ρ_R	0.99	[0.96, 0.99]
Persistent shock s.d.	σ_η^R	0.315	[0.266, 0.477]
Transitory shock s.d.	σ_ε^R	0.201	[0.002, 0.245]
Idea flows and joint speed distribution			
Calibrated stock-growth scale	$\tilde{q}$	0.0054	calibrated
Threshold growth‡	ν	1.06	[1.05, 1.09]
Gumbel copula parameter	θ_{cop}	2.82	[2.19, 3.70]
Kendall's rank correlation	τ_K	0.64	[0.54, 0.73]

Notes: This table reports the estimated and calibrated structural parameters of the quantitative model, estimated on the 63-country 10-decade sample over 1500 to 1999, with 95% parametric-bootstrap confidence intervals. The output block is estimated by maximum likelihood on log GDP per capita: ξ^A is the across-continent technology discount, ζ^A is the shape index governing the dispersion of the country adoption speeds φ_n^A (reported in Table IA4), (ρ_1, ρ_2) is the AR(2) output-productivity process, and $(\sigma_\eta^y, \sigma_\varepsilon^y)$ are its persistent and transitory shock standard deviations. The innovation block is estimated by maximum likelihood on innovation counts jointly with the output block: the labor-supply elasticity σ, the shape index ζ^R of the knowledge speeds φ_n^R, the across-continent insight discount ξ^R, and the AR(1) research-productivity process ρ_R with its persistent and transitory shock standard deviations. The idea-flow scale reports $\tilde{q}$, which is calibrated so the world knowledge stock grows in line with long-run frontier output growth; ν, the growth factor of the innovation threshold ($r_t = r_0\nu^t$), profiled by likelihood; and the Gumbel copula parameter θ_{cop} and Kendall's τ_K, which summarize the cross-country dependence between the separately estimated φ_n^A and φ_n^R. †The labor-supply elasticity σ and the across-continent insight discount ξ^R are weakly identified; their bootstrap intervals are wide and asymmetric, which is why β is fixed. ‡Because ν enters nonlinearly and is estimated by a likelihood profile rather than inside the joint optimizer, the interval reported for it is a 95% profile-likelihood confidence interval (likelihood-ratio inversion, χ_1^2), not the parametric bootstrap used for the other rows.

Table IA4 – Per-Country Geographic Frictions

Country	φ_n^A	φ_n^R	Country	φ_n^A	φ_n^R
Europe			*Americas (cont.)*		
United Kingdom	0.857	0.918	Chile	0.391	0.023
Germany	0.602	0.770	Uruguay	0.450	0.063
France	0.626	0.482	Cuba	0.024	0.002
Switzerland	0.971	0.842	Peru	0.194	0.000
Russia	0.285	0.142	Ecuador	0.072	0.000
Netherlands	0.830	0.437	Jamaica	0.132	0.036
Sweden	0.699	0.642	Panama	0.253	0.036
Italy	0.776	0.704	Bolivia	0.208	0.036
Austria	0.674	0.230	*Asia*		
Belgium	0.942	0.288	Japan	0.354	0.258
Denmark	0.885	0.320	China	0.081	0.013
Finland	0.491	0.532	India	0.030	0.033
Norway	0.724	0.585	Taiwan	0.223	0.000
Poland	0.268	0.084	South Korea	0.156	0.002
Hungary	0.301	0.125	Philippines	0.042	0.001
Spain	0.534	0.003	Turkey	0.121	0.000
Czechoslovakia	0.372	0.073	Singapore	0.318	0.110
Ireland	0.557	0.161	Indonesia	0.090	0.000
Portugal	0.470	0.096	Thailand	0.049	0.000
Greece	0.430	0.011	Malaysia	0.168	0.000
Bulgaria	0.181	0.004	Sri Lanka	0.064	0.000
Former Yugoslavia	0.056	0.007	Myanmar	0.005	0.000
Romania	0.100	0.005	*Africa*		
Iceland	0.579	0.182	South Africa	0.410	0.046
Albania	0.015	0.000	Kenya	0.008	0.001
Americas			Ghana	0.019	0.000
United States	0.913	0.054	Nigeria	0.011	0.000
Canada	0.803	0.395	Liberia	0.035	0.001
Brazil	0.110	0.016	Uganda	0.001	0.036
Mexico	0.237	0.028	Zambia	0.000	0.036
Argentina	0.513	0.019	Malawi	0.003	0.036
Colombia	0.143	0.009	*Oceania*		
Venezuela	0.336	0.039	Australia	0.750	0.356
			New Zealand	0.650	0.205

Notes: This table reports the per-country adoption speed φ_n^A and the separately estimated knowledge speed φ_n^R for all 63 countries in the estimation sample, grouped by continent and ordered by innovation count within each continent; the list is split into two side-by-side blocks to fit on one page. Both speeds are bounded above at one (full-speed access to the world pool); the frontier economies sit at or near the ceiling. Within-continent foreign access is governed by the country speed alone; across-continent access is slowed by an additional uniform discount, $\xi^A = 2.05$ on φ^A and $\xi^R = 1.32$ on φ^R. The knowledge speeds are estimated freely; for countries with very few recorded innovations φ_n^R is a near-zero extrapolation pinned by rank and only loosely identified (e.g. Peru, Ecuador, Myanmar), and is better read as "effectively closed to foreign insight" than as a precise level. The speeds are estimated by maximum likelihood on the full 63-country sample over 1500 to 1999.

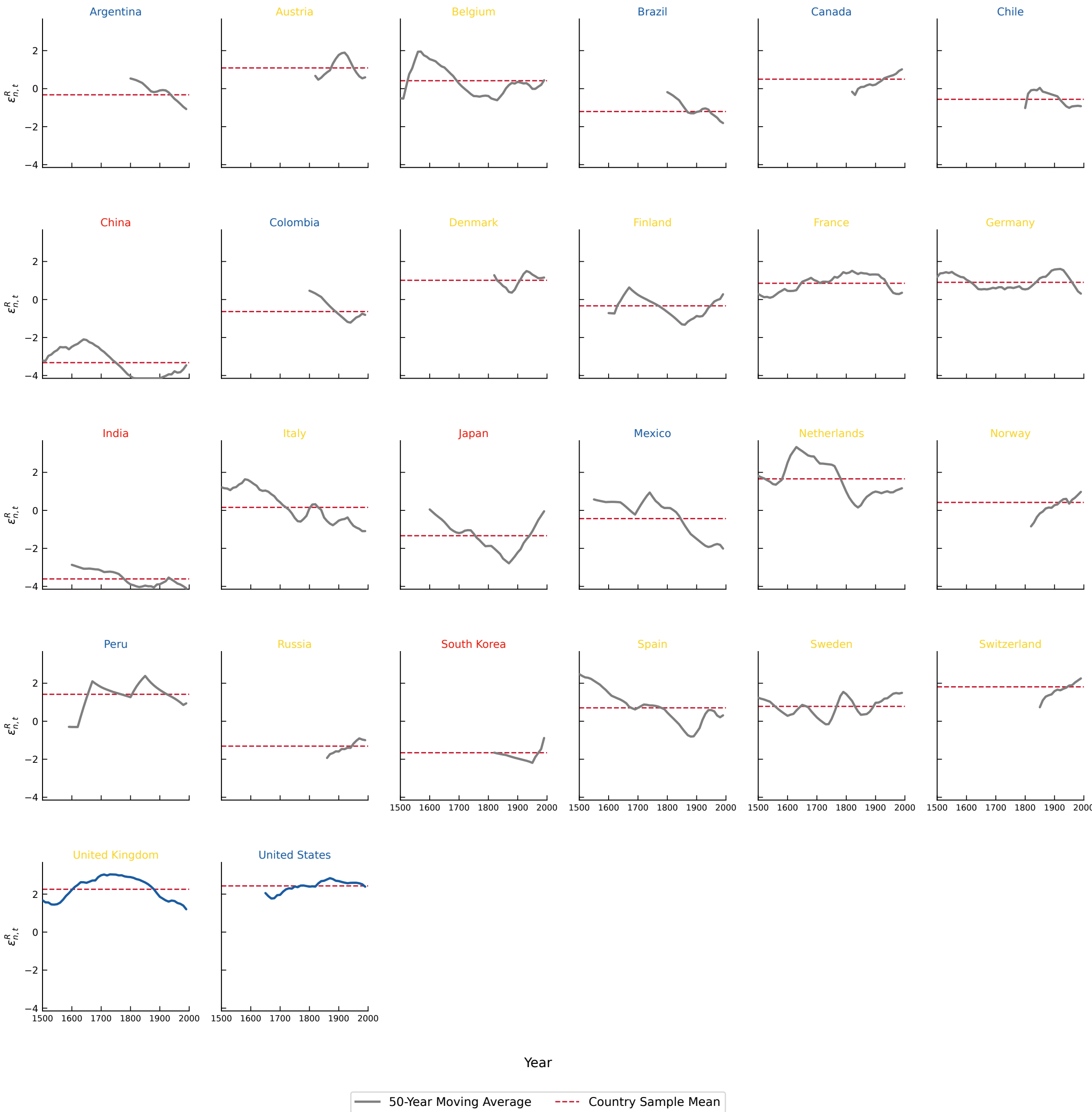
Table IA5 – Structural Variance Decomposition: Innovation and Output per Capita

	Innovations per capita			Output per capita		
Period	Frictions	Persistent	Transitory	Frictions	Persistent	Transitory
1500	3%	84%	13%	53%	46%	0%
1550	8%	73%	19%	25%	74%	1%
1600	14%	66%	20%	19%	80%	1%
1650	10%	83%	7%	31%	69%	1%
1700	4%	90%	6%	21%	78%	1%
1750	2%	89%	9%	38%	62%	0%
1800	2%	87%	11%	43%	56%	1%
1850	17%	83%	0%	54%	46%	1%
1900	39%	64%	-3%	87%	13%	0%
1950	52%	47%	1%	77%	22%	1%

Notes: This table decomposes the cross-country variance of log innovations per capita and log output per capita into three structural components, by 50-year period: *frictions* (the deterministic diffusion-pool term, i.e. differential access through φ^A and φ^R), *persistent* (the filtered AR research- and output-productivity states), and *transitory* (residual and, for innovation, Poisson sampling). Each share is a covariance decomposition, $\mathrm{Cov}(\text{component, total})/\mathrm{Var}(\text{total})$, and the three sum to 100% within each block; small negative entries arise when a component covaries negatively with the total. For output, the frictions channel grows to explain nearly nine-tenths of the modern cross-country output gap by 1900. For innovation it is small until the 19th century—persistent, country-specific research productivity dominates—and then rises to about 52% by the mid-20th century, so access governs the geography of output well before it governs the geography of ideas. Figure 13 plots the same shares over time, combining the persistent and transitory components into a single *shocks* channel. Estimates use the full 63-country sample.

The figures that follow report the implied research-productivity panel and the structural variance decomposition. Both use the full 63-country estimation over 1500 to 1999; the research-productivity figure displays the 26 example countries.

Figure IA18. Implied Research Productivity: Persistent Component ϵ^R



Notes: Each panel plots the implied research productivity of one of the 26 example countries, ordered alphabetically, on a common (uniform) vertical scale (panel titles colored by continent: Europe in yellow, the Americas in blue, Asia in red). Research productivity $u^R_{n,t}$ is the part of log innovations not accounted for by population scale or by access to the diffusion pools, $u^R_{n,t} = \log I_{n,t} - [\log L_{n,t} + \beta(1+\sigma)\log \mathcal{R}_{n,t} - \sigma \log \mathcal{A}_{n,t}]$ (defined for country-decades with positive counts), and decomposes into a common decade effect, a persistent AR(1) component, and a transitory component, $u^R_{n,t} = \alpha^R_t + \epsilon^R_{n,t} + v^R_{n,t}$, as in Section 5.1. Each panel plots the persistent component $\epsilon^R_{n,t}$, filtered from the Poisson likelihood—net of the common decade effect α^R_t and excluding the transitory shock $v^R_{n,t}$. The solid line is a 50-year trailing moving average of $\epsilon^R_{n,t}$ (the United States and the United Kingdom in blue, other countries in gray); the dashed red line is each country's sample mean of that series. Pools and diffusion speeds are taken from the full 63-country estimation.

IA6.2 Likelihood and Identification

Both blocks are estimated by maximum likelihood, simultaneously, taking as given the population path $L_{n,t}$ (Maddison Project Database) and the measured innovation counts $I_{n,t}$ (Section 2). We maximize the sum of the two log-likelihoods over the full parameter vector at once; the two blocks are coupled through the output access pool $\mathcal{A}$, which enters the innovation log-mean (through the $-\sigma \log \mathcal{A}$ congestion term and the insight pool), so the innovation data also bear on the adoption-speed parameters. The diffusion pools inherit the within- versus across-continent structure of Section 4.

Diffusion pools Ideas accumulate as latent flows proportional to the measured counts, $x_{n,t} = \tilde{q}_t \, \Lambda_{t-1} \, I_{n,t}$ with $\tilde{q}_t = \tilde{q} \, \nu^{-t}$ (equivalently, $\tilde{q}_t = 1/r_t$ for the innovation threshold $r_t = r_0 \nu^t$ of Section 5.1) and world stock $\Lambda_t = \Lambda_{t-1}\Big(1 + \tilde{q}_t \sum_n I_{n,t}\Big)$. The technology pool is the Eaton–Kortum access aggregate

$$\mathcal{A}_{n,t} = \Lambda_{t-1} - \sum_{\tilde{n}} g^A_{n\tilde{n},t}, \qquad g^A_{n\tilde{n},t} = \Big(1 - \varphi^A_{n\tilde{n}}\Big)\Big(x_{\tilde{n},t-1} + g^A_{n\tilde{n},t-1}\Big), \qquad \varphi^A_{n\tilde{n}} = \begin{cases} 1 & \tilde{n} = n, \\ \varphi^A_n & c(\tilde{n}) = c(n), \\ \varphi^A_n / \xi^A & c(\tilde{n}) \neq c(n), \end{cases} \tag{IA4}$$

and the insight pool $\mathcal{R}_{n,t}$ is identical with φ^A replaced by φ^R (country levels φ^R_n, across-continent discount ξ^R). A pre-1500 warm-up discounts the inherited stock so the pools are already differentiated in 1500. Because the world stock Λ_t grows at a friction-independent rate, the scale $\tilde{q}$ can be calibrated to a stock-growth target—matching long-run frontier output growth—before the friction likelihood is evaluated, and ν is estimated over a grid.

Output block Output per capita satisfies

$$\log y_{n,t} = \tfrac{1}{\theta} \log \mathcal{A}_{n,t} + \alpha^y_t + \epsilon^y_{n,t} + \upsilon^y_{n,t}, \tag{IA5}$$

with α^y_t a common decade effect, $\epsilon^y_{n,t}$ a country AR(2) productivity state, and $\upsilon^y_{n,t}$ a transitory shock, as in Section 5.1. With $\theta = 2$ fixed, the block estimates $(\zeta^A, \xi^A, \rho_1, \rho_2, \sigma^y_\eta, \sigma^y_\varepsilon)$ by maximizing the Gaussian likelihood of the log-GDP-per-capita panel, filtering ϵ^y by a Kalman recursion and removing α^y_t by within-decade demeaning. Each country's φ^A_n is placed at the rank-quantile of a power-function(ζ^A) distribution on $(0, 1]$ (equivalently, $1/\varphi^A_n$ is Pareto with tail ζ^A) implied by its rank in a preliminary free fit, so a single shape index governs the whole cross-section. The cross-section of GDP levels identifies ζ^A (faster access raises output); the within- versus across-continent split of GDP concentration identifies ξ^A.

Innovation block Innovations per capita satisfy

$$\log\Big(x_{n,t}/L_{n,t}\Big) = \beta(1+\sigma)\log\mathcal{R}_{n,t} - \sigma\log\mathcal{A}_{n,t} + \alpha_t^R + \epsilon_{n,t}^R + \upsilon_{n,t}^R, \tag{IA6}$$

with ϵ^R an AR(1) research-productivity state and υ^R a transitory shock; the outcome is the measured count $I_{n,t}$, modeled as Poisson with this structural log-mean. With $\beta = 0.5$ fixed, the block estimates $(\zeta^R, \xi^R, \sigma, \rho_R, \sigma_\eta^R, \sigma_\varepsilon^R)$ by maximum likelihood, filtering ϵ^R by a Poisson–Kalman recursion, with φ_n^R placed at power-function(ζ^R) rank-quantiles from a preliminary free fit. The shape index ζ^R is identified by how much more concentrated innovation is than output; the labor-supply elasticity σ and the across-continent insight discount ξ^R are only weakly identified (wide, asymmetric bootstrap intervals), which is why we fix β.

The joint distribution of diffusion speeds The two blocks deliver separately estimated speed schedules $\{\varphi_n^A\}$ and $\{\varphi_n^R\}$. We summarize their cross-country dependence by a one-parameter Gumbel copula fit to the recovered rank pairs, reporting Kendall's τ_K and the implied copula parameter. The dependence is strong but imperfect ($\tau_K = 0.65$): a country with slow adoption typically has slow knowledge diffusion too, and $\varphi_n^R < \varphi_n^A$ for most countries.

Implied research productivity and diagnostics The innovation block decomposes research productivity $u_{n,t}^R = \log I_{n,t} - [\log L_{n,t} + \beta(1+\sigma)\log\mathcal{R}_{n,t} - \sigma\log\mathcal{A}_{n,t}]$ into a common decade effect α_t^R, a persistent AR(1) component $\epsilon_{n,t}^R$ with persistence ρ_R, and a transitory component $\upsilon_{n,t}^R$; Figure IA18 plots the persistent component $\epsilon_{n,t}^R$ filtered from the Poisson likelihood.

IA7 Proofs

This appendix collects the proofs of the formal results stated in Section 4.

IA7.1 Proof of the Royalty-Payments Formula (11)

Under the model's same-period adoption assumption, contemporaneous ideas from location ℓ are accessible only to adopters in ℓ, so royalty payments accrue only locally.

For $q_1 \geq q_2$, define

$$F_{\ell t}^{\text{new}}(q_1, q_2) = e^{-x_{\ell t} q_2^{-\theta}}\left[1 + x_{\ell t}\left(q_2^{-\theta} - q_1^{-\theta}\right)\right]$$

as the joint CDF of the best and second-best contemporaneous ideas discovered in ℓ at time t. Let $F_{\ell t}^{\text{old}}(q) \equiv e^{-\left(\sum_{\tilde{\ell}}\sum_{k=1}^{\infty}[1-(1-\varphi_{\ell\tilde{\ell}}^A)^k]x_{\tilde{\ell},t-k}\right)q^{-\theta}} = e^{-(\mathcal{A}_{\ell t} - x_{\ell t})q^{-\theta}}$ denote the CDF of the best idea from the pre-existing stock available to adopters in ℓ.

Proof. The probability density that, in a randomly chosen domain accessible to adopters in ℓ, (i) the best idea is a contemporaneous discovery in ℓ with productivity q_1, and (ii) the second-best technique across all sources has productivity below q_2, is

$$\frac{\partial F_{\ell t}^{\text{new}}(q_1,q_2)}{\partial q_1}\,F_{\ell t}^{\text{old}}(q_2)=x_{\ell t}\theta q_1^{-\theta-1}\,e^{-x_{\ell t}q_2^{-\theta}}\,e^{-(\mathcal{A}_{\ell t}-x_{\ell t})q_2^{-\theta}}=x_{\ell t}\theta q_1^{-\theta-1}\,e^{-\mathcal{A}_{\ell t}\,q_2^{-\theta}},$$

where the consolidated exponent is the location-level technology pool $\mathcal{A}_{\ell t}$ defined in (7). Differentiating this expression with respect to q_2 gives the joint density of the best contemporaneous idea from ℓ at q_1 and the second-best available idea at q_2:

$$x_{\ell t}\theta q_1^{-\theta-1}\,\mathcal{A}_{\ell t}\theta q_2^{-\theta-1}e^{-\mathcal{A}_{\ell t}q_2^{-\theta}}.$$

The profit (before royalty payments) an entrepreneur in ℓ earns when producing with productivity q is

$$Y_{\ell t}P_{\ell t}^{\varepsilon}\left(\frac{\varepsilon}{\varepsilon-1}w_{\ell t}^{Y}\right)^{1-\varepsilon}\frac{1}{\varepsilon}\,q^{\varepsilon-1}.$$

Following Bernard et al. (2003), the royalty payment to the inventor is $\eta_{\ell t}$ times the profit advantage of the best idea over the second-best available idea. Summing across all $N_{\ell t}$ varieties produced in ℓ, total payments to contemporaneous inventors in ℓ are

$$\begin{aligned}T_{\ell t}&=N_{\ell t}\,\eta_{\ell t}\,Y_{\ell t}P_{\ell t}^{\varepsilon}\left(\frac{\varepsilon}{\varepsilon-1}w_{\ell t}^{Y}\right)^{1-\varepsilon}\frac{1}{\varepsilon}\,x_{\ell t}\\&\quad\times\int_0^\infty\int_0^{q_1}\left[q_1^{\varepsilon-1}-q_2^{\varepsilon-1}\right]\theta q_1^{-\theta-1}\,\mathcal{A}_{\ell t}\,\theta q_2^{-\theta-1}\,e^{-\mathcal{A}_{\ell t}q_2^{-\theta}}\,dq_2\,dq_1.\end{aligned}$$

The double integral splits into two terms. For the $q_1^{\varepsilon-1}$ piece, the inner integral over $q_2\in[0,q_1]$ produces $e^{-\mathcal{A}_{\ell t}q_1^{-\theta}}$, giving $\mathcal{A}_{\ell t}^{(\varepsilon-1)/\theta-1}\,\Gamma(1-(\varepsilon-1)/\theta)$. For the $q_2^{\varepsilon-1}$ piece, swapping the order of integration and integrating $q_1\in[q_2,\infty)$ first produces $q_2^{-\theta}$; the remaining q_2 integral evaluates to $\mathcal{A}_{\ell t}^{(\varepsilon-1)/\theta-1}\,\Gamma(2-(\varepsilon-1)/\theta)=\mathcal{A}_{\ell t}^{(\varepsilon-1)/\theta-1}(1-(\varepsilon-1)/\theta)\,\Gamma(1-(\varepsilon-1)/\theta)$. Subtracting yields

$$\int_0^\infty\int_0^{q_1}\left[q_1^{\varepsilon-1}-q_2^{\varepsilon-1}\right]\theta q_1^{-\theta-1}\,\mathcal{A}_{\ell t}\,\theta q_2^{-\theta-1}\,e^{-\mathcal{A}_{\ell t}q_2^{-\theta}}\,dq_2\,dq_1=\mathcal{A}_{\ell t}^{\frac{\varepsilon-1}{\theta}-1}\,\frac{\varepsilon-1}{\theta}\,\Gamma\!\left(1-\frac{\varepsilon-1}{\theta}\right).$$

Substituting and using $P_{\ell t}^{\varepsilon-1}=N_{\ell t}^{-1}\,\left(\frac{\varepsilon}{\varepsilon-1}w_{\ell t}^{Y}\right)^{\varepsilon-1}/[\Gamma(1-(\varepsilon-1)/\theta)\,\mathcal{A}_{\ell t}^{(\varepsilon-1)/\theta}]$ collapses the prefactors to

$$T_{\ell t}=\eta_{\ell t}\,Y_{\ell t}P_{\ell t}\,\frac{\varepsilon-1}{\theta}\,\frac{1}{\varepsilon}\,\frac{x_{\ell t}}{\mathcal{A}_{\ell t}}.$$

Finally, using $Y_{\ell t} P_{\ell t} = \frac{\varepsilon}{\varepsilon - 1} w^Y_{\ell t} L^Y_{\ell t}$ (production-side market clearing in ℓ),

$$T_{\ell t} = \eta_{\ell t} \frac{1}{\theta} w^Y_{\ell t} L^Y_{\ell t} \frac{\lambda_{\ell t} - \lambda_{\ell, t-1}}{\mathcal{A}_{\ell t}},$$

which is equation (11). □

IA7.2 Measured Innovations

Suppose that there are K domains. In each location, there is a stock of ideas given by $\lambda_{\ell t}$. The probability that an idea is in domain k is $\frac{1}{K}$.

The frontier idea in domain k follows a distribution with distribution function $\Pr(q^*_{k,t-1} \leq q) = e^{-\Lambda_{k,t-1} q^{-\theta}}$, where $\Lambda_{k,t} = \frac{1}{K} \sum_\ell \lambda_{\ell t}$. The productivity of the best newly discovered idea in domain k in country n at time t is $\Pr(q_{n,k,t} \leq q) = e^{-\frac{1}{K} x_{nt} q^{-\theta}}$. The probability that the best newly discovered idea in domain k is recorded as an innovation is

$$\begin{aligned}
\Pr\left(q^{new}_{n,k,t} > b_t q^*_{k,t-1}\right) &= \int \Pr\left(q^{new}_{n,k,t} > b_t q^*_{k,t-1} | q^*_{k,t-1} = q\right) d \Pr\left(q^*_{k,t-1} = q\right) \\
&= \int \left[1 - e^{-\frac{1}{K} x_{nt} (b_t q)^{-\theta}}\right] \Lambda_{k,t-1} \theta q^{-\theta - 1} e^{-\Lambda_{k,t-1} q^{-\theta}} dq \\
&= 1 - \frac{\Lambda_{k,t-1}}{\frac{1}{K} x_{nt} b_t^{-\theta} + \Lambda_{k,t-1}} \\
&= \frac{x_{nt} b_t^{-\theta}}{x_{nt} b_t^{-\theta} + K \Lambda_{k,t-1}} \\
&= \frac{x_{nt} b_t^{-\theta}}{x_{nt} b_t^{-\theta} + \Lambda_{t-1}}.
\end{aligned}$$

where $\Lambda_t \equiv \sum_k \Lambda_{kt} = \sum_\ell \lambda_{\ell t}$. Since domains are independent, the number of domains in which country n records an innovation is binomial, with K draws and probability of success $\frac{x_{nt} b_t^{-\theta}}{x_{nt} b_t^{-\theta} + \Lambda_{t-1}}$.

Define $r_t = K b_t^{-\theta}$. Consider the limit in which $K \to \infty$ and $b_t \to \infty$, holding r_t fixed. In this limit,

$$K \frac{x_{nt} b_t^{-\theta}}{x_{nt} b_t^{-\theta} + \Lambda_{t-1}} = \frac{x_{nt} r_t}{x_{nt} \frac{r_t}{K} + \Lambda_{t-1}} \to \frac{x_{nt}}{\Lambda_{t-1}} r_t$$

By the Poisson limit of the binomial distribution, the number of innovations is Poisson with mean $\frac{x_{nt}}{\Lambda_{t-1}} r_t$.

IA7.3 Proof of Proposition 3

Proof. Suppose that z^R grows at rate g_R. Let $(1+g_x) = [(1+\gamma)(1+g_R)]^{\frac{1}{1-\beta}}$. The technology pool can be expressed as

$$\mathcal{A}_{nt} = \sum_{k=1}^{\infty} x_{nt-k} + \sum_{\tilde{n}\neq n}\sum_{k=1}^{\infty}\left[1-\left(1-\varphi_n^A\right)^k\right] x_{\tilde{n}t-k}$$

Along a BGP, $x_{nt-k} = (1+g_x)^{-k} x_{nt}$, so we can express the spillover pool as

$$\begin{aligned}
\mathcal{A}_{nt} &= \sum_{k=1}^{\infty}(1+g_x)^{-k} x_{nt} + \sum_{\tilde{n}\neq n}\sum_{k=1}^{\infty}\left[1-\left(1-\varphi_n^A\right)^k\right](1+g_x)^{-k} x_{\tilde{n}t} \\
&= \frac{1}{g_x} x_{nt} + \sum_{\tilde{n}\neq n}\left(\frac{1}{g_x} - \frac{1-\varphi_n^A}{g_x+\varphi_n^A}\right) x_{\tilde{n}t} \\
&= \frac{1}{g_x}\left[x_{nt} + \sum_{\tilde{n}\neq n}\left(\frac{1+g_x}{g_x+\varphi_n^A}\varphi_n^A\right) x_{\tilde{n}t}\right]
\end{aligned}$$

Define $X_t \equiv \sum_{\tilde{n}} x_{\tilde{n}t}$ and $s_{nt} = \frac{x_{nt}}{X_t}$. Then this is

$$\begin{aligned}
\mathcal{A}_{nt} &= \frac{X_t}{g_x}\left[s_{nt} + \left(\frac{1+g_x}{g_x+\varphi_n^A}\varphi_n^A\right)(1-s_{nt})\right] \\
&= \frac{X_t}{g_x}\left[1-(1-s_{nt})\frac{\left(1-\varphi_n^A\right)g_x}{g_x+\varphi_n^A}\right]
\end{aligned}$$

Define $\kappa_n(g_x) \equiv \log\left[1-(1-s_{nt})\frac{\left(1-\varphi_n^A\right)g_x}{g_x+\varphi_n^A}\right]$, so that $\log\mathcal{A}_{nt} = \log\frac{X_t}{g_x} + \kappa_n(g_x)$.

The variance of log output per capita is

$$\begin{aligned}
\mathrm{Var}(\log y_{nt}) &= \mathrm{Var}\left(u_n^Y + \frac{1}{\theta}\log\mathcal{A}_{nt}\right) \\
&= \mathrm{Var}\left(u_n^Y\right) + \frac{2}{\theta}\mathrm{Cov}\left(u_n^Y, \log\mathcal{A}_{nt}\right) + \frac{1}{\theta^2}\mathrm{Var}(\log\mathcal{A}_{nt}) \\
&= \mathrm{Var}\left(u_n^Y\right) + \frac{2}{\theta}\mathrm{Cov}\left(u_n^Y, \kappa_n(g_x)\right) + \frac{1}{\theta^2}\mathrm{Var}(\kappa_n(g_x))
\end{aligned}$$

We now take a first-order approximation around $g_x \searrow 0$. As $g_x \searrow 0$, the innovation shares converge to the shares implied by research fundamentals:

$$\lim_{g_x\to 0} s_{nt} = \frac{\exp u_n^R}{\sum_{\tilde{n}}\exp u_{\tilde{n}}^R}.$$

Since $\kappa_n(0) = 0$ and

$$\kappa_n'(0) = -\left(1 - \lim_{g_x \to 0} s_{nt}\right) \frac{1 - \varphi_n^A}{\varphi_n^A} = -\left(1 - \frac{\exp u_n^R}{\sum_{\tilde{n}} \exp u_{\tilde{n}}^R}\right) \frac{1 - \varphi_n^A}{\varphi_n^A}$$

we have the following limiting expressions:

$$\begin{aligned}
\lim_{g_x \to 0} \mathrm{Var}\left(\kappa_n(g_x)\right) &= \frac{1}{N} \sum_n \kappa_n(g_x)^2 - \left(\frac{1}{N} \sum_n \kappa_n(g_x)\right)^2 = 0 \\
\lim_{g_x \to 0} \frac{d}{dg_x} \mathrm{Var}\left(\kappa_n(g_x)\right) &= \lim_{g_x \to 0} \frac{1}{N} \sum_n 2\kappa_n(g_x)\kappa_n'(g_x) - 2\left(\frac{1}{N} \sum_n \kappa_n(g_x)\right) \frac{1}{N} \sum_n \kappa_n'(g_x) = 0 \\
\lim_{g_x \to 0} \mathrm{Cov}\left(u_n^Y, \kappa_n(g_x)\right) &= \frac{1}{N} \sum_n u_n^Y \kappa_n(g_x) - \frac{1}{N} \sum_n u_n^Y \frac{1}{N} \sum_n \kappa_n(g_x) = 0 \\
\lim_{g_x \to 0} \frac{d}{dg_x} \mathrm{Cov}\left(u_n^Y, \kappa_n(g_x)\right) &= \lim_{g_x \to 0} \frac{1}{N} \sum_n u_n^Y \kappa_n'(g_x) - \frac{1}{N} \sum_n u_n^Y \frac{1}{N} \sum_n \kappa_n'(g_x) = \mathrm{Cov}\left(u_n^Y, \kappa_n'(0)\right)
\end{aligned}$$

Together, these imply that to a first-order approximation around $g_x \searrow 0$,

$$\begin{aligned}
\mathrm{Var}\left(\log y_{nt}\right) &\approx \mathrm{Var}\left(u_n^Y\right) + g_x \frac{2}{\theta} \mathrm{Cov}\left(u_n^Y, \kappa_n'(0)\right) \\
&= \mathrm{Var}\left(u_n^Y\right) - g_x \frac{2}{\theta} \mathrm{Cov}\left(u_n^Y, \left(1 - \frac{\exp u_n^R}{\sum_{\tilde{n}} \exp u_{\tilde{n}}^R}\right) \frac{1 - \varphi_n^A}{\varphi_n^A}\right)
\end{aligned}$$

Under the maintained covariance condition, this first-order approximation implies that $\mathrm{Var}(\log y_{nt})$ increases with g_x near zero.

Further, since Var$(\log y_{nt})$ =Var$\left(u_n^Y + \frac{1}{\theta} \log \mathcal{A}_{nt}\right)$, We use the following covariance decomposition:

$$1 = \frac{\mathrm{Var}\left(u_n^Y\right) + \mathrm{Cov}\left(u_n^Y, \frac{1}{\theta} \log \mathcal{A}_{nt}\right)}{\mathrm{Var}\left(\log y_{nt}\right)} + \frac{\mathrm{Cov}\left(u_n^Y, \frac{1}{\theta} \log \mathcal{A}_{nt}\right) + \mathrm{Var}\left(\frac{1}{\theta} \log \mathcal{A}_{nt}\right)}{\mathrm{Var}\left(\log y_{nt}\right)}$$

Using the covariance decomposition for $u_n^Y + \frac{1}{\theta} \log \mathcal{A}_{nt}$, we can assign the share accounted for by the technology-pool component as

$$\begin{aligned}
\text{Fraction Accounted for by } \frac{1}{\theta} \log \mathcal{A}_{nt} &= \frac{\mathrm{Cov}\left(u_n^Y, \frac{1}{\theta} \log \mathcal{A}_{nt}\right) + \mathrm{Var}\left(\frac{1}{\theta} \log \mathcal{A}_{nt}\right)}{\mathrm{Var}\left(\log y_{nt}\right)} \\
&\approx \frac{-g_x \frac{1}{\theta} \mathrm{Cov}\left(u_n^Y, \left(1 - \frac{\exp u_n^R}{\sum_{\tilde{n}} \exp u_{\tilde{n}}^R}\right) \frac{1 - \varphi_n^A}{\varphi_n^A}\right)}{\mathrm{Var}\left(u_n^Y\right) - g_x \frac{2}{\theta} \mathrm{Cov}\left(u_n^Y, \left(1 - \frac{\exp u_n^R}{\sum_{\tilde{n}} \exp u_{\tilde{n}}^R}\right) \frac{1 - \varphi_n^A}{\varphi_n^A}\right)}
\end{aligned}$$

This is increasing in g_x.

□

IA7.4 Proof of Proposition 4

Proof. A country's response to a foreign innovation is $\frac{d\log y_{nt+1}}{dx_{j,t}} = \frac{1}{\theta}\frac{d\log \mathcal{A}_{nt+1}}{dx_{j,t}} = \frac{1}{\theta}\frac{\varphi^A_{nj}}{\mathcal{A}_{nt+1}}$. The response to its own innovation is $\frac{d\log y_{nt+1}}{dx_{n,t}} = \frac{1}{\theta}\frac{d\log \mathcal{A}_{nt+1}}{dx_{n,t}} = \frac{1}{\theta}\frac{1}{\mathcal{A}_{nt+1}}$. The ratio is φ^A_{nj}, which is, of course, increasing in φ^A_{nj}. □

IA7.5 Proof of Proposition 5

Proof. Suppose that countries 1 and 2 have the same history of innovations $x_{1t} = x_{2t}$ and that country 1 has proportionally lower adoption probabilities, so that

$$\varphi^A_{1n} = \begin{cases} 1 & n = 1 \\ \alpha\varphi^A_{21} & n = 2 \\ \alpha\varphi^A_{2n} & n \neq 1, 2 \end{cases}$$

for a constant $\alpha \in (0, 1)$. Country 2 has higher output per capita than country 1:

$$\begin{aligned}
\mathcal{A}_{1t} &= \sum_n \sum_{k=1}^{\infty} \left[1 - \left(1 - \varphi^A_{1n}\right)^k\right] x_{nt-k} \\
&= \sum_{k=1}^{\infty} x_{1t-k} + \sum_{k=1}^{\infty} \left[1 - \left(1 - \varphi^A_{12}\right)^k\right] x_{2t-k} + \sum_{n\neq 1,2} \sum_{k=1}^{\infty} \left[1 - \left(1 - \varphi^A_{1n}\right)^k\right] x_{nt-k} \\
&= \sum_{k=1}^{\infty} x_{2t-k} + \sum_{k=1}^{\infty} \left[1 - \left(1 - \alpha\varphi^A_{21}\right)^k\right] x_{1t-k} + \sum_{n\neq 1,2} \sum_{k=1}^{\infty} \left[1 - \left(1 - \alpha\varphi^A_{2n}\right)^k\right] x_{nt-k}
\end{aligned}$$

We can bound $\mathcal{A}_{1t}$ above by using $\alpha < 1$

$$\begin{aligned}
\mathcal{A}_{1t} &\leq \sum_{k=1}^{\infty} x_{2t-k} + \sum_{k=1}^{\infty} \left[1 - \left(1 - \varphi^A_{21}\right)^k\right] x_{1t-k} + \sum_{n\neq 1,2} \sum_{k=1}^{\infty} \left[1 - \left(1 - \varphi^A_{2n}\right)^k\right] x_{nt-k} \\
&= \mathcal{A}_{2t}
\end{aligned}$$

and we can bound $\mathcal{A}_{1t}$ below using $\left[1-\left(1-\alpha\varphi_{n\tilde{n}}^{A}\right)^{k}\right] \geq \alpha\left[1-\left(1-\varphi_{n\tilde{n}}^{A}\right)^{k}\right]$:[12], giving

$$\begin{aligned} \mathcal{A}_{1t} &\geq \sum_{k=1}^{\infty} x_{2t-k} + \alpha \sum_{k=1}^{\infty}\left[1-\left(1-\varphi_{21}^{A}\right)^{k}\right] x_{1t-k} + \alpha \sum_{n\neq 1,2}\sum_{k=1}^{\infty}\left[1-\left(1-\varphi_{2n}^{A}\right)^{k}\right] x_{nt-k} \\ &\geq \alpha \mathcal{A}_{2t} \end{aligned}$$

The first inequality, $\mathcal{A}_{1t} \leq \mathcal{A}_{2t}$ implies that country 2 is richer than country 1. The second, $\mathcal{A}_{1t} \geq \alpha\mathcal{A}_{2t}$, implies that country 2 has a larger one-period-ahead response to a third country j's innovation

$$\frac{d\log y_{1t+1}}{dx_{j,t}} = \frac{1}{\theta}\frac{d\log \mathcal{A}_{1t+1}}{dx_{j,t}} = \frac{1}{\theta}\frac{\varphi_{1j}^{A}}{\mathcal{A}_{1t+1}} = \frac{1}{\theta}\frac{\alpha\varphi_{2j}^{A}}{\mathcal{A}_{1t+1}} \leq \frac{1}{\theta}\frac{\varphi_{2j}^{A}}{\mathcal{A}_{2t+1}} = \frac{1}{\theta}\frac{d\log \mathcal{A}_{2t+1}}{dx_{j,t}} = \frac{d\log y_{2t+1}}{dx_{j,t}}$$

$\square$

[12]For $\alpha \in (0,1)$, $\left[1-\left(1-\alpha\varphi_{n\tilde{n}}^{A}\right)^{k}\right] \geq \alpha\left[1-\left(1-\varphi_{n\tilde{n}}^{A}\right)^{k}\right]$ because $1-\left(1-\alpha\varphi_{n\tilde{n}}^{A}\right)^{k}$ is concave in α, which implies

$$1-\left(1-\alpha\varphi_{n\tilde{n}}^{A}\right)^{k} = 1-\left(1-\left[\alpha\varphi_{n\tilde{n}}^{A}+(1-\alpha)\,0\right]\right)^{k} \geq \alpha\left[1-\left(1-\varphi_{n\tilde{n}}^{A}\right)^{k}\right]+(1-\alpha)\left[1-(1-0)^{k}\right] = \alpha\left[1-\left(1-\varphi_{n\tilde{n}}^{A}\right)^{k}\right]$$