WORKING GROUP ON THE **FOUNDATIONS** OF
**LONG-RUN PROSPERITY**
HOOVER INSTITUTION | STANFORD UNIVERSITY

# What Are the Foundations of Long-Run Prosperity?

By Alexander Galetovic, Stephen Haber, and Amit Seru

## The Puzzle of Prosperity

In 1776, Adam Smith remarked that "China has been long one of the richest, that is, one of the most fertile, best cultivated, most industrious, and most populous countries in the world. It seems, however, to have been long stationary."[1] Smith's insight was prescient. Within a few decades, Britain's rapid industrialization allowed it easy victories over China in the Opium Wars. China's policy response, the Self-Strengthening Movement of 1865–95, which sought to build the manufacturing base necessary to defend itself, proved to be too little, too late. Over the next four decades it was invaded three times by rapidly industrializing Japan. Those defeats catalyzed the collapse of the Chinese state in 1911, ushering in decades of civil war that only ended with the communist victory of 1949.

The Sino-British and Sino-Japanese reversals in prosperity are not sui generis. On the one hand, countries that were stunningly poor within modern memory are now among the world's most prosperous. In 1960, Spain was poorer than Mexico, South Korea was poorer than the Central African Republic, Singapore was poorer than El Salvador, Taiwan was poorer than the Democratic Republic of the Congo, and Ireland was poorer than Venezuela.[2] On the other hand, countries that were on the way to prosperity stagnated. Italy and Greece have the same per capita

[1] Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations* (Hartford, CT: Cooke & Hale, 1818), 50, https://www.google.com/books/edition/An_Inquiry_Into_the_Nature_and_Causes_of/1v8-AAAAYAAJ?hl=en&gbpv=1.

[2] Per capita income data from Penn World Tables, version 9.1. Comparisons across countries in 1960 are output-side real GDP per capita at current PPPs in 2011 US dollars.

incomes today as they did in 2000. Ukraine is now poorer than it was at independence from the USSR in 1991. The Venezuelan economy has been shrinking since 1977, a contraction that is especially remarkable considering that Venezuela's proven oil reserves exceed those of Saudi Arabia and that it began two decades before the Bolivarian socialist revolution of Hugo Chávez.[3]

What is true about reversals in prosperity across countries is also true about regions within them. Circa 1920, Cleveland's automobile, machine tool, steel, electrical machinery, and chemical industries made it one of the wealthiest cities in the United States. Until the 1960s, Birmingham, Alabama, was an industrial powerhouse, earning it the nickname "The Magic City." Upstate New York boasted flourishing manufacturing hubs, such as Rochester, Buffalo, and Schenectady. During the heyday of those Rust Belt cities, the region now known as Silicon Valley—one of the most prosperous places on the planet today—was known primarily for its apricot orchards.

The rise and decline of prosperity is not solely a modern phenomenon. There was, for example, a long period of economic efflorescence in the Eastern Mediterranean in antiquity. As Josiah Ober has shown, regardless of the metric that one applies, standards of living in Greek city states during the fourth and fifth centuries BC approached that of the Netherlands in the mid-eighteenth century AD.[4] A similar economic efflorescence took place in Rome from the third century BC through the second century AD. By the end of that period, the city of Rome housed one million people, a population not reached again until the middle of the twentieth century. These periods of Eastern Mediterranean prosperity are especially striking considering that they occurred without the benefit of fossil-fuel-based production and transport technologies. Rather, the technological changes that underpinned them were organizational, legal, and political.

Understanding how and why economies prosper, stagnate, and wither is a question of first-order importance, but our knowledge about this process is surprisingly thin. Indeed, if economic growth was well understood as a scholarly matter, policies based on theory and empirics would have produced *convergence* in levels of economic development across countries and regions over

[3] Per capita income data from Penn World Tables, version 9.1. Time series comparisons within countries are real GDP at constant 2011 national prices per capita in 2011 US dollars.

[4] Josiah Ober, *Democracy and Knowledge: Innovation and Learning in Classical Athens* (Princeton, NJ: Princeton University Press, 2010).

the past century, rather than *divergence*. To be concrete, the per capita income today of Australia would not be fifty times that of Liberia—a gap that is especially puzzling considering that Australia was founded as a forced-labor penal camp run by the dregs of the British army, while Liberia was founded as a republic with a constitution modeled on that of the United States, and prior to that had remained outside of the Atlantic slave trade.

**How Scholars Learned More and More About Less and Less**

The question of why some countries are rich while others are poor was one of the central concerns of the late-eighteenth- and early-nineteenth-century field of political economy. The basic answer offered by Adam Smith, David Ricardo, and the other scholars who defined that field—now usually referred to as "classical economics"—is that growth occurs when there are opportunities for people and firms to specialize, and specialization happens when they are free to trade with one another. A necessary condition for that to happen is a government that is a neutral arbiter of contracts and that does not itself prey upon markets.

As economics, political science, history, and sociology began to emerge as independent disciplines during the late nineteenth and early twentieth centuries, economic growth faded as a central concern of scholarship. Growth, at least in those countries with universities hosting the new departments, appeared to be taking care of itself. "Neoclassical economics," as the dominant strain of economic research and teaching came to be known, instead focused on the mechanics by which markets determine equilibrium output and prices.

National Accounting and the Study of Productivity Growth

It was not until the 1950s that the discipline of economics began to focus once again on the puzzle of economic growth. This interest was driven, in part, by one of the central lessons of the Second World War: victory was largely a function of which side could produce more guns, planes, tanks, and ships. It was also driven, however, by the Cold War; impoverished countries were fertile ground for the spread of Communism.

The focus of research on economic growth during the immediate post-war period was largely empirical, drawing heavily on the newly developed field of national accounting. One line of scholarship, most closely identified with Simon Kuznets, generated estimates of the rate, structure, and spread of modern economic growth (the increase in output that came from shifting

capital and labor toward manufacturing, mining, and services).[5] A central finding of this research was that the fast rates of economic growth observed in the United States, Western Europe, and Japan during the nineteenth and twentieth centuries were associated with a marked increase in capital intensity. That finding generated, however, another puzzle: if the answer to the question of growth was more capital, then why did capital tend to remain in countries that were capital abundant, such as the United States, rather than flow to places that were capital scarce, where its marginal productivity would be higher, such as El Salvador?

Another prominent line of research, pioneered by Moses Abramovitz and Robert Solow, decomposed economic growth among the contributions of factor accumulation (capital and labor) and total factor productivity (TFP, the efficiency with which economies use capital and labor).[6] A central finding of this research is that economic growth is driven mostly by TFP growth. This finding raised yet another puzzle: if improvements in TFP are, as Charles R. Hulten puts it, "costless improvements in the way an economy's resources of labor and capital are transformed into real GDP (the proverbial manna from heaven),"[7] how do we explain why this happens in some economies but not others?

Zvi Griliches and Dale Jorgenson provided a resolution to this puzzle in the 1960s and 1970s.[8] By correcting for measurement errors in the earlier literature on TFP, they found that most economic growth could be accounted for by factor accumulation. This result required, however, that researchers include knowledge as a factor of production in addition to capital and labor. It also prompted yet another puzzle: why does knowledge not move freely across countries and regions within them

---

[5] Simon Kuznets, *Modern Economic Growth: Rate, Structure, and Spread* (New Haven, CT: Yale University Press, 1966).

[6] Moses Abramovitz, "Resource and Output Trends in the United States since 1870," *American Economic Review* 46, no. 2 (1956): 5–23, http://www.jstor.org/stable/1910656; Robert Solow, "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics* 39, no. 3 (1957): 312–20, https://doi.org/10.2307/1926047.

[7] Charles R. Hulten, "Total Factor Productivity. A Short Biography," in *New Developments in Productivity Analysis*, ed. Charles R. Hulten, Edwin R. Dean and Michael J. Harper(University of Chicago Press, 2001), 8n5.

[8] Dale Jorgenson and Zvi Griliches, "The Explanation of Productivity Change," *Review of Economic Studies* 34, no. 3 (1967): 249–83, https://doi.org/10.2307/2296675.

The Absence of a Consensus Theory

Economic theorists advanced two explanations to the knowledge puzzle, but no consensus emerged. The challenge was that theorists had to write mathematical models that were tractable, which pushed them in the direction of single explanators and representative agents.[9] One line of theorizing, most closely associated with Robert Lucas, is that the accumulation of skills by agents, referred to as "human capital," is a source of growth independent from labor and physical capital.[10] A related line of theorizing, most closely associated with Paul Romer, is that investments by agents in ideas and knowledge give rise to the accumulation of technology that is nonrival and not subject to diminishing returns, which he termed "endogenous economic growth."[11] Neither theory explains, however, why agents invest in human capital in some countries but not in others, or why knowledge accumulates as a factor of production in some countries but not in others. They also do not explain why reversals in prosperity are so common.

The response to this puzzle came not from economic theory but from economic history. In a line of research that harkens back to classical economics, but whose modern iteration is associated with Douglass North, economic historians pointed out that investments—and hence factor accumulation—do not take place in the absence of legal and political institutions that specify and enforce private property rights.[12] A related line of thinking, most closely associated with Kenneth Sokoloff, is that investments that generate knowledge do not take place in the absence of legal and political institutions that specify and enforce intellectual property rights, most particularly US-style patent systems.[13]

[9] A representative agent model assumes that agents act in such a way that the sum of their choices is mathematically equivalent to the decision of one individual.

[10] Robert Lucas, "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22, no. 1 (1988): 3–42, https://doi.org/10.1016/0304-3932(88)90168-7.

[11] Paul Romer, "Endogenous Technological Change," *Journal of Political Economy* 98, no. 5 (1990): S71–102, http://www.jstor.org/stable/2937632.

[12] See, for example, Douglass North, *Institutions, Institutional Change and Economic Performance* (Cambridge: Cambridge University Press, 1990).

[13] B. Zorina Khan and Kenneth Sokoloff, "The Early Development of Intellectual Property Institutions in the United States," *Journal of Economic Perspectives* 15, no. 3 (2001): 233–46, http://www.jstor.org/stable/2696565.

Quasi Experiments and the Search for Single Causes

The absence of a consensus theoretical framework meant that the empirical study of growth went in several different directions in the 1990s and early 2000s. Partially responding to the single explanator models written by theorists, and partly responding to a general trend in microeconomics toward causal identification, the empirical literature tended to focus on explanations built around a single cause of growth. One line of empirical research focused on institutions as the fundamental cause,[14] another focused on human capital, a third on technology, a fourth on inequality, and a fifth on geography as the fundamental cause of growth.

These different lines of research tended to have one thing in common: a quasi experiment based on econometric techniques such as difference in differences, instrumental variables, or regression discontinuity. As clever as some of these designs are, they could not overcome the problem of data that had been gathered by institutional actors for ends other than social science research—and that were therefore weak proxies for the variables of interest. How, for example, might a researcher with a data set on years of schooling by country know whether the data captured the concept of human capital (specialized skills owned by agents), the concept of technology (knowledge that is nonrival and only partially excludable), or the concept of institutions (publicly funded education)? Making matters more difficult, the data sets tended to be truncated with respect to time, truncated with respect to space, and characterized by nonrandom missingness and nonrandom measurement error.

Problems with data quality were compounded by uncertainty about model specification. The goal of a quasi experiment is to replicate a medical experiment in which a treatment causes an outcome, holding all else constant. In a medical setting, experiments satisfy the treatment condition through random assignment, and satisfy the "all else constant" condition through

[14] See, for examples, Robert Hall and Charles Jones, "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics* 114 (1999): 83–116, https://doi.org/10.1162/003355399555954; Daron Acemoglu, Simon Johnson, and James Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review* 91, no. 5 (2001): 1369–1401, http://www.jstor.org/stable/2677930; Dani Rodrik, Arvind Subramanian, and Francisco Trebbi, "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development," *Journal of Economic Growth* 9, no. 2 (2004): 131–65, http://www.jstor.org/stable/40212696.

random sampling. In studies of long-run economic growth researchers must satisfy these requirements by making simplified assumptions about the nature of the data-generating process and about the mechanics of the larger system in which the treatment and outcome are variables. Researchers are basically forced to assume that a reduced-form regression can cleanly identify the effect of one variable on another without knowing the identity of all the exogenous variables and their distributions, the identity of all the endogenous variables and their distributions, and the functional form of the interactions among the exogenous variables and the feedbacks among the endogenous variables. In the early 1990s researchers pointed out that the results of reduced-form growth regressions were fragile.[15] More recently, scholars have pointed out that the results from quasi experiments are no less fragile.[16]

Turning toward Field Experiments . . . and Away from the Study of Economic Growth

The response of the discipline to the challenges posed by observational data and model uncertainty has been randomized controlled trials (RCTs) in the field. While much has been learned through field experiments, it is difficult as a practical matter to satisfy the "all else constant" requirement when there is unobserved heterogeneity across study sites. As a result, the results of many field experiments cannot be replicated.[17]

More fundamentally, field experiments are not designed to answer the question of why some countries are rich while others are poor. Rather, they are designed to estimate the average effect of a specific treatment on a specific outcome over the short run, such as the provision of mosquito nets on school attendance. While one might conceivably add up the estimated causal

[15] Ross Levine and David Renelt, "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review* 82, no. 4 (1992): 942–63, http://www.jstor.org/stable/2117352.

[16] Morgan Kelly, "Understanding Persistence," working paper 2020, available at SSRN: https://ssrn.com/abstract=3688200; Alwyn Young, "Leverage, Heteroskedasticity and Instrumental Variables in Practical Application," working paper 2021, available at https://personal.lse.ac.uk/YoungA/LeverageandIV.pdf; Apoorva Lal, Max Lockhart, Yiqing Xu, and Ziwen Zu, "How Much Should We Trust Instrumental Variables Estimates in Political Science? Practical Advice Based on Over 60 Replicated Studies," working paper 2021, available at https://yiqingxu.org/papers/english/2021_iv/LLXZ.pdf.

[17] Nancy Cartwright and Jeremy Hardie, *Evidence Based Policy: A Practical Guide to Doing It Better* (Oxford: Oxford University Press, 2012); Dean Karlan and Jacob Appel, *Failing in the Field: What We Can Learn When Field Research Goes Wrong* (Princeton, NJ: Princeton University Press, 2016).

effects from thousands of experimental treatments, doing so would neither account for the fifty-to-one difference in per capita income we observe across countries nor explain how and why those differences emerged and persist.[18]

**The Sources of Prosperity Redux: Opportunities from Cognate Fields**

Recent advances in several fields of scholarship have opened opportunities for research into the study of long-run prosperity. Some of these advances—such as the analysis of complex adaptive systems, machine learning, spatial econometrics using geographic information system (GIS) and remote sensing data, and the emerging field of comparative political economy—have come from scholarly cross-pollination. Other advances, such as a renewed focus on the theory of economic rent, have come from within the discipline of economics itself.

Complex Adaptive Systems

There is significant debate among scientists about the proper way to determine causality. One view is that the task is to identify the independent effect of a specific phenomenon (a cause variable) on an outcome (an effect variable) by analyzing the response of the effect variable when the magnitude of the cause variable is changed.

Another view is that the task is to identify a vector of exogenous factors that increase the probability that a particular constellation of features would emerge from a complex adaptive system. In this approach, the features of the system—its endogenous, observable outcomes—are understood as emergent properties (simultaneously determined phenomena that have properties not found in the system's component parts).[19] It is therefore not meaningful to identify the marginal effect of the features on one another. An example from evolutionary biology illustrates the intuition. For most biologists, it is not meaningful to identify how much of rabbits' speed derives from the fact that coyotes hunt them and how much derives from the fact that they live

[18] Angus Deaton, "Randomization in the Tropics Revisited: A Theme and Eleven Variations," NBER Working Paper 27600 (2020), https://doi.org/10.3386/w27600.

[19] Emergence happens because of interactions among exogenous factors and feedbacks among endogenous variables over time. A canonical example is a pancake: its fluffy texture is not to be found in any of its ingredients; it is the outcome of interactions among the ingredients and feedbacks that take place in the batter in the presence of heat.

on grasslands. As Deborah Gordon points out, the task is to identify the vector of exogenous factors that increased the probability of a specific process of coevolution whose emergent properties include fast rabbits, clever coyotes, and tall grass—such that one place has lots of them, while another place has none.[20]

The etiology of identifying single, fundamental causes has underpinned much of the empirical research in economics and political science over the past two decades. It is why, for example, scholars have been engaged in a two-decade-long debate as to whether *the* fundamental cause of economic growth is inequality, human capital, institutions, technology, or geography (see p. 5).

It would not take lengthy argumentation to show that four of the five variables listed above as fundamental causes of growth are themselves endogenous outcome variables. Much like fast rabbits and clever coyotes on grasslands, levels of inequality, stocks of human capital, rates of technological absorption, distributions of power, and the properties of governance institutions fed back on one another over time within societies; they coevolved. Estimating the marginal effect of a change in any one of them (as a cause variable) on an outcome (an effect variable), without taking into account how that change ramified through the entire system will yield biased results. This is especially the case when there exists a substantial time lag between a cause and an effect, as is the case in the study of long-run economic growth. It would also not take lengthy argumentation to show that the one exogenous variable, geography, cannot by itself be a cause of long-run prosperity—for the simple reason that a variable that does not exhibit time series variation cannot by itself explain variance across units over time.

The task at hand is to think about the set of outcomes that we call poverty and prosperity as simultaneously determined emergent properties of different complex adaptive systems. The application of this way of thinking to the question of economic growth is, in fact, obvious once one considers that a market economy is perhaps the quintessential example of a complex adaptive system: no one guides it in any meaningful sense of the word. Rather, millions of agents make countless small decisions every day over long periods of time without any central

[20] Deborah Gordon, "The Ecology of Collective Behavior" *PLoS Biology* 12, no. 3 (2014): e1001805, https://doi.org/10.1371/journal.pbio.1001805.

coordination. The challenge is to identify the vector of factors that push systems to evolve in one direction rather than another.[21]

Machine Learning, Remote Sensing, and Geospatial Analysis

Scholarly cross-pollination at the intersection of computer science and statistics has given rise to new tools such as machine learning, remote sensing, and geospatial analysis. Machine learning is usually thought of in the context of "big data" (the identification of subtle patterns and relationships in massive data sets). Machine learning also, however, allows researchers to draw more reliable inferences than would be possible through traditional econometric methods when sample sizes are small and there is model uncertainty. Remote sensing allows researchers to measure phenomena with a spatial dimension at high resolution, and to then measure variance in that spatial dimension over time. The related field of geospatial analysis allows researchers to acquire, manage, analyze, and visualize data that were previously unrelatable by integrating location data (where things are) with descriptive information (what things are like there), and temporal data (the date and time of an occurrence).

The combination of machine learning, remote sensing, and geospatial analysis has opened opportunities for research that previously did not exist. Remote sensing has produced prodigious amounts of data. Machine learning applications allow researchers to mimic the interactions among endogenous variables in a complex system. Geospatial analysis, when coupled to machine learning, allows researchers to specify geographic or climatologic data as exogenous variables in a complex system.

The Emerging Field of Comparative Political Economy

Scholarly cross-pollination among political scientists, economic historians, and financial economists has yielded an emerging new field of comparative political economy.

[21] See, for example, Stephen Haber, Roy Elis, and Jordan Horrillo, "The Ecological Origins of Economic and Political Systems," working paper (2021), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3958073.

Economic historians and political scientists interested in the question of why Western Europe came to dominate the rest of the world—and why Britain came to dominate Western Europe—made a convincing case in the 1990s that the ability to specify and enforce financial contracts was a crucial factor.[22] Financial economists interested in the sources of economic growth made a convincing case in the 1990s and early 2000s that the size, structure, and efficiency of financial systems have first-order impacts on poverty, income distribution, entrepreneurship, and firm entry. They also found that financial systems do not respond automatically to the demand for finance: the regulatory and supervisory policies that govern banks and securities markets have a first-order effect on the size, structure, and efficiency of financial systems.[23] Scholarly cross-pollination among economic historians, political scientists, and financial economists subsequently yielded a body of research showing that the regulatory and supervisory policies that govern markets are not the product of robots programmed to maximize social welfare. They are the results of deals among interested parties that determine which laws are passed, which regulations are enacted, and which groups are given forbearance. It follows that banks and securities markets may be regulated and supervised according to technical criteria, and banking contracts are enforced according to abstruse laws, but those criteria and laws are the outcomes of a political process—a game as it were—with the stakes being wealth and power.[24]

Over the past two decades comparative political economy become a lively field of scholarly inquiry addressing multiple research areas related to the foundations of long-run growth. One line of research, strongly influenced by the quasi-experimental growth literature, has used econometric tools to historicize the study of institutions and economic growth.[25]Another line of

[22] See, for example, Douglass North and Barry Weingast, "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England," *Journal of Economic History* 49, no. 4 (1989): 803–32, http://www.jstor.org/stable/2122739.

[23] See, for examples, Robert King and Ross Levine, "Finance and Growth: Schumpeter Might Be Right," *Quarterly Journal of Economics* 103, no. 3 (1993): 717–37, https://doi.org/10.2307/2118406; Thorsten Beck, Norman Loayza, and Ross Levine, "Finance and the Sources of Growth," *Journal of Financial Economics* 58, nos. 1–2 (2000): 261–300, https://doi.org/10.1016/S0304-405X(00)00072-6.

[24] See, for example, James Barth, Gerard Caprio, and Ross Levine, *Rethinking Bank Supervision and Regulation: Until Angels Govern* (Cambridge: Cambridge University Press, 2006); Charles Calomiris and Stephen Haber, *Fragile by Design: The Political Origins of Financial Crises and Scarce Credit* (Princeton: NJ: Princeton University Press, 2014).

[25] See for examples, Daron Acemoglu, Simon Johnson, and James Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review* 91, no. 5 (2001):

research, influenced by sociology but using the tools of history and economics, has focused on the origins of cultural values and their consequences for long-run economic outcomes.[26] Yet a third line of research has focused on markets, and the gains from specialization they permit, as emergent phenomena that are more likely to occur under forms of political organization associated with freedom of individual action.[27]

New Insights from Classical Economics: Ricardian Rents as a Source of Innovation

A widely held belief is that innovation is driven by market power rents. The purpose of intellectual property, according to this view, is to allow innovators to create temporary monopolies. Some scholars have gone so far as to argue that market power is necessary for technological growth to happen. Innovation therefore has two faces: on the one hand, it expands the variety of goods and services that are available, thereby raising living standards; but on the other hand, innovators have an incentive to constrain output to generate the market power rents that compensate them for investments in research and development (R&D). This "intellectual property as monopoly" view of innovation provides the basis for antitrust intervention by the Federal Trade Commission and the Department of Justice in technology markets as well as antitrust actions against firms that specialize in technology development.

A quite different view of the source of innovation—that they are driven by the search for Ricardian rents, not market power rents—has emerged in recent years.[28] According to the

1369–1401, https://www.aeaweb.org/articles?id=10.1257/aer.91.5.1369; Abhijit Banerjee and Lakshmi Iyer, "History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India," *American Economic Review* 95 (2005), http://www.jstor.org/stable/4132711; Nathan Nunn, "The Long-Term Effects of Africa's Slave Trades," *Quarterly Journal of Economics* 123 (2008), http://dx.doi.org/10.1162/qjec.2008.123.1.139.

[26] See, for examples, Luigi Guiso, Paola Sapienza, and Luigi Zingales, "Long Term Persistence," *Journal of the European Economic Association* 14 (2016); Vasilki Fouka and Alain Schlaepfer, "Agricultural Labor Intensity and the Origins of Work Ethics," *Economic Journal* 130 (2020): 1081–1113, https://doi.org/10.1093/ej/ueaa029.

[27] See, for example, S. R. Epstein, *Freedom and Growth: The Rise of States and Markets in Europe, 1300–1750* (London and New York: Routledge, 2000).

[28] For its clearest articulation to date, see Alexander Galetovic, "Patents in the Semiconductor Industry: The Ricardian Hypothesis," in *The Battle over Patents: History and the Politics of Innovation,* ed. Stephen Haber and Naomi Lamoreaux (Oxford: Oxford University Press, 2021), 27–68.

Ricardian rents hypothesis, firms operate in highly competitive markets and compete through innovations designed to produce more revenue per dollar of input than the least productive producer in that same market. They do so by creating new products and services, as well as by making improvements to existing products and services. Importantly, firms will refrain from undertaking the costly investments in R&D that yield new (or better) products and services if their competitors can free ride by copying their innovations without compensation. In the Ricardian rent view of innovation, therefore, intellectual property does not generate market power; it prevents free riding and creates the basis for markets in patents.

The intellectual origins of the Ricardian rents hypothesis dates to the observation by classical economist David Ricardo that differential levels of output in agriculture are a function of differences in land quality, which translate, in turn, into differential rental rates for land. In the 1990s, Gene M. Grossman and Elhanan Helpman foreshadowed the application of Ricardian rents to modern innovative industries by observing that the upgrading of consumer products—and, importantly, the inputs to produce them—plays a central role in raising standards of living.[29] Researchers at the Hoover Institution in the 2010s then found no evidence that the innovative firms that develop essential technologies in the smartphone industry—one of the world's canonical, intellectual-property-intense industries—earn market power rents. This prompted them to probe the theory that underpins the market power hypothesis. They found that it is neither logically consistent nor logically complete.[30]

The Ricardian rents hypothesis explains why you can read, and even listen to, this document on a laptop computer, tablet, or smartphone made by any number of manufacturers that integrate technologies from thousands of patents developed by dozens of firms. If any of the firms that

[29] See Gene M. Grossman and Elhanan Helpman, "Quality Ladders in the Theory of Growth," *Review of Economic Studies* 58 (1991): 43–61, https://doi.org/10.2307/2298044. A canonical example of the phenomenon is the replacement of horses and carriages with automobiles, and then the successive improvements in the comfort, safety, power, and fuel efficiency of automobiles.

[30] Alexander Galetovic and Stephen Haber, "The Fallacies of Patent Holdup Theory," *Journal of Competition Law and Economics* 13:1 (2017), https://doi.org/10.1093/joclec/nhx006; Alexander Galetovic and Stephen Haber, "SEP Royalties: What Theory of Value and Distribution Should Courts Apply?" *Ohio State Technology Law Journal* 17 (2021), https://www.law.berkeley.edu/wp-content/uploads/2021/05/Galetovic_Haber.pdf.

owned those patents operated as a monopolist, the royalty paid by the manufacturer of your device to the patent owner would reflect that monopoly, and the manufacturer would pass along the cost to you. This is not what researchers (and consumers) observe; the quality-adjusted prices of information technology devices have been falling at better than 10 percent per year for more than a decade.[31] Indeed, they are now so low that parents give them to children as toys.

The Ricardian rents hypothesis also sheds light on human capital and its accumulation. Most Ricardian rents are generated by an entrepreneur seeing a market opportunity and then recruiting teams of individuals with diverse talents and complementary skill sets—including design, engineering, finance, law, government affairs, accounting, manufacturing, and marketing—into a single firm. Those individuals made necessary and costly investments in their human capital in the expectation that they will share in the Ricardian rents generated by the firm.

Finally, the Ricardian rents hypothesis sheds light on why market economies inevitably generate demand for redistribution through the political system. If the rents from innovation are earned by individuals with high levels of human capital, then there exists a fundamental tension within societies between innovation and the distribution of income. It therefore explains why there are no high-innovation economies that do not, in one form or another, engage in redistributive transfers. What varies across those societies is not the existence of the transfers; it is whether the form it takes throttles the engine that generates prosperity.

[31] See Alexander Galetovic, Stephen Haber, and Ross Levine, "An Empirical Examination of Patent Holdup," *Journal of Competition Law and Economics* 11: 3 (2015), https://doi.org/10.1093/joclec/nhv024.