This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: The Rate and Direction of Inventive Activity Revisited

Volume Author/Editor: Josh Lerner and Scott Stern, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-47303-1; 978-0-226-47303-1 (cloth)

Volume URL: http://www.nber.org/books/lern11-1

Conference Date: September 30 - October 2, 2010

Publication Date: March 2012

Chapter Title: The Consequences of Financial Innovation: A Counterfactual Research Agenda

Chapter Authors: Josh Lerner, Peter Tufano

Chapter URL: http://www.nber.org/chapters/c12370

Chapter pages in book: (p. 523 - 575)

11

The Consequences of Financial Innovation
A Counterfactual Research Agenda

Josh Lerner and Peter Tufano

The significance of financial innovation has been widely touted. Many leading scholars, including Miller (1986) and Merton (1992), highlight the importance of new products and services in the financial arena, sometimes characterizing these innovations as an "engine of economic growth."

At several levels, these arguments are plausible. Financial innovations can be seen as playing a role akin to that of the "general purpose technologies" delineated by Bresnahan and Trajtenberg (1995) and Helpman (1998): not only do these breakthroughs generate returns for the innovators, but they have the potential to affect the entire economic system and can lead to far-reaching changes. For instance, these innovations may have broad implications for households, enabling new choices for investment and consumption, and reducing the costs of raising and deploying funds. Similarly, financial innovations enable firms to raise capital in larger amounts and at a lower cost than they could otherwise, and in some cases (for instance, biotechnology start-ups) to obtain financing that they would otherwise simply be unable to raise. This latter idea is captured in a recent model of economic growth by

Josh Lerner is the Jacob H. Schiff Professor of Investment Banking at Harvard Business School, with a joint appointment in the Finance and the Entrepreneurial Management Units, and a research associate and codirector of the Productivity, Innovation, and Entrepreneurship Program at the National Bureau of Economic Research. Peter Tufano is the Peter Moores Dean and professor of finance at the Saïd Business School at the University of Oxford and a research associate and codirector of the Household Finance Working Group at the National Bureau of Economic Research.

We would like to thank Bob Hunt, Bill Janeway, Joel Mokyr, Antoinette Schoar, Scott Stern, and participants at the American Economic Association's 2010 Meeting, the National Bureau of Economic Research's Rate and Direction of Inventive Activity Preconference and Conference, and Brown University's Conference on Financial Innovation for their helpful comments. We thank the Division of Faculty Research and Development at the Harvard Business School for support of this project.

Michalopoulos, Laeven, and Levine (2010), who argue that growth is driven not just by profit-maximizing entrepreneurs who spring up to commercialize new technologies, but also by the financial entrepreneurs who develop new ways to screen and fund the technologists.

Moreover, it appears that financial innovation is ubiquitous. Tufano (1995, 2003) shows that far from being confined to the last few decades, financial innovation has been part of the economic landscape for centuries. Goetzmann and Rouwenhorst (2005) document nineteen major financial innovations that span the past 4,000 years, ranging from the innovation of interest to creation of Eurobonds. Not only is financial innovation an historical phenomena, it is also a widespread one. For example, Tufano (1989) shows that of all public offerings in 1987, 18 percent (on a dollar-weighted basis) consisted of securities that had not been in existence in 1974.

But at the same time, claims of the beneficial impacts of financial innovations must be approached with caution. One reason is that despite the acknowledged economic importance of financial innovation, the sources of such innovation remain poorly understood, particularly empirically. In a recent review article, Frame and White (2004) are able to identify only thirty-nine empirical studies of financial innovation. Moreover, this literature concentrates largely on the "back end" of the innovation process, focusing on the diffusion of these innovations, the characteristics of adopters, and the consequences of innovation for firm profitability and social welfare. Frame and White identify only two papers on the origins of innovation, namely, Ben-Horim and Silber (1977) and Lerner (2002).

The paucity of research in this area contrasts sharply with the abundant literature on the sources of manufacturing innovation. This neglect is particularly puzzling given the special circumstances surrounding financial innovation. Several considerations—discussed in detail in section 11.3—suggest that the dynamics of financial innovation are quite different from those in manufacturing. Together, these considerations suggest the need to examine financial innovation as a phenomenon in its own right.

The second reason for caution has been the recent crisis in the global financial system, which has shaken many economists' faith in the positive effects of financial innovation. Certainly, in many post mortems of the crisis, financial innovation was seen as far from an "engine of economic growth." For instance, Levitin characterized recent changes in retail financial services as "negative innovations," such as "opaque pricing, including billing tricks and traps . . . that encourag[e] unsafe lending practices." A similar theme was sounded by Krugman (2007) in regards to securities regulation:

> [T]he innovations of recent years—the alphabet soup of C.D.O.'s and S.I.V.'s, R.M.B.S. and A.B.C.P.—were sold on false pretenses. They were promoted as ways to spread risk, making investment safer. What they did instead—aside from making their creators a lot of money, which they didn't have to repay when it all went bust—was to spread confusion, luring investors into taking on more risk than they realized.

Given this unsettled but huge territory, it is premature to provide definitive answers regarding the causes and consequences of financial innovations and how they differ from the much better understood innovation process in the manufacturing sector. Indeed, a number of observers have pointed out recently that financial innovations are neither all bad nor all good, but contain a mixture of elements (e.g., Johnson and Kwok 2009; Litan 2010; Mishra 2010).

There are many different research approaches to understanding financial innovation, including empirical studies, theoretical models, and traditional historical descriptions. Each has advantages and disadvantages, which we discuss later. In this chapter, our goal is to lay out a complementary research agenda, which we hope will encourage subsequent scholars. After we review the definition of financial innovation, we turn to three general observations about how financial innovation is similar to and different from other forms of innovation—and which inform the limitations of standard research methods. We then consider three case studies of particular innovations and highlight both what is known and unknown about their consequences.

The original *Rate and Direction* volume was published in 1962. Just two years later, Robert W. Fogel, a future Nobel laureate in economics, published his masterpiece *Railroads and American Economic Growth.* In it, Fogel advanced a method, now used in history, political science, and economic history, to consider counterfactual histories. In a counterfactual analysis, the researcher (a) posits a set of plausible counterfactuals and how they might have come to pass; and (b) evaluates metrics to establish the implications of these alternative historical paths. We suggest how this method, while seemingly imprecise and controversial, can be used to better understand financial innovation. We also discuss the limitations of this method. In our conclusion, we suggest avenues for future exploration.

11.1 Background on Financial Innovation

Much of the theoretical and empirical work in financial economics considers a highly stylized world in which there are few types of securities (e.g., debt and equity) and a handful of simple financial institutions, such as banks or exchanges. In reality there are a vast range of different financial products, many different types of financial institutions, and a variety of processes that these institutions employ to do business. The literature on financial innovation must grapple with this real-world complexity.

Financial innovation is the act of creating and then popularizing new financial instruments, as well as new financial technologies, institutions, and markets. The innovations are sometimes divided into product or process variants, with product innovations exemplified by new derivative contracts, new corporate securities, or new forms of pooled investment products, and process improvements typified by new means of distributing securities, pro-

cessing transactions, or pricing transactions. In practice, even this innocuous differentiation is not clear, as process and product innovations are often linked. Innovation includes the acts of invention and diffusion, although in point of fact these two are related, as most financial innovations are evolutionary adaptations of prior products.

As noted before, one of the major challenges associated with the study of financial innovation is the lack of data. Studies of manufacturing innovation traditionally focus on R&D spending and patenting. Given the rarity with which financial service firms report R&D spending and the fact that financial patents were used only infrequently until recently, these measures are unlikely to be satisfactory in this context. Most alternatives are also troubling. Consider, for instance, the listings of new securities compiled by Thomson Reuters' Securities Data Company (SDC), which maintains the leading database of corporate new issues. First, much of the innovation in financial services has taken place outside the realm of publicly traded securities, such as new Automatic Teller Machines and insurance products. Second, as Tufano (2003) points out, many of the "novel" securities identified in the SDC database are minor variants of existing securities, often promulgated by investment banks seeking to differentiate themselves from their peers.

Thus, saying much systematically about the variation in the rate of financial innovation across time and space is challenging. Lerner (2006) takes a first step toward addressing this gap by developing a measure of financial innovation based on news stories in the *Wall Street Journal.* The analysis finds that financial innovation is characterized by a disproportionate role of smaller firms. More specifically, a doubling in firm size is associated with less than a doubling in innovation generation. Moreover, firms that are less profitable in their respective sectors are disproportionately more innovative. These results are consistent with depictions by Silber (1975, 1983) that more marginal firms will contribute the bulk of the financial innovations. In addition, older, less leveraged firms located in regions with more financial innovation appear to be more innovative. Few patterns are seen over time, though this may reflect the fact that the analysis is confined to the years 1990 through 2002. Financial innovations seem to be disproportionately associated with US-based firms, though this may reflect the use of a US-based publication to identify the innovations.

A major focus of writings on financial innovations has been the attempt to catalog the inventions. Goetzmann and Rouwenhorst (2005) group the nineteen financial innovations they study into three categories, based on whether they (a) facilitate the transfer of value through time; (b) allow the ability to contract on future values; and (c) permit the negotiability of claims. There are almost as many schemes as authors, but many of these share the feature of looking through to the underlying functions performed by the innovations. Merton's (1992) and Crane et al.'s (1995) schemes are illustrative. In

particular, they identify six functions that innovations—and more generally, economies—perform:

1. Moving funds across time and space (e.g., savings accounts)
2. The pooling of funds (e.g., mutual funds)
3. Managing risk (e.g., insurance and many derivatives products)
4. Extracting information to support decision making (e.g., markets that provide price information, such as extracting default probabilities from bonds or credit default swaps)
5. Addressing moral hazard and asymmetric information problems (e.g., contracting by venture capital firms)
6. Facilitating the sale or purchase of goods and services through a payment system (e.g., cash, debit cards, credit cards)

Not surprisingly, no classification scheme is perfect, and more importantly, given their complexity of design and use, many innovations span multiple categories in this scheme and its alternatives.

In many respects, financial innovations resemble any other kind of invention. Among the points of commonality are:

- These innovations are not easy or cheap to develop and diffuse. While the cost of developing many security innovations is considerably smaller than for manufacturing or scientific innovations, investment banks frequently retain many highly compensated PhDs and MBAs and lawyers to design new products and services. Furthermore, innovators must frequently expend considerable resources developing distribution channels for their products.
- These innovations are risky. Tufano (1989) documents that the vast majority of security discoveries do not lead to more than a handful of subsequence issuances.
- Innovation is frequently linked closely with the competitive dynamics between incumbents and entrants, as suggested by the work just cited.
- Firms have struggled, at least until recently (and perhaps temporarily), to obtain intellectual property protection, akin to many emerging industries.

But in other respects, financial innovation is quite different. It is to these dissimilarities that we turn in the next section.

11.2 What Is Different—and Challenging—about Financial Innovation?

In general, economists' thinking about financial innovation has been shaped by their experience with innovation in manufacturing industries. Assessments of the nature and consequences of innovation in the service sector are rarer. Financial innovation illustrates the limitations of our understanding of nonmanufacturing innovation in particularly sharp relief.

At first glance, it might be unclear why financial innovation should differ from other types of new product development. In the canonical accounts of financial innovation (most importantly, Ross [1976] and Allen and Gale [1994]), innovation is driven by investor demand for a particular set of cash flows. Astute intermediaries recognize this demand and engineer securities with the desired characteristics. By splitting up or combining cash flows of existing securities, the intermediaries can create profits (at least in the short run) for themselves and increase social welfare. Described in this way, the financial innovation process seems little different from Apple's decision to introduce a tablet that combined features of a laptop and a cell phone, or Tropicana's introduction of orange juice with added calcium.

But these similarities between financial and other forms of innovation can be deceptive. In this section, we posit three sets of issues that make the study of financial innovation particularly challenging:

- The financial system is highly interconnected. As a result, a financial innovation is likely to generate a complex web of externalities, both positive and negative. Therefore, assessing the social consequences of financial innovation can be very challenging.
- Financial innovations are highly dynamic. As an innovation diffuses from pioneering adopters to more general users, these products frequently change in their underlying structure, the way that they are marketed, and how they are used. These transformations mean that the consequences of an innovation may change over time.
- While certainly many forms of innovation, such as pharmaceuticals, are subject to regulation, the regulation of new financial products and services is particularly complex and dynamic, and new financial reform has an uncertain impact on the pace and direction of financial innovation.

11.2.1 The Challenge Measuring Social Welfare

Since the pioneering work of Trajtenberg (1990), economists have understood that the benefits of innovation can be empirically quantified. These studies have focused on products whose features can be reduced to a relatively modest number of attributes and price. Each innovation can then be understood as offering a different combination of attributes. Often within the context of a discrete choice model, economists then use data on actual attributes, prices, and sales to estimate the underlying demand and utility functions of the representative consumer. The benefits from an innovation can then be quantified as the increase in social welfare associated with having the new set of choices compared to the ones available in the earlier period.[1]

1. Other important papers in the literature on the quantification of the economic benefits of innovations and new goods more generally include Berry, Levinsohn, and Pakes (1995); Bresnahan (1986); Hausman (1997); and Petrin (2002).

At least in theory, such a framework would allow one to assess whether innovations tend to significantly boost social welfare, or whether much of the spending on new product development is socially wasteful, motivated instead by the rent-seeking behavior and the desire to steal market share from competitors as Dasgupta and Stiglitz (1980) suggest.

To be sure, many innovations give rise to externalities that would resist this type of straightforward analysis. For instance, the widespread diffusion of cellular telephones and text messaging has led by many accounts to an increase in automobile accidents caused by distracted drivers—and has led to regulations to prohibit these uses of the innovations. Similarly, medical advances that prolong the lives of cancer patients may have the consequence of putting greater financial pressures on Social Security and Medicare as the longevity (and associated medical costs) of senior citizens increase.

The particular challenge associated with assessing the social impact of financial innovation lies in the fact that so many of its consequences are in the form of externalities. On the positive side of the ledger, many financial innovations address broad social needs. For example, venture capitalists provide a blend of money and expertise to help young firms succeed; credit cards extend credit but also simplify the process of purchasing goods and services. Moreover, in many instances, the decisions of early adopters have important consequences for others. For instance, as the pool of mutual funds has proliferated and funds have grown, upfront and annual fees associated with these products have generally fallen. As a result, the decision to partake of a financial innovation changes the attractiveness of the innovation for others.

But at the same time, in many instances these innovations have consequences to nontransacting parties that may be less desirable. To return to the subject of Krugman's quote earlier, the collapse in the markets for many of the complex securities based on mortgages contributed to a dramatic reduction in credit availability throughout the economy. Thus, these innovations indirectly may have led to numerous small businesses facing much higher interest rates or being unable to access credit at all, even though they had no involvement with the mortgage market. Even "well-meaning" innovations, such as process innovations that reduce the costs and effort of refinancing mortgages can lead to unintended consequences in the economy, a point emphasized by Khandani, Lo, and Merton (2009).

These detrimental effects are frequently referred to as "systemic risk." One immediate challenge is that systemic risk itself is a poorly defined notion. This confusion is captured by the following quote from Alan Greenspan (1995):

> It would be useful to central banks to be able to measure systemic risk accurately, but its very definition is still somewhat unsettled. It is generally agreed that systemic risk represents a propensity for some sort of

> significant financial system disruption, . . . [but] until we have a common theoretical paradigm for the causes of systemic stress, any consensus of how to measure systemic risk will be difficult to achieve. (7)

Schwarcz (2008), after compiling the various definitions that have been used in policy circles, suggests the following definition:

> [T]he risk that (i) an economic shock such as market or institutional failure triggers (through a panic or otherwise) either (X) the failure of a chain of markets or institutions or (Y) a chain of significant losses to financial institutions, (ii) resulting in increases in the cost of capital or decreases in its availability, often evidenced by substantial financial-market price volatility. (204)

Given the interconnected nature of the financial system, it would be surprising if the most widely adopted financial innovations did not contribute to systemic risk as defined earlier, as well as "systemic benefits." When the bulk of the social impact is through positive and negative externalities, it is unclear how one should seek to assess welfare consequences of innovations.

11.2.2 The Challenge of Dynamic Impacts

The word "innovation" is used by economists to indicate a change, and financial innovation must be understood as part of a process of change. Financial innovations—especially systemically important ones—demonstrate two related dynamic features: the innovation spiral and a change in the how products are used over time.

Merton (1992) coined the term "innovation spiral" to describe the process whereby one financial innovation begets the next. Sometimes this spiral has one successful innovation providing the raw material, or building blocks, for another. For example, the innovation of a futures market in a particular commodity can allow financial engineers to build specialized and more complex over-the-counter (OTC) products using dynamic trading strategies. An innovation need not be successful, however, to be part of the innovation spiral. Tufano (1995) and Mason et al. (1995) describe a sequence of financial innovations, most of which were unsuccessful, but nonetheless provided information that led to a subsequent wave of newer products. Persons and Warther (1997) formally model this spiral process. The innovation spiral is not unique to financial innovations; elsewhere one innovation can produce follow-on effects including lowering the barriers to subsequent innovation. For example, in electronics, semiconductor innovations have made possible a host of products ranging from personal computers to industrial applications to handheld devices. Similarly, the technology developed for unsuccessful pioneering personal digital assistants, such as Go's Pen Operating System and Apple's Newton, ultimately led to the success of the BlackBerry and

iPhone. Once one acknowledges the existence of an innovation spiral, one must recognize that actions that might discourage a certain innovation could have implications for the development of subsequent innovations.

Much of the research on innovation deals with the dynamics of the adoption process; that is, how a new product, process, or service is taken up, first by innovators, then early adopters, early majority, late majority, and laggards. This adoption process is typically characterized by an S-curve (or logistic function), which plots the number of adopters as a function of time. There is a substantial body of work on adoption rates, but Rogers (1962) is generally credited with codifying and advancing this literature. An S-curve adoption pattern suggests that, almost by definition, an innovation is unlikely to have economy-wide or systemic implications until it has been adopted fairly widely.

Most of the work on the diffusion of innovations deals with the characteristics of the population of potential adopters and of the actual adopters. Generally, more knowledgeable, sophisticated, and risk-taking individuals adopt innovations earlier. Generalizing across the landscape of innovations (not just financial breakthroughs), Rogers highlights five types of adopters:

- Innovators, the initial ones to take up the innovation. These are typically younger, better educated, and have higher social status than later adopters.
- Early adopters, who often serve as opinion leaders in shaping others' decision to adopt the product.
- The early majority, who adopt an innovation after a varying time lag.
- The late majority, who approach innovations with skepticism and wait until most of society has adopted the innovation.
- The laggards, who are the last to adopt an innovation, and tend to be older and of lower social status and with limited resources.

The mechanisms behind these broad patterns have attracted extensive research in subsequent years. For instance, Coleman, Katz, and Menzel (1966) highlighted how these patterns are driven by direct social ties between potential adopters; Burt (1987) has emphasized more diffuse connections with third parties; and Granovetter (1978) explained many of the differences because of differing psychological thresholds.

Not only do the identities of adopters change over time, but sometimes the way in which products are used can evolve. Early adopters may not only be more aware of the features—and limitations—of new products, but use them differently. For example, it is typically difficult to get an issuer and set of investors to be the first to issue and buy a new security. These innovation partners are often informally part of the product development process, consulted by the bankers who are trying to bring the product to market. They would typically be much more informed about the strengths and weaknesses

of a product than a late majority adopter, who might take a product's widespread usage to signal its lack of flaws. For example, in litigations involving "failed" financial products, it seems anecdotally that later adopters are more likely to sue, claiming that they were unaware of the potential flaws with the product, sometimes even claiming they never even read the security documents. (Consistent with these claims, Lerner [2010] shows that those who litigate patented financial innovations are disproportionately smaller, more marginal firms, with less financial resources. Similarly, studies of litigation of new securities offerings suggest that much of the litigation is initiated by relatively unsophisticated individual investors [Alexander 1991].)

This challenge is captured in a model of financial innovation by Gennaioli, Shleifer, and Vishny (2010). The paper argues that a financial innovation can address the demand for clients for a particular set of cash flows and thus be socially beneficial. But they suggest that the risks associated with these new products' cash flows may be systematically underestimated by these investors. In this case, they show, there may be excessive issuance of novel securities by financial institutions. Once the investors suddenly realize these risks, there will be an exodus back to traditional, safer products. In this way, financial innovation can add to the fragility of the overall financial system.

Given the importance of externalities in financial innovation, the changing awareness of adopters may have broad implications. While some late adopters of smartphones might use only a portion of their newest gadget's technology, the social costs of their ignorance might be minimal. However, a late-adopting, unsophisticated investor or borrower using a new complex instrument might find himself with an exposure or liability that sophisticated earlier adopters fully appreciated. Understanding the dynamics of adoption provides some insight into the potential for financial innovations to give rise to externalities and systemic risks. We may need to understand especially the processes whereby innovations become widely accepted—by whom and for what purpose—to understand systemic risks.

An appreciation of the innovation spiral and the diffusion processes for financial innovations highlights the challenges facing much traditional empirical work on financial innovation. First, to understand social welfare, it is problematic to study a single financial innovation out of context, as any one innovation—whether successful or not—will tend to influence the path of future innovations. Second, most empirical studies, but especially structured interventions like randomized control trials, document the experiences of early adopters, and the way in which the product is used by these sophisticated adopters. However, the experiences of later adopters—and the ways in which innovations are adapted for multiple uses as they are diffused more broadly—may give greater clues as to the social welfare implications of financial innovations. Finally, the long time spans over which financial innovations diffuse and the innovation spiral that an initial innovation often

engenders suggest that the researcher needs an extended time frame, or an historical approach to studying financial innovations.

11.2.3 The Interaction between Regulation and Innovation

The relationship between financial innovation and regulation is complex. There has been much written about regulation (and taxes) as being important stimuli for financial innovation. Miller (1986) expounds on this link at some length, and it is fairly easy to find financial products whose origins can be tied, at least in part, to regulations or taxes. For example, in the nineteenth century, the innovation of low-par stock was an outgrowth of state securities taxes (Tufano 1995). In the 1980s, the growth—and preferred stock form—of various adjustable rate products was stimulated by intercorporate dividend deduction rules. More recently, bank capital rules have encouraged the creation or adaptation of a variety of capital securities.

Not only does regulation give rise to certain innovations, but then regulators need to "catch up" with the products, in a cat-and-mouse process that Kane (1977) labels the regulatory dialectic. Innovators look for opportunities that exploit regulatory gaps, regulators impose new regulations, and each new regulation gives rise to new opportunities for more innovation. In this back and forth, the regulatory system can be at a disadvantage for a variety of reasons. First, many regulatory bodies have mandates that are defined by product or by institution, rather than by function. For example, consider just a few of the products that deliver equity-index exposure: baskets of stocks, index funds, exchange-traded funds (ETFs), futures contracts, index-linked annuities, indexed-linked certificates of deposit, and various structured notes. Suppose that one wanted to regulate equity exposures broadly. One would have to coordinate activities between the Securities and Exchange Commission (SEC), Commodity Futures Trading Commission, banking regulators, and state insurance regulators just for a start. Without broad mandates or functional jurisdictions, opportunities for regulatory arbitrage through innovation will occur. Second, even a well-staffed, reasonably well-paid, and highly talented regulatory agency is up against a world of potential entrepreneurs and innovators. Inevitably, regulation will tend to react to innovations, typically with a lag. From the perspective of systemic risk, this responsive approach may be appropriate, as innovations early in their S-curve adoptions are unlikely to pose economy-wide risks, and are probably bought and sold by the more sophisticated set of adopters.

11.3 A Counterfactual Approach to Studying the Social Welfare Implications of Systemic Financial Innovations

In the wake of the events of the past few years, there have been numerous calls to limit or even ban financial innovation. For example, in a 2009 *Business Week* article entitled "Financial Innovation Under Fire," Coy notes:

> [S]ome economists go further and argue that any financial innovation is guilty until proven innocent. Former International Monetary Fund chief economist Simon Johnson and James Kwak, authors of the popular Baseline Scenario blog, wrote in the summer issue of the journal *Democracy* that innovation often generates unproductive or even destructive transactions. "The presumption should be that innovation in financial products is costly . . . and should have to justify itself against those costs," they wrote.

In April 2009, Fed Chairman Bernanke, while defending financial innovation, noted its precarious state in public debates:

> The concept of financial innovation, it seems, has fallen on hard times. Subprime mortgage loans, credit default swaps, structured investment vehicles, and other more-recently developed financial products have become emblematic of our present financial crisis. Indeed, innovation, once held up as the solution, is now more often than not perceived as the problem.[2]

An interesting sign of the mood is the Security and Exchange Commission's creation of the first new division in thirty years, a Division of Risk, Strategy, and Financial Innovation, implicitly joining "financial innovation" and "risk."[3]

Against this chorus of anti-innovation rhetoric, it is important to carry out rigorous scholarly research to establish the social costs and benefits of financial innovation. Given the large number of financial innovations, it is important to come up with a research strategy that can address the important policy issues of the day. These debates seem to be of various forms: financial innovations' potential to give rise to systemic risks; financial innovations' potential to harm consumers; and "wasteful" use of private resources by financial innovators in rent-seeking behavior. Against this potential list of costs we must analyze innovation's benefits, both direct and indirect.

In this chapter, we focus on the systemic risks and benefits imposed by financial innovations. If an innovation is to have system-wide implications, it must be broadly adopted. This research strategy permits us to focus on widely adopted innovations, rather than narrowly adopted ones or others that were never or barely adopted by users. To study potentially wasteful rent-seeking or some aspects of consumer damage, one would need to include these latter innovations, but they strike us as not being the likely locus of systemic risks or benefits.

How do we define a "systemically important" or "broadly adopted" financial innovation? We use top-down data on the economy to identify these innovations. For example, if one studies the balance sheet of the US

2. http://www.federalreserve.gov/newsevents/speech/bernanke20090417a.htm.
3. http://www.sec.gov/news/press/2009/2009-199.htm.

household over the past sixty years, a number of striking trends emerge, in particular the economic importance of money market mutual funds, mutual funds more generally, and retirement plans. Clearly, these are innovations that were adopted widely in the postwar period.

Then, for a subset of these innovations, we detail the elements of their welfare implications. Using a technique of historians, we not only detail actual outcomes, but also discuss counterfactual histories: What would the economy have been like had this innovation not been invented or popularized? While this method is inherently judgmental, it frames a discussion or debate that attempts to tease out not only the direct costs and benefits, but also the externalities—both positive and negative—associated with each innovation. In the following section, we provide the logic of our selection of these case studies of systemic innovations and a brief primer on the methods of counterfactual history.

11.3.1 Methodology: Criteria for Selection of Case Studies and the Counterfactual Approach

We need a disciplined way to scan the economy to select our case studies. To do this, we consider the major changes in the way that financial functions are delivered to each of the major nongovernmental sectors in the economy. The sectors are (a) households, (b) nonfinancial corporations, (c) financial firms, and (d) public entities. As noted earlier, the functions include six activities: (a) pooling, (b) payments, (c) moving funds across time and space, (d) managing risk, (e) resolving information asymmetries, and (f) extracting information from markets. Our primary frame of reference for our exercise is the United States in the postwar period. (See table 11.1.)

We focus on three case studies: venture capital and private equity, mutual funds and exchange-traded funds, and securitization. This allows us to focus on three of the functions and three of the sectors. In addition, the selection

Table 11.1 **Typology of case studies**

	Households	Nonfinancial firms	Financial firms
Pooling	Mutual funds and exchange-traded funds	Venture capital and private equity	Securitization
Moving money across time and space			
Payments	Card products		
Managing risk	Retirement accounts	Derivatives	
Resolving information asymmetries		Venture capital and private equity	
Extracting information from markets			Derivatives

of these case studies suggests the strengths and weaknesses of a counterfactual approach, as the first two case studies are more amenable to this approach than is the third.

Most economic analyses attempt to measure outcomes of interventions relative to some alternative. We typically exploit cross-sectional and time-series variation—often with large-sample data—to tease out the relative effects of some intervention or innovation. We use control/treatment approaches or randomized control trials to minimize noise and identify phenomena. These methods work well when we have large samples or natural experiments.

Unfortunately, systemic innovations do not lend themselves well to these methods. Because they are systemic, it is difficult to find adequate "control" states. Pre- and posttests are problematic because innovations are adopted over long periods of time. These tests are also difficult because early adopters may not be representative of late adopters—and the way the product is used may vary over time. Randomized control trials do not tend to capture the systemic effects when products are broadly adopted. This is not to say that econometric methods are not useful in understanding financial impacts, but they have meaningful limits, and that complementary approaches can be valuable.

A meaningful alternative is to adopt a historical approach to understanding systemic innovations that span years or decades. There are a number of excellent studies of financial history and economic history, with a few that specialize in financial innovation. Goetzmann and Rouwenhorst's edited volume (2005) contains a set of nineteen essays on particular innovations, including the invention of interest in Sumerian Loans, the creation of Roman Shares, the origins of paper money in China, Dutch perpetuities, modern European annuities, inflation indexed bonds in early America, and the first Eurobonds in the nineteenth century. Davis' (1994) book, *A History of Money,* spans 3000 BC to the twentieth century. With this wide sweep, it covers a number of innovations in its scope. Beyond documenting the various forms of instruments created over time, Davis traces the evolution of financial institutions, for example, the working-class financial institutions of friendly societies, cooperatives, and building societies in Europe in the nineteenth century. Kindleberger (1984), Cameron et al. (1967), and Cameron (1972) are other fine examples of centuries-long, multicountry historical studies of the evolution of financial systems. Cameron (1972), studying banking in the early stages of industrialization, notes that financial innovation is "necessary for the realization of (technical innovation)" and in combination can achieve "the pooling of risks and economies of scale in finance as well as in manufacture." However, the role of some innovations in creating financial crises is highlighted in Kindleberger (1984, 270):

> Time and again in these pages it has been stressed that when the macroeconomic system is constrained by a tight supply of money, it creates more, at least for a time. Shortage of gold and silver has led to substitution of copper, pepper, salt, that is, to more primitive commodity monies, or to more sophisticated substitutes such as various forms of paper (and plastic): bank money, bank notes, bills of exchange, especially chains of bills of exchange, bank deposits, open-book credits, credit cards, certificates of deposit, Euro-currencies and so on.

In his analysis, these expansions of the money supply sometimes lead to overextension, distress, speculation, and at times panics and crashes.

While a historical approach has the advantage of intensely studying phenomena, it may not address relative performance implications, unless one adopts a comparative historical approach, for example comparing one period or country to another. Unfortunately, however, these comparisons often suffer from a great deal of endogeneity that makes interpretation difficult. For example, while we could compare economies with considerable financial innovation to those less innovative, it is highly unlikely that these comparisons would be ceteris paribus. Financial innovation, and certainly financial development more generally, is not unrelated to economic development, so these types of comparisons are problematic. For example, the adoption of mutual funds is related to a number of metrics of financial development and to the state of legal institutions (Khorana, Servaes, and Tufano 2005).

Scholars have used various approaches to deal with these inevitable issues by studying counterfactual or virtual histories. In essence, a counterfactual approach requires the analyst to posit "what would have happened if . . . had happened (or not happened)." This method has been used—and debated—by historians, economic historians, political scientists, sociologists, and philosophers. For reviews, see, for example, Ferguson (1997, chapter 1), Cowan and Foray (2002), Sylvan and Majeski (1998), Bunzl (2004), and Tetlock and Lebow (2001). While dismissed by some as a "mere parlour game" (Carr 1987), and referred to in scatological terms by others (see ref. in Ferguson 1999), the method has been used extensively.

Counterfactual reasoning seems to have been adopted most extensively in international relations and politics. For example, Ferguson (2000) is a collection of papers that study a variety of counterfactuals: "What if Charles I had avoided the Civil War? What if there had been no American Revolution? What if Germany had invaded Britain in May 1940?" Perhaps the most well-known example of the method (among economists) is Fogel's groundbreaking 1964 book *Railroad and American Economic Growth: Essays in Econometric History.* The book studies the impact of the railroads by trying to assess how the economy would have developed in their absence, as noted in Fogel's preface:

> The pages that follow contain a critical evaluation of the proposition that railroads were indispensible to American economic growth during the nineteenth century . . . [I] estimate the amount by which production possibilities of the nation would have been reduced if agricultural commodities could not have been shipped by railroads. (vii)

Fogel combines counterfactual reasoning with empirical estimates of development. He compares observed gross domestic product (GDP) increases with three counterfactuals: no railroads at all, an extension of internal navigation (canals), and the improvement of country roads. In essence, Fogel's work demonstrates the core elements of counterfactual analysis. First, he identifies an important topic where the facts do not permit the type of replicability that underscores much of scientific inquiry. Second, he identifies a set of alternative paths of history. Subsequent work has differentiated between "miracle worlds" and "plausible worlds," based on the likelihood of the alternative to have played out. In Fogel's examples, he did not posit air travel (a miracle, to be sure, in the nineteenth century), but rather quite plausible alternative transportation developments. Finally, Fogel rigorously attempts to analyze the economic implications of these alternatives using a well-defined metric.

Fogel was awarded the Nobel Prize in 1993 for having given birth to cliometrics, or new economic history. In its award, the Nobel committee made clear the importance of Fogel's pioneering of the counterfactual approach:

> Robert W. Fogel's scientific breakthrough was his book (1964) on the role of the railways in the American economy. Joseph Schumpeter and Walt W. Rostow had earlier, with general agreement, asserted that modern economic growth was due to certain important discoveries having played a vital role in development. Fogel tested this hypothesis with extraordinary exactitude, and rejected it. The sum of many specific technical changes, rather than a few great innovations, determined the economic development. We find it intuitively plausible that the great transport systems play a decisive role in development. Fogel constructed a hypothetical alternative, a so called counterfactual historiography; that is he compared the actual course of events with the hypothetical to allow a judgment of the importance of the railways. He found that they were not absolutely necessary in explaining economic development and that their effect on the growth of GNP was less than three per cent. Few books on the subject of economic history have made such an impression as Fogel's. His use of counterfactual arguments and cost-benefit analysis made him an innovator of economic historical methodology.[4]

Fogel's use of counterfactual arguments flew in the face of "common sense" and demonstrated the power of this method. Just as the innovation

4. Taken from http://nobelprize.org/nobel_prizes/economics/laureates/1993/press.html.

of railroads gave rise to some changes in the economy, but perhaps not as much as originally thought, financial innovations are long-lived phenomena that can substantially alter the economic landscape—but perhaps not as much as originally thought. We use counterfactual reasoning and methods to structure our exploration of the social consequences of these innovations. Our goal is not to definitively determine whether recent financial innovations were or were not socially valuable. Rather, we lay out an approach to make progress on this problem, and challenge others to use this approach systemically. Our three case studies provide some factual background on the innovations, then lay out—for debate—counterfactual histories and thoughts about the implications of each. A full analysis, à la Fogel, of the counterfactual history of each innovation would be beyond the scope of this chapter.

While we adopt this method as a complement to existing historical, experimental, and econometric methods, we acknowledge the many strong and legitimate criticisms of it, and of Fogel's work. There are important, and quite specific, critiques of the calculations employed by Fogel by Nerlove (1966), McClelland (1968), and David (1969), among others. In particular, David's (1969) critique focuses on problems caused by inadequate consideration of complementaries (e.g., passenger transportation or changes in inventories to reflect different transport speeds), path-dependent adoption processes (e.g., learning effects or returns to scale), finding the correct scaled metric for measuring social benefits, or the challenge of taking a partial (vs. general) equilibrium approach. This latter general point lies at the heart of the criticism—and appeal—of counterfactual analysis, as summarized by Goldin (1995, 195):

> The notion of a counterfactual was hard for many historians to swallow. It involved the hypothetical removal of the largest enterprise at the time, the first big business in America, one of the most productive sectors, and some of the wealthiest Americans, to mention just a few parts of the mental experiment. But, noted Fogel, those who were making claims about the indispensability of the railroad were implicitly invoking precisely this experiment. He was merely making the claim explicit and subjecting it to hard evidence.

In some sense, the instances where counterfactual analysis is most problematic—where an innovation is intrinsically bound up with the rest of the economy for decades—are precisely those instances where it is useful to complement existing research with this more provocative method. A full quantification of the social welfare consequences of removing railroads (or mutual funds or venture capital) from the economy is daunting, but the audacity of asking the question forces our profession to try to address the many issues that bedeviled Fogel and his critics. If this method provokes debate and criticism (and additional work) we will have achieved some of

our objectives, moving the discussion beyond simplistic notions about financial innovation.

11.3.2 Venture Capital and Private Equity

A Brief History

Long before the creation of the venture capital and private equity industry, fast-growing firms were able to raise financing. Banks provided debt in the form of loans, and for more long-run, riskier investments, wealthy individuals provided equity.

But by the time of the Great Depression of the 1930s, there was a widespread perception that the existing ways of financing fast-growing young firms were inadequate. Not only were many promising companies going unfunded, but investors with high net worth frequently did not have the time or skills to work with young firms to address glaring management deficiencies. Nor were the alternatives set up by the Roosevelt administration during the New Deal—such as the Reconstruction Finance Corporation—seen as satisfactory. The rigidity of the loan evaluation criteria, the extensive red-tape associated with the award process, and the fears of political interference and regulations all suggested a need for an alternative.

The first formal venture capital firm was established with both private and social returns in mind. American Research and Development (ARD) grew out of the concerns that the United States, having been pushed out of the depression by the stimulus of wartime spending by the federal government, would soon revert to economic lethargy when the war ended. In October 1945, Ralph Flanders, then head of the Federal Reserve Bank of Boston, argued that if this danger was to be addressed, a new enterprise was needed, with the goal of financing new businesses. He argued that the enterprise would not only need to be far more systematic in "selecting the most attractive possibilities and spreading the risk" than most individual investors had been, but would need to tap into the nation's "great accumulation of fiduciary funds" (i.e., pension funds and other institutional capital) if it was to be successful in the long term.

The ARD was formed a year later to try to realize this vision. Flanders recruited a number of civic and business leaders to join in the effort, including MIT president Karl Compton. But the day-to-day management of the fund fell on the shoulders of Harvard Business School professor Georges F. Doriot. The ARD in its communications emphasized that its goal was to fund and aid new companies in order to generate "an increased standard of living for the American people."

Flanders, Doriot, and their contemporaries realized that the financing of young, growing, and restructuring companies was a risky business. Information problems made it difficult to assess these companies and permitted

opportunistic behavior by entrepreneurs after the financing was received. These risks had deterred investors from providing capital to these firms.

To illustrate such problems, if the firm raises equity from outside investors, the manager has an incentive to engage in wasteful expenditures (e.g., lavish offices) because he may benefit disproportionately from these but does not bear their entire cost. Similarly, if the firm raises debt, the manager may increase risk to undesirable levels. Because providers of capital recognize these problems, outside investors demand a higher rate of return than would be the case if the funds were internally generated. Additional problems may appear in the types of more mature companies in which private equity firms invest. For instance, entrepreneurs might invest in strategies or projects that have high personal returns but low expected monetary payoffs to shareholders.

Even if the manager wants to maximize firm value, information gaps may make raising external capital more expensive or even preclude it entirely. Equity offerings of companies may be associated with a "lemons" problem: that is, if the manager is better informed about the company's investment opportunities and acts in the interest of current shareholders, then he will only issue new shares when the company's stock is overvalued. Indeed, numerous studies have documented that stock prices decline upon the announcement of equity issues, largely because of the negative signal sent to the market. This "lemons" problem leads investors to be less willing to invest at attractive valuations in young or restructuring companies, or even to invest at all.

The ARD established an approach to addressing these problems that venture capital and private equity groups have followed ever since. First, by intensively scrutinizing companies before providing capital, and only funding a small fraction of those seeking funds, they could alleviate some of the information gaps and reduce capital constraints. Second, they employed a variety of tools that allowed them to monitor and control firms after the transactions. These included the use of convertible securities with powerful control rights, the syndication and staging of investments, the provision of oversight through formal board seats and information rights, the incentivization of management through extensive equity holdings, and informal coaching of management. Finally, there was a real effort to certify the funded entrepreneurs as being different from their peers, which facilitated their ability to enter into alliances, get access to investment bankers, and so forth. The tools that venture capital and private equity investors use in this difficult environment enable companies ultimately to receive the financing that they cannot raise from other sources.

The activity in the private equity industry increased dramatically in the late 1970s and early 1980s. Industry observers attributed much of the shift to the US Department of Labor's clarification of the Employee Retirement

Income Security Act's "prudent man" rule in 1979. Prior to this year, the legislation limited the ability of pension funds to invest substantial amounts of money into venture capital or other high-risk asset classes. The Department of Labor's clarification of the rule explicitly allowed pension managers to invest in high-risk assets, including private equity. Numerous specialized funds—concentrating in areas such as leveraged buyouts, mezzanine transactions, and such hybrids as venture leasing—sprang up during these years.

The subsequent years saw both very good and trying times for private equity investors. On the one hand, the 1980s saw venture capitalists back many of the most successful high-technology companies, including Cisco Systems, Genentech, Microsoft, and Sun Microsystems. Numerous successful buyouts—such as Avis, Beatrice, Dr. Pepper, Gibson Greetings, and McCall Pattern—garnered considerable public attention during that period. At the same time, commitments to the private equity industry during this decade were very uneven. The annual flow of money into venture capital funds increased by a factor of ten during the first half of the 1980s, but steadily declined from 1987 through 1991. Buyouts underwent an even more dramatic rise through the 1980s, followed by a precipitous fall at the end of the decade.

Much of this pattern was driven by the changing fortunes of private equity investments. Returns on venture capital funds had declined sharply in the mid-1980s after being exceedingly attractive in the 1970s. This fall was apparently triggered by overinvestment in a few industries, such as computer hardware, and the entry of many inexperienced venture capitalists. Buyout returns underwent a similar decline in the late 1980s, due in large part to the increased competition between groups for transactions. Kaplan and Stein (1993) documented that of the sixty-six largest buyouts completed during the market peak (between 1986 and 1988), 38 percent experienced financial distress, which they define as default or an actual or attempted restructuring of debt obligations due to difficulties in making payments, and 27 percent actually did default on debt repayments, often in conjunction with a Chapter 11 filing. Kaplan and Schoar (2005) and other papers provide indirect supporting evidence showing that the performance of both venture and private equity funds is negatively correlated with inflows into these funds. Funds raised during periods of high capital inflows—which are typically associated with market peaks—perform far worse than their peers.

The 1990s and 2000s saw these patterns repeated on an unprecedented scale. The second half of the 1990s saw dramatic growth and excellent returns in venture capital investments; the 2000s saw tremendous growth of private equity funds. This recovery was triggered by several factors. The exit of many inexperienced investors after the earlier collapse ensured that the remaining groups faced less competition for transactions. The healthy market for initial public offerings during much of the 1990s meant that it

was easier for venture funds to exit transactions, leading to high returns. Meanwhile, the extent of technological innovation—particularly in information technology-related industries—created extraordinary opportunities for venture capitalists. The mid-2000s saw unprecedented availability of debt on favorable terms, which enabled buyout groups to highly leverage firms and make high returns likely. New capital commitments to both venture and buyout funds rose in response to these changing circumstances, increasing to record levels. Once the enabling condition deteriorated, the level of fundraising and investment dropped sharply. Funds were left with large numbers of transactions that could not be exited, and investors faced the certainty of a sharp drop in returns.

The Broader Social Impact: Venture Capital

Clearly, the innovations of venture capital (VC) and private equity funds exert a major impact on the fates of individual companies. But does all this fundraising and investing influence the overall economic landscape as well? We will look at evidence regarding venture capital first, and then private equity funds. One caveat should be noted upfront: all these studies examine the last three decades, with a particular emphasis on the experience of the United States, a time and place that are certainly not representative of the entirety of economic history. There is little choice, however, given the relative youth of these intermediaries and the lack of data on earlier, pioneering funds.

To assess this question, we can look at studies of the experience of the market with the most developed and seasoned venture capital industry, the United States. Despite the fact that venture activity is particularly well-developed in this nation, the reader might be skeptical as to whether this activity would noticeably impact innovation: for most of past three decades, investments made by the entire venture capital sector totaled less than the research-and-development and capital-expenditure budgets of large, individual companies such as IBM, General Motors, or Merck.

One way to explore this question is to examine the impact of venture investing on wealth, jobs, and other financial measures across a variety of industries. Though it would be useful to track the fate of every venture capital-financed company and find out where the innovation or technology ended up, in reality only those companies that have gone public can be tracked. Consistent information on venture-backed firms that were acquired or went out of business simply does not exist. Moreover, investments in companies that eventually go public yield much higher returns than support given to firms that get acquired or remain privately held.

These firms have had an unmistakable effect on the US economy. In late 2008, 895 firms were publicly traded on US markets after receiving their private financing from venture capitalists (this does not include the firms that went public, but were subsequently acquired or delisted). One way to

assess the overall impact of the venture capital industry is to look at the economic "weight" of venture-backed companies in the context of the larger economy.[5] By late 2008, venture-backed firms that had gone public made up over 13 percent of the total number of public firms in existence in the United States at that time. And of the total market value of public firms ($28 trillion), venture-backed companies came in at $2.4 trillion—8.4 percent.

Venture-funded firms also made up over 4 percent (nearly $1 trillion dollars) of total sales ($22 trillion) of all US public firms at the time. Contrary to the general perception that venture-supported companies are not profitable, operating income margins for these companies hit an average of 6.8 percent—close to the average public-company profit margin of 7.1 percent. Finally, those public firms supported by venture funding employed 6 percent of the total public-company workforce—most of these jobs were high-salaried, skilled positions in the technology sector. Clearly, venture investing fuels a substantial portion of the US economy.

This impact is quite modest in industries dominated by mature companies such as the manufacturing industries. But contrast those industries with highly innovative ones, and the picture looks completely different. For example, companies in the computer software and hardware industry that received venture backing during their gestation as private firms represented more than 75 percent of the software industry's value. Venture-financed firms also play a central role in the biotechnology, computer services, and semiconductor industries. In recent years, the scope of venture groups' activity has been expanding rapidly in the critical energy and environmental field, though the impact of these investments remains to be seen. Presumably, these are industries where the externalities generated by new activity are the greatest.

It might be thought that it would not be difficult to address the question of the impact of venture capital on innovation in a more rigorous manner. For instance, one could seek to explain across industries and time whether, controlling for R&D spending, venture capital funding has an impact on various measures of innovation. But even a simple model of the relationship between venture capital, R&D, and innovation suggests that this approach is likely to give misleading estimates.

This is because both venture funding and innovation could be positively related to a third unobserved factor, the arrival of technological opportunities. Thus, there could be more innovation at times that there was more venture capital, not because the venture capital caused the innovation, but rather because the venture capitalists reacted to some fundamental technological shock that was sure to lead to more innovation. To date, only a handful of papers have attempted to address these challenging issues.

5. This analysis is based on the authors' tabulation of unpublished data from SDC Venture Economics, with supplemental information from Compustat and the Center for Research into Securities Prices (CRSP) databases.

The first of these papers, by Hellmann and Puri (2002), examines a sample of 170 recently formed firms in Silicon Valley, including both venture-backed and nonventure firms. Using questionnaire responses, they find evidence that venture capital financing is related to product market strategies and outcomes of startups. They find that firms that are pursuing what they term an innovator strategy (a classification based on the content analysis of survey responses) are significantly more likely and faster to obtain venture capital. The presence of a venture capitalist is also associated with a significant reduction in the time taken to bring a product to market, especially for innovators (probably because these firms can focus more on innovating and less on raising money). Furthermore, firms are more likely to list obtaining venture capital as a significant milestone in the life cycle of the company as compared to other financing events. There seems to be a link between this form of financial innovation and more traditional product innovation.

The results suggest significant interrelations between investor type and product market dimensions, and a role of venture capital in encouraging innovative companies. But this does not definitively answer the question of whether venture capitalists cause innovation. For instance, we might observe personal injury lawyers at accident sites, handing out business cards in the hopes of drumming up clients. But just because the lawyer is at the scene of the car crash does not mean that he caused the crash. In a similar vein, the possibility remains that more innovative firms choose to finance themselves with venture capital, rather than venture capital causing firms to be more innovative.

Kortum and Lerner (2000) visit the same question. Here, the study looks at the aggregate level: did the participation of venture capitalists in any given industry over the past few decades lead to more or less innovation? It might be thought that such an analysis would have the same problem as the aforementioned personal injury lawyer story. Put another way, even if we see an increase in venture funding and a boost in innovation, how can we be sure that one caused the other?

The authors address these concerns about causality by looking back over the industry's history. In particular, as we discussed earlier, a major discontinuity in the recent history of the venture capital industry was the US Department of Labor's clarification of the Employee Retirement Income Security Act in the late 1970s, a policy shift that freed pensions to invest in venture capital. This shift led to a sharp increase in the funds committed to venture capital. This type of external change should allow one to figure out what the impact of venture capital was, because it is unlikely to be related to how many or how few entrepreneurial opportunities there were to be funded.

Even after addressing these causality concerns, the results suggest that venture funding does have a strong positive impact on innovation. The estimated coefficients vary according to the techniques employed, but on average a dollar of venture capital appears to be three to four times more potent in stimulating patenting than a dollar of traditional corporate R&D.

The estimates therefore suggest that venture capital, even though it averaged less than 3 percent of corporate R&D in the United States from 1983 to 1992, is responsible for a much greater share—perhaps 10 percent—of US industrial innovations in this decade.

A natural worry with the aforementioned analysis is that it looks at the relationship between venture capital and patenting, not venture capital and innovation. One possible explanation is that such funding leads entrepreneurs to protect their intellectual property with patents rather than other mechanisms such as trade secrets. For instance, it may be that the entrepreneurs can fool their venture investors by applying for large number of patents, even if the contributions of many of them are very modest. If this is true, it might be inferred that the patents of venture-backed firms would be lower quality than nonventure-backed patent filings.

How could this question of patent quality be investigated? One possibility is to check the number of patents that cite a particular patent.[6] Higher-quality patents, it has been shown, are cited by other innovators more often than lower-quality ones. Similarly, if venture-backed patents are lower quality, then companies receiving venture funding would be less likely to initiate patent-infringement litigation. (It makes no sense to pay money to engage in the costly process of patent litigation to defend low-quality patents.)

So, what happens when patent quality is measured with these criteria? As it happens, the patents of venture-backed firms are more frequently cited by other patents and are more aggressively litigated—thus it can be concluded that they are high quality. Furthermore, the venture-backed firms more frequently litigate trade secrets, suggesting that they are not simply patenting frantically in lieu of relying on trade-secret protection. These findings reinforce the notion that venture-supported firms are simply more innovative than their nonventure-supported counterparts.

Mollica and Zingales (2007), by way of contrast, focus on regional patterns: as a regional unit, they use the 179 Bureau of Economic Analysis economic areas, which are composed by counties surrounding metropolitan areas. They exploit the regional, cross-industry, and time-series variability of venture investments in the United States to study the impact of venture capital activity on innovation and the creation of new businesses. Again, they grapple with causality issues by using an instrumental variable: as an instrument for the size of VC investments, they use the size of a state pension fund's assets. The idea is that state pension funds are subject to political pressure to invest some of their funds in new businesses in the states. Hence, the size of the state pension fund triggers a shift in the local supply of VC investment, which should help identify the effect of VC on patents.

6. Patent applicants and examiners at the patent office include references to other relevant patents. These serve a legal role similar to that of property markers at the edge of a land holding.

Even with these controls, they find that VC investments have a significant positive effect both on the production of patents and on the creation of new businesses. A one standard deviation increase in VC investment per capita generates an increase in the number of patents of between 4 and 15 percent. An increase of 10 percent in the volume of VC investment increases the total number of new business by 2.5 percent.

The Broader Social Impact: Private Equity

Turning to private equity (PE), in the past decade the growth of this industry has triggered anxiety about the impact of buyouts in markets as diverse as China, Germany, South Korea, the United Kingdom, and the United States. This anxiety is not unreasonable. While the leveraged buyout transactions of the 1980s were scrutinized in a number of important academic analyses, these studies had two important limitations. First, the bulk of the older research focused on a relatively small number of transactions involving previously publicly traded firms based in the United States. But these represent only a very modest fraction of all buyouts. The second limitation of the older research relates to the fact that the industry has grown and evolved tremendously since the 1980s.

A variety of recent research has sought to assess the consequences of private equity investments over a more comprehensive sample. Each study has looked at a particular consequence of the investment process.

First, Strömberg (2008) examined the nature and outcome of the 21,397 private equity transactions worldwide between 1970 and 2007. In the most straightforward possible outcome, the author simply sought to understand the consequences of these transactions. The key findings were:

- Of the exited buyout transactions, only 6 percent end in bankruptcy or financial restructuring. This translates into an annual rate of bankruptcy or major financial distress of 1.2 percent per year. This rate is a lower default rate than for US corporate bond issuers, which has averaged 1.6 percent per year.
- Holding periods for private equity investments have increased, rather than decreased, over the years. Fifty-eight percent of the private equity funds' investments are exited more than five years after the initial transaction. So-called "quick flips" (i.e., exits within two years of investment by private equity fund) account for 12 percent of deals and have also decreased in the last few years.

This study, of course, only examines one small fraction of what would be the consequences of these transactions. It cannot answer the question of whether the bulk of the firms would be worse or better off because of these transactions.

Bloom, Sadun, and Van Reenen (2009) examine management practices across 4,000 PE-owned and other firms in a sample of medium-sized

manufacturing firms in Asia, Europe, and the United States using a unique double-blind management survey to score firms across eighteen dimensions. The main goal of the study is to determine whether private equity ownership, relative to other ownership firms, is a way to achieve improved management practices within firms through the introduction of new managers and better management practices.

They find that private equity-owned firms are, on average, the best-managed ownership group. The PE-owned firms are significantly better managed across a wide range of management practices than government, family, and privately-owned firms. This is true even controlling for a range of other firm characteristics such as country, industry, size, and employee skills. The PE-owned firms are particularly strong at operations management practices, such as the adoption of modern lean manufacturing practices, using continuous improvements, and a comprehensive performance documentation process. But because the survey is only a cross-sectional one, they cannot determine whether the private equity groups turned these firms into better managed ones, or simply purchased firms that were better managed in the first place.

Lerner, Sorenson, and Stromberg (2008) examine long-run investments by firms. This work was motivated by the lively debate about the impact of private equity investors on the time horizons of the companies in their portfolios. The private status, according to some, enables managers to proceed with challenging restructurings without the pressure of catering to the market's demands for steadily growing quarterly profits, which can lead to firms focusing on short-run investments. Others have questioned whether private equity-backed firms take a longer-run perspective than their public peers, pointing to practices such as special dividends to equity investors.

In this study, one form of long-run investment was examined: investments in innovation. Innovation offers an attractive testing ground for the issues delineated earlier due to various factors. These factors include the long-run nature of R&D expenditures, their importance to the ultimate health of firms, and the extensive body of work in the economics literature that has documented that the characteristics of patents can be used to assess the nature of both publicly and privately held firms' technological innovations.

The key finding is that patenting levels before and after buyouts are largely unchanged. But firms that undergo a buyout pursue more economically important innovations, as measured by patent citations, in the years after private equity investments. In a baseline analysis, the increase in the key proxy for economic importance is 25 percent. This results from firms focusing on and improving their research in their technologies, where the firms have historically focused.

In a pair of studies, Davis et al. (2008, 2009) have examined the impact of these investments on employment and productivity. The former question

has aroused considerable controversy. Critics have claimed huge job losses, while private equity associations and other groups have released several recent studies that claim positive effects of private equity on employment. While efforts to bring data to the issue are highly welcome, many of the prior studies have significant limitations, such as the reliance on surveys with incomplete responses, an inability to control for employment changes in comparable firms, the failure to distinguish cleanly between employment changes at firms backed by venture capital and firms backed by other forms of private equity, and an inability to determine in which nation jobs are being created and destroyed.

The authors constructed and analyzed a data set in order to overcome these limitations and, at the same time, encompass a much larger set of employers and private equity transactions from 1980 to 2005. The study utilizes the Longitudinal Business Database (LBD) at the US Bureau of the Census to follow employment at virtually all private equity-backed companies, before and after private equity transactions.

Among the key results were:

- Employment grows more slowly at establishments that are bought out than at the control group in the year of the private equity transaction and in the two preceding years. The average cumulative employment difference in the two years before the transaction is about 4 percent in favor of controls.
- Employment declines more rapidly in bought-out establishments than in control establishments in the wake of private equity transactions. The average cumulative two-year employment difference is 7 percent in favor of controls. In the fourth and fifth years after the transaction, employment at private equity-backed firms mirrors that of the control group.
- But firms backed by private equity have 6 percent more greenfield job creation, that is, at new facilities in the United States, than the peer group. It appears that the job losses at bought-out establishments in the wake of private equity transactions are largely offset by substantially larger job gains in the form of greenfield job creation by these firms.

In their follow-on study, the authors focus on whether and how labor productivity changed at US manufacturing firms that were targets of private equity transactions in the United States from 1980 to 2005. The interpretation of the patterns regarding employment changes needed to be cautious, because we did not examine productivity changes at these establishments.

The authors find that while firms acquired by private equity groups had higher productivity than their peers at the time of the original acquisition, they experienced in the two-year period after the transaction productivity growth 2 percentage points more than at controls. About 72 percent of this out-performance differential reflects more effective management of existing facilities, rather than the shut-down and opening of firms. (It should

be noted that private equity investors are much more likely to close underperforming establishments at the firms they back, as measured by labor productivity.)

A Counterfactual Approach

As noted before, one form of analysis increasingly popular among economic historians is counterfactual reasoning. We can seek to understand the impact of venture capital and private equity by considering the possibilities that these sectors had not developed.[7]

A crucial argument offered by the functional perspective (Merton 1992) is that in the absence of a financial institution, other actors may evolve to play the same function. There are at least three alternative institutions that could have played these roles of venture capitalists and private equity investors: individual investors, governments, and integrated financial institutions. The evidence suggests that in some respects, these entities could have substituted for the missing institutions. But evidence also appears to suggest that these substitute institutions would have faced significant limitations, which are likely to have reduced their effectiveness.

As we mentioned earlier, angel investors were well-established as financiers to entrepreneurs long before the establishment of venture funds. By the last decades of the nineteenth century and the first decades of the twentieth century, wealthy families had established offices to manage their investments. Families such as the Phippes, Rockefellers, Vanderbilts, and Whitneys invested in and advised a variety of business enterprises, including the predecessor entities to AT&T, Eastern Airlines, and McDonnell Douglas.

Lamoreaux, Levenstein, and Sokoloff (2007) examine the financing of entrepreneurial ventures in Cleveland at the turn of twentieth century when, they argue, the region had a status not unlike that of Silicon Valley later in the century. They document that the entrepreneurs largely relied on personal connections to finance breakthroughs, whether through friends, family members, or mentors from earlier employment. These investors provided a bundle of services not unlike those of contemporary venture capitalists, including capital, certification of the new enterprise to strategic partners and other potential investors, and sometimes protection against exploitation by would-be opportunists.

But other evidence suggests that angels have important limitations. Hoberg et al. (2009) obtained access to a remarkable data set of entrepreneurial firms: the legal records of clients of Brobeck, Phleger & Harrison, a prestigious San Francisco law firm that filed for bankruptcy in 2003. They find that among the transactions that required a smaller amount of financ-

7. Another approach would be to identify the evolution of industries where these intermediaries were not active. Because the industries where investments took place were not randomly selected, this approach is fraught with interpretive issues.

ing (which, they argue, was largely a function of exogenous considerations such as the fundamental nature of the technology), the performance of the angel-backed and venture-backed firms were about equal: while the angel deals had a somewhat lower incidence of failure, many of these are inactive. The probability of initial public offerings and acquisitions, outcomes that are most often associated with financial success, was about the same. But among larger transactions, the venture-backed firms were more successful on all dimensions examined. The authors suggest that capital constraints may explain the differences: both types of financing can work for small deals, but the requirements of larger deals makes venture capital a superior mode of financing.

A second alternative source of financing is government funding. This substitute for traditional venture financing has been employed widely, but probably nowhere more extensively than Europe. Dozens of national and region-wide initiatives in recent decades have sought to promote funding for entrepreneurs and venture capital funds. To cite just one of many examples, in 2001, the European Commission provided more than 2 billion euros to the European Investment Fund (EIF); making it Europe's largest venture investor overnight. This amount is very significant relative to the roughly 4 billion euros that were invested by European venture funds in that year.

Through this large investment, the EIF intended to stimulate entrepreneurship. Europe had seen a low level of venture activity for many decades: when the ratio of venture investment to gross domestic product is computed for leading industrialized nations, the European nations are invariably among the lowest.[8] The lack of activity reflected the miserable returns that European venture investments have yielded. Venture Economics' calculations suggest that from the beginning of the industry through the end of 2009, the average European venture fund had an annual return of 1.6 percent: hardly a number to warm the hearts of investors![9] (The comparable number for US-based funds over the same period is 15.0 percent.) Thus, policymakers have argued, the low levels of fund-raising and low historical returns create a need for public financing.

Unfortunately, the numerous efforts launched by the European Union to encourage the financing of new firms have followed a depressingly familiar pattern. Even if the intention of the initiative is to create reasonable-sized funds, by the time every country, and every region in each country, gets its "fair share" of the government's money, the pie has been sliced in very thin pieces indeed. The European Seed Capital Fund Scheme is one telling

8. These calculations are compiled from various publications and websites of the Canadian, European, Israeli, and US (National) venture capital associations, as well as those of the *Asian Venture Capital Journal.* In some nations where venture capital investments are not clearly delineated, we employ seed and start-up investments. The GDP data are from the Central Intelligence Agency (2009).

9. Return data taken from http://banker.thomsonib.com/ta/.

example. As Gordon Murray (1998) points out, these funds (which typically had under €2 million in capital) were so undercapitalized that even if they did nothing besides pay for the salary of an investment professional and an administrative assistant, rent for a modest office, and travel, and never invested a single dollar, they would run out of capital long before their assigned ten-year life was up. Moreover, with so few euros to disperse, the investments they could make were tiny. Certainly, they were insufficient to get the typical entrepreneurial company to the point where it could go public, or even, in many cases, to the point where it would be interesting to a corporate acquirer. For a number of groups, their best hope of achieving any return from their investments was to sell the stakes back to the companies they had bought them from. This is hardly a way to achieve the European Commission's goal of providing capital to needy entrepreneurs.

A final alternative, seen particularly among latter-stage investors, are integrated financial institutions. In a number of nations, such as Japan, the bulk of the financing to rapidly growing and restructuring entities are provided by large integrated financial institutions. Even in the United States, where the independent private equity industry was founded, over one-quarter of all private equity transactions involve a bank-affiliated fund (Fang, Ivashina, and Lerner 2010).

It might be thought that these diversified financial institutions, in addition to substituting adequately for private equity groups, might actually be able to undertake investment more successfully. Such a conclusion is suggested by the literature on internal capital markets. Stein (1997), for example, sees organizational diversification across activities (in this context, banks that can engage in either underwriting or investing) as an important element of efficient capital allocation. When opportunities are poor in one industry, he argues, managers can maintain their overall capital budget (which they value in and of itself) while still making good investments in their other industries. By contrast, managers of narrowly focused firms with poor investment opportunities have no place else to invest and, in an effort to maintain their capital budgets, may end up investing in negative net present value projects.

Empirical data suggests, however, that the effectiveness of these institutionalized investors is far less effective in practice. In particular, the share of transactions affiliated with banks is procyclical, peaking at times of big capital inflows into the private equity market. Transactions done at the top of the market are most likely to experience subsequent distress, and this pattern is especially pronounced for transactions involving banks' private equity groups. This result is particularly striking because prior to the transaction, targets of bank-affiliated investments generally have significantly better operating performance than other buyout targets, though their size and other features are similar. The results suggest that incentive problems and an inability to add value to portfolio companies have limited the success of bank-affiliated funds.

These plausible counterfactual histories, in which venture and private equity investors were replaced by angels, governments, or integrated financial institutions, suggest that while important aspects of the venture capital and private equity process can be duplicated, the alternative approaches also have their own challenges, which makes it hard to duplicate the free-standing investment organizations. While we must be cautious in our interpretation, the counterfactual analysis suggests that these institutions could not have been readily replaced. Unlike the railroads, which could have been replaced by alternative transportation modes, these financial innovations may have had a larger unique contribution to economic growth.

Taking Stock

It should be noted, however, that all of these studies have important limitations. First, these studies consider venture capital and private equity in aggregate. As alluded to earlier, both industries have been characterized by highly "lumpy" fund-raising, where a few years account for the peak of the activity. These years are also characterized by poorer private returns and higher rates of bankruptcy, which might suggest that the social returns from these periods are modest as well.

These limitations are particularly acute in the case of the private equity studies. None of these studies can grapple with the consequences of the 2005 to 2008 market peak, which accounted for fully 47 percent of the private equity raised (in inflation-adjusted dollars) between 1969 and 2008.

Moreover, the findings that have been completed to date raise questions about what goes on during these boom periods. Axelson et al. (2009) document the cyclical use of leverage in buyouts. Using a sample of 1,157 transactions completed by major groups worldwide between 1985 through 2008, they show that the level of leverage is driven by the cost of debt, rather than the more industry- and firm-specific factors that affect leverage in publicly traded firms. The availability of leverage is also strongly associated with higher valuation levels in deals.

Similarly, Davis et al. (2009) find that the positive productivity growth differential at target firms (relative to controls) is not even. Rather, it is larger in periods with an unusually high interest rate spread between AAA-rated and BB-rated corporate bonds, and virtually nonexistent during periods with low spreads. One interpretation of this pattern is that private equity groups are committed to adding value to their portfolio only during periods when making money through other means (e.g., through leverage and financial engineering) is not feasible; that is, during periods when private equity activity is relative quiescent.

If firms completing buyouts at market peaks employ leverage excessively and are less likely to focus on adding value, as their findings suggest, we may expect industries with heavy buyout activity to experience more intense subsequent downturns. Moreover, the effects of this overinvestment would be exacerbated if private equity investments drive rivals not backed by private

equity to aggressively invest and leverage themselves. (Chevalier [1995] shows that in regions with supermarkets receiving private equity investments, rivals responded by entering and expanding stores.)

But this claim remains unproven. A counterargument, originally proposed by Jensen (1989), is that the high levels of debt in private equity transactions force firms to respond earlier and more forcefully to negative shocks to their business. As a result, private equity-backed firms may be forced to adjust their operations earlier, at the beginning of an industry downturn, enabling them to better weather a recession. Even if some private equity-backed firms eventually end up in financial distress, their underlying operations may thus be in better shape than their peers, which facilitates an efficient restructuring of their capital structure and lowers the deadweight costs on the economy. Consistent with this argument, Andrade and Kaplan (1998) study thirty-one leveraged buyouts from the 1980s that became financially distressed, and found that the value of the firms postdistress was slightly higher than the value before the buyout, suggesting that even the leveraged buyouts that were hit most severely by adverse shocks added some economic value. Thus, the extent to which the steady-state findings are weakened and undone by the intense cyclicality in these markets remains an open question.

11.3.3 Mutual Funds and Exchange-Traded Funds

Just as venture capital and private equity have become important components of the modern US economy, mutual funds (including exchange-traded funds) have become a dominant force in the investment management arena. While there has been substantial work on mutual funds, little of it directly addresses the social welfare consequences of this innovation. To lay out the approach for studying its implications, we (a) provide a brief history of the US mutual fund industry; (b) demonstrate its economic importance; (c) highlight the areas in which funds may have positively and negatively influenced social welfare; and (d) sketch out a counterfactual history to draw out these consequences.

A Brief History of the Innovations in the US Mutual Fund Industry

While mutual funds have antecedents in nineteenth-century British Unit Investment Trusts (comparable to closed-end funds today) and earlier European structures, the "modern" open-end mutual fund was created in 1924.[10] The Massachusetts Investment Trust, launched in March 1924, was followed in quick succession by the State Street Investment Corporation in July and the Investment Corporation in November 1925. Like the investment trusts

10. For a history of the fund industry, see Fink (2008) and the references therein. For a useful list of innovations in the fund industry, see http://www.icifactbook.org/fb_appd.html. For the early predecessors of modern mutual funds, see Goetzmann and Rouwenhorst (2005), chapter 15, "The Origins of Mutual Funds" by Rouwenhorst.

that preceded them, these new funds were pooled investment vehicles offering professional active investment management services. The key innovations were the structure of the funds, as well as the manner in which redemptions were handled. Open-ended mutual funds, as they would come to be known, had a single class of investor claims in the form of equity, rather than a levered structure (still common in closed-end funds). More importantly, they allowed investors to buy or redeem shares on a daily basis at net asset value, unlike the prior investment trusts, which traded on exchanges and were (and are) typically sold at discounts or premia to net asset value. The offer of shares and redemptions was daily and continuous, as opposed to the infrequent issuance of new shares by prior investment trusts.

The next major wave of innovation in mutual funds took place in the early 1970s. Up until this time, funds had held portfolios of stocks, and, to a far lesser degree, bonds. No fund had primarily held short-term money market instruments and designed itself to maintain a stable net asset value. In September 1972, the Reserve Fund was launched, followed a few weeks later by a competing fund, the Capital Preservation Fund, and in 1974 by offerings by Dreyfus and Fidelity. The latter allowed shareholders to redeem shares through a check-writing feature. The innovation of money market funds was not the holding of short-term instruments per se, but their mechanisms to maintain stable net asset values through either rounding their net asset values (NAVs) to the nearest penny (penny rounding funds), by valuing their portfolio at amortized cost (versus market value), or by adding or subtracting realized gains and losses from accrued income on a daily basis. (See Fink (2008, 84). These practices would eventually be memorialized into regulation through section 2a7 of the 1940 Act, which would permit amortized cost accounting and penny rounding methods for money market funds.

At about the same time, in the early 1970s, the first municipal bond funds by Kemper and Fidelity were offered, expanding the asset classes in which fund shareholders could invest. In the early 1970s, institutional index funds were first offered. Rather than use active management or a completely unmanaged fixed portfolio, these investments offered investors the return of a stock index (including the occasional rebalancing due to additions/deletions by the index). The next major retail innovation would take place in 1976, with the creation of the first indexed mutual fund, Jack Bogle's Vanguard First Index Investment Trust. The First Index Investment Trust brought the indexing concept to retail investors in a mutual fund structure, wrapped around a low-cost, high-service business model that was informed by Bogle's experiences, beginning with his 1951 Princeton college thesis, "The Economic Role of Investment Companies" (see Slater 1997).

A more recent innovation, similar in spirit to index funds but with a different institutional structure, was created in 1992 by Leland O'Brien Rubinstein in the form of SuperTrust and rapidly followed by a similar offering by the American Stock Exchange in the form of SPDRs (see Tufano

and Kyrillos 1994). The products, which would later morph into exchange-traded funds, had features of the old fixed-portfolio investment trusts and closed-end funds, in that they passively managed funds that were bought and sold on exchanges. The key innovation was to find a way to keep these funds trading at fundamental value or net asset value, rather than at fluctuating discounts and premia. The traditional open-end fund did so by contract form, allowing shareholders to buy and redeem shares at the NAV. The ETF innovation kept the link to NAV by allowing institutions to assemble the portfolio of underlying securities and create new ETFs (and disassemble the ETF portfolio into its underlying components). By creating a direct link between the security and its underlying components, ETFs minimize discounts or premia to NAVs. Overall, the fund industry has witnessed a high level of innovation over the past decades.

The Economic Importance of the US Mutual Fund Industry

Over the past few years, policymakers have been debating whether mutual funds, or at least money market mutual funds, are "systemically important" and should be regulated by others beyond the SEC. Regardless of the outcome of this regulatory debate, there is little question that mutual funds are one of the most successful financial innovations of the twentieth century. Whether measured by their growth rates, adoption rates, fraction of capital intermediated in the economy, or importance to household balance sheets, mutual funds are critical to the economy. Furthermore, evidencing the innovation spiral, the original actively managed stock and bond mutual fund structure has been the chassis on which we have seen innovations such as index funds, exchange-traded funds, sector funds, and money market funds.

On an absolute level, the US mutual fund industry is simply enormous. As of October 2009, industry assets (excluding ETFs) exceeded $10 trillion, as shown in table 11.2 from the Investment Company Institute's data on the 7,762 funds in operation.

These absolute numbers, while staggering in size, do not put the economic

Table 11.2 **Total net assets of US domiciled mutual funds, October 2009 (billions of dollars)**

Stock funds	4,596.2
Hybrid funds	604.5
Taxable bond funds	1,682.5
Municipal bond funds	443.9
Taxable money market funds	2,951.3
Tax-free money market funds	409.9
Total	10,688.3

Source: http://www.ici.org/research/stats/trends/trends_10_09. This total excludes exchange-traded funds, with $738 billion in assets.

Table 11.3 Composition of US household financial market assets, 1950 and 2008

	1950	2008	Gain/Loss
Bank-system deposits	28.1	18.2	–9.9
Money market mutual funds	0.0	4.5	4.5
Direct holdings of stocks and bonds	51.1	29.0	–22.1
Mutual funds (stock, bond, balanced)	0.7	10.0	9.3
Pension reserves (incl. DB and DC plans)	5.2	30.4	25.2
Other	14.9	7.9	–7.0
Total financial market assets	100	100	
Total mutual fund share	0.7	14.5	13.8

importance of the fund industry into context. One way to do so is to examine their adoption, in aggregate, by an important sector of the economy: households. Table 11.3 shows the breakdown of aggregate financial assets held by the US household (and nonprofit) sector in 1950 and 2008, as calculated by the Federal Reserve's Flow of Funds accounts.[11]

The pervasive impact of mutual funds can be seen in this aggregate balance sheet. First, from 1950 through 2008, households held far fewer "deposits," defined broadly, with the deposit-like share going from 28.1 percent of financial assets to 22.7 percent. Of this 22.7 percent, money market funds accounted for 4.5 percent, or nearly one-fifth. Secondly, in 1950, slightly over half of all household financial assets were in direct holdings of stocks and bonds. By 2008, this figure had dropped to 29.0 percent, but 10 percent were held in long-term stock and bond mutual funds, which increased from 0.7 percent to 10.0 percent over fifty-eight years. Finally, the decline in direct holdings of stocks and bonds was more than offset by an increase in holdings in pension reserves, which rose from 5.2 percent to 30.4 percent of all household financial assets. A large fraction of these pension assets are in defined contribution plans, which in turn are invested in mutual funds. Putting these three elements together, mutual funds have had a profound impact on the household balance sheet.

The Social Welfare Implications of Mutual Funds

While there is little question that mutual funds have not only been a financial innovation, but a successful one in terms of adoption, how can we gauge the social welfare implications of this sector? Unlike the venture capital and private equity innovations, where researchers have documented employment, business formation, product innovation, and productivity impacts, there is far less done at a macro level on the social welfare impacts of the

11. http://www.federalreserve.gov/releases/z1/ or the various data series. These numbers include financial assets, excluding equity in unincorporated businesses, to reflect financial market claims.

fund industry. In part, this may reflect the fact that funds are not typically involved with portfolio firms in the same direct way as private equity or venture capital firms. Their impact on social welfare would come from benefits to investors who seek low-cost diversified portfolios, or to capital markets, as information processors and as deep pools of capital. Our discussion focuses primarily on the former—the costs and benefits to investors.

It is clear from the past six decades of history that households' revealed preference has been to hold funds more than to hold individual securities—and to hold securities more than bank deposits. If one were to assume that these choices were the direct result of the existence of mutual funds, one could provide a crude estimate of the return differential earned by investors as a result of the mutual fund innovation, one portion of the social welfare gains from innovation. For the purpose of this thought exercise, suppose that households allocated their assets between cash (earning the risk-free rate) and the market (stocks and bonds), which earns a premium over the risk-free rate.

Define:

r_f = The risk-free rate, a proxy for the return on deposits.
RP = The equity risk premium on an unmanaged portfolio of assets.
M = The fraction of assets held in securities (market), prior to the introduction of mutual funds.
ΔM = The incremental fraction of assets held in securities (market) as a result of mutual fund introduction. Presumably, $\Delta M > 0$, based on the decrease in deposits over the postwar period.
f = The weighted average incremental fee charged by funds in excess of the embedded fees in direct holdings of equities, where the weight is given by the mix of mutual fund holdings as a fraction of all market holdings. The sign of f is unclear: while funds have explicit fees, there are implicit fees with holdings in banks (in the form of deposit-loan spreads) as well as explicit fees.

Before and after the introduction of mutual funds, the household sector's return would be

$$E(R_{\text{pre}}) = (1 - M)r_f + (M)(r_f + \text{ERP})$$
$$E(R_{\text{post}}) = (1 - M - \Delta M)r_f + (M + \Delta M)(r_f + \text{ERP-}f).$$

Taking the difference between these two and combining terms, we could calculate a net increase in return equal to

$$-Mf + \Delta M(\text{ERP-}f),$$

where the first term is the decrease in private return due to incremental weighted average fees and the second term is the net increase in return due to the increased holdings of risky market assets.

Even a quick inspection of this naïve formula makes clear some of the challenges with estimating this differential. First, an increase in returns that is accompanied by a commensurate increase in risk does not increase social welfare, unless we can show that the representative investor was better able to move closer to some optimal level of risk taking.

Second, it assumes that the introduction of funds does not affect the risk-free rate or the market risk premium. However, if in aggregate institutional and individual investors moved more funds into the market and away from banks and other low risk investments, these returns, and other market-wide elements such as liquidity, could easily be affected. The increased demand for riskier assets from the deeper pool of potential market investors could lower costs of capital for firms. The more intensive alpha-seeking behavior of funds could make prices more reflective of efficient market levels and reduce bid-ask spreads. However, both of these assertions would need to be proven.

Third, it attributes the change in deposit holdings entirely to funds and does not consider secondary influences of the innovation. For example, holdings of higher-risk portfolios would tend to increase household wealth but lead to greater fluctuations in wealth. The former would tend to increase the willingness to hold risky assets, and the latter might depress this willingness. Also, the introduction of money market funds might have led households to hold more in low-risk assets.

Fourth, while mutual funds clearly charge fees, and ample research demonstrates that funds cannot persistently beat the market, we need to calculate the incremental fees incurred by household investors. While the absolute level of mutual fund expenses is greater than zero, and while turnover is far higher than a passively managed portfolio, the relevant comparison for our purposes would be the incremental fees and turnover relative to the benchmark prefund portfolio, composed of bank deposits and direct holdings of securities. A directly held portfolio would have individual investors (or a bank trust department) managing their own investments, paying retail commissions, and implementing their own trading strategy. Most likely, this alternative would also have households less well diversified.

This simple specification makes clear some of the elements left out of this analysis. On the positive side, we would need to capture:

1. Greater development of capital and debt markets as a result of new institutions. There is extensive literature on financial development and economic development. While there are ongoing debates about the causality and magnitude of these relationships, one would have to acknowledge that mutual funds have been a substantial element of financial development.
2. Greater holding of foreign securities to counteract home bias. French (2008) documents a substantial increase in US holdings of foreign securities, which partially might be attributed directly or indirectly to mutual fund holdings.

3. Greater savings overall. While it is purely speculative, one wonders what the savings rates of individuals would have been in the absence of mutual funds.
4. Institutional competition for the fragmented and regulated banking industry. On this latter point, the development of money market funds was an explicit reaction to the interest rate caps imposed by Regulation Q.

Finally, without a mutual fund sector, would we have seen the development or widespread adoption of defined contribution (DC) pension plans, where the employee selects his or her investments from a menu largely consisting of retail funds? Technically, it would have been possible to move to this system were no retail funds in place by offering retail investors the option of investing in institutional products. However, realistically, this might have been difficult because along with the investment management aspects of mutual funds came the record-keeping systems that would support DC plans. Also, retail offerings of funds likely made it easier for firms and employees to understand and to get comfortable with workplace-based defined contribution plans. By the time DC plans were introduced (spurred by the 1974 Employee Retirement Income Security Act [ERISA] rules and the 1978 Revenue Act), consumers had extensive experience with funds, with over 10 million mutual fund accounts in America.[12]

On the negative side, this specification would not capture (a) "Excessive" rent seeking by mutual fund companies and the associated transfer of wealth from investors to the industry; (b) "Excessive" or "insufficient" savings by individuals; (c) "Excessive" risk taking by individuals; and (d) "Costly" disintermediation of the banking sector, including the relative loss of regulatory control over the money supply that bank regulators had traditionally enjoyed. All of these costs, and benefits, are difficult to measure because in many instances we lack models to determine the optimal levels of these quantities. The optimal level of risk taking in the economy, for example, depends on preferences and risk aversion, which are not exogenous.

Counterfactual Histories

Trying to untangle any of these issues is difficult enough, but the specification also makes clear that one cannot analyze the social welfare consequences of the fund industry except in context. Had mutual funds not been invented (or adopted), what counterfactual history might have emerged? Which are plausible and "miracle" alternatives? Surely investing and pooling would have continued as core functions in a financial system, but the institutional arrangements would have been different without mutual funds, index funds, and ETFs. Some possible alternatives include:

12. See http://www.icifactbook.org/fb_data.html, table 1.

1. Continuation of the prefund status quo, involving banks, bank trust departments, direct holdings of stocks and bonds, brokers and financial advisors, closed-end funds, and opaque holdings of securities through intermediaries such as insurance companies.
2. Modified status quo outcomes, where some of these institutions came to dominate others (e.g., a movement to greater intermediation but in the form of insurance-wrapped investments).
3. "Miraculous" innovations, such as fractional shares and bonds that would permit individual investors to create diversified portfolios at a small scale.

The first possibility of the prefund status quo is largely a banking and direct security holding alternative. Closed-end funds would have remained a minor player in the economy. Fink (2008) argues that closed-end funds became marginalized in the wake of the events of 1929, and direct holdings of securities—sold by brokers—were preferred as the means by which households acquired exposure to the "market." In this counterfactual world, households would hold poorly diversified, rarely rebalanced portfolios of a small number of securities. They would have been advised by bank trust departments (for the very wealthy), securities brokers, and popular periodicals. One could not assume index funds or ETFs in this counterfactual, as they were part of the innovative process we are analyzing. One may not even be able to assume low-cost brokerage models, as they too, were a relatively recent financial innovation.

While the actively managed mutual fund industry is often criticized for failing to produce reliably positive excess returns or alpha, it is less likely that investors would have performed better on their own employing this direct-ownership counterfactual. Perhaps the most complete analysis of the social welfare impacts of mutual funds, in the context of active investing, can be found in French's (2008) AFA Presidential Address. In it, he documents the perpetual, and costly, search for alpha, estimating the deadweight loss to be about 67 basis points per year relative to passive investing. French convincingly documents that actively traded mutual funds are considerably more expensive than passive portfolios, but assumes virtually zero costs for direct-held portfolios: "I assume the only expenses individuals incur when they hold shares directly are trading costs, which are included in the aggregate estimates below. I ignore, for example, the time they spend managing their portfolios and the cost of subscriptions to Value Line and Morningstar" (1543). It is unclear if he includes noncommission payments to financial advisors, bank trust departments, or others who would facilitate the direct investing activities of investors. In our counterfactual, we would need to include these costs, which were likely sizable. Furthermore, it is not clear that the direct buyers of securities would receive excellent investing advice. Recent evidence by Bergstresser, Chalmers, and Tufano (2009), for example,

shows that broker-sold mutual funds consistently underperform direct-sold funds. If this is any indication, replacing thousands of fund managers with millions of even less well-informed brokers is not likely to increase household wealth. One might imagine that household portfolios might show even greater home bias and would virtually certainly not contain index-like fund holdings.

The second alternative is that an intermediated solution other than open-end funds and ETFs could have emerged. Despite their lack of popularity in the 1930s, perhaps closed-end funds might have enjoyed renewed popularity. While they provide pooling and liquidity to investors, it is unlikely that this path would have led to higher social welfare for investors. First, closed-end funds routinely trade at discounts and premia to net asset values, and investors would need to bear this additional discount risk (in addition to the risk of fluctuations in the portfolio's NAV). Second, by their nature, closed-end funds have a fixed amount of assets under management, versus an open-end fund, which can expand or contract assets in response to demand. Closed-end funds would therefore benefit less from economies of scale due to growth than would open-end funds. Closed-end funds also require incremental distribution expenses and legal expenses to start new funds to accommodate new demand, whereas open-end funds can accept new assets at virtually no administrative costs.

Another possible intermediated solution would have been that insurance-based investments would have met demand had the fund innovation not taken place. Revealed preference suggests that there is greater demand for funds than for bundled insurance-cum-investment products. As of the end of 2008, mutual funds (excluding ETFs) held $9.6 trillion in assets; by comparison, total assets held by life insurers was $4.6 trillion, with much of the latter backing noninvestment term insurance products.[13] Given this sizable difference in revealed demand, it is difficult to believe that a bundled insurance-investment product would have satisfied investor preferences as well as funds have. Furthermore, the bundling of these products makes them more difficult to explain and sell, likely leading to higher transaction costs (and possibly poorer matching of products to consumer needs). For a discussion of these problems and the resultant market failures that can give rise to regulation, see Campbell et al. (2010).

Another, more miraculous possibility is that an alternative functional substitute for funds would have emerged, providing low-cost pooling and investment management, small lot sizes for diversified portfolios, and liquid-

13. Data from http://www.icifactbook.org/fb_data.html (table 1) and http://www.acli.com/ACLI/Tools/Industry+Facts/Assets+and+Investments/. The mutual fund number includes $3.8 trillion in money market funds and $5.8 trillion in long-term investments. However, the life insurance figure includes both assets held to back term policies (with no investment element) and other policies (universal, whole life, etc.) with an investment component. In 2008, about three-quarters of all life insurance purchases were for term insurance, which does not have an investment component (http://www.acli.com/NR/rdonlyres/0BFEABCA-1E2A-4F4C-A879-95CF104238AB/22608/FB0709LifeInsurance1.pdf, table 7.2).

ity in the form of daily trading. By 2000 or so, this alternative history might not have seemed far-fetched. A number of startups offered products of this sort, allowing investors to directly buy pools of securities, including fractional shares, and provided a high level of liquidity. For example, one of the first of these innovators, folioFN permitted investors to buy folios (portfolios) of stocks (as well as mutual funds) in fractional shares. However, it would take the development of the Internet, and the adoption of Internet-based transacting, to make this counterfactual a reality. Even so, one wonders about the ultimate returns earned by direct investors. Recent behavioral finance work, by Odean (1999), Barber and Odean (2000, 2001a, 2001b, 2002, 2004), and others call into question the investing acumen of individual investors.

Overall, a more extensive consideration of these and other counterfactuals would likely suggest that the innovation of mutual funds, index funds, and ETFs likely were beneficial for investors, relative to other reasonable counterfactuals.

11.3.4 Securitization

Pooling is a timeless function of financial systems, and our first two case studies focus on pooling vehicles using different forms of intermediation. Over the past four decades, another major innovation that performs the pooling function has been securitization.[14]

Like other pooled vehicles, which assemble portfolios of assets (stakes in new companies, shares in firms, or holdings of bonds) and sell claims against them, securitization vehicles bundle a variety of financial claims, often in the form of retail IOUs (mortgages, auto loans, student loans, credit card receivables) and sell claims against them. In venture capital and private equity, there are often multiple classes of claims (general partners and limited partners), in open-end funds a single claim (equity holders), and in closed-end funds often multiple claimants (equity and debt, sometimes.) Securitized vehicles can have a single class of investors (if purely pooled vehicles) or can create multiple classes of investors. While early securitization used the former method, much of modern securitization gives different investors varying exposures to credit or prepayment risk. Even more complicated structures create tranched structures using already pooled structures (collateralized debt obligation [CDO]-squareds) or using derivatives (synthetic CDOs). For the purpose of this discussion, we will focus on "simple" securitization structures, recognizing that some of the most vociferous criticism was directed at the more complex structures.

The History and Extent of Securitization

While there was securitization of a sort in the 1920s, the practice as we know it came into widespread adoption in the 1970s and 1980s, beginning

14. For a general discussion of the pooling function, see Sirri and Tufano (1995).

with the securitization of home mortgages.[15] Before that time, most home mortgages were originated, funded, and serviced by banks and credit unions or, if they were government-insured mortgages, were bought by government-owned Fannie Mae.[16] In some instances, loans were sold from one party to another, but this whole loan market was fairly illiquid.

The first major development in securitization was the introduction of the pass-through mortgage backed security (MBS), first issued by Ginnie Mae in 1970. In a pass-through, a portfolio of mortgages are bundled together and investors receive all principal and interest payments. Pass-through MBS or participating certificates combined the sale of loans, the bundling of mortgages into a pool, and the use of an off-balance-sheet structure. Unlike later securitizations, these instruments had a single class of investors, who shared proportionally in the portfolio's risks and returns, including prepayment risks.

The next major innovation in securitization was the development of tranched structures, first used in the Collateralized Mortgage Obligation (CMO) issued by Freddie Mac in 1983. In this multiclass security, a set of rules predetermined which investors got which cash flows. A major concern in these structures was to allocate prepayment risk among investors. Because borrowers have the right to prepay their mortgages and would tend to do so when it was to their advantage (and to the disadvantage of the lenders), prepayment risk (or the embedded call option in mortgages) was an unattractive feature from the perspective of investors. Under CMO structures, certain investors willing to take on greater prepayment risk would accordingly earn higher promised returns, while other investors would be the last to be prepaid and therefore earn lower promised returns. Other structures would modify the division of prepayment risk (and credit risk) among investors in more complex ways. With the passage of the Tax Reform Act of 1986, which allowed the Real Estate Mortgage Investment Conduit (REMIC) tax vehicle, CMO issuance and securitization expanded dramatically.

The volume of securitized home mortgages grew from $28 billion in 1976 to $4.2 trillion in 2003.[17] Government-sponsored entities (i.e., Fannie Mae and Freddie Mac) played an important role in this process by standardizing mortgage products, pooling mortgages into mortgage-backed securities, and guaranteeing investors against losses.[18] Securitization supported the development of mortgage brokers and specialized mortgage originators who developed a new "originate-to-distribute" model, as well as third-party servicing.

15. The material in this section draws heavily upon Ryan, Trumbull, and Tufano (2010).
16. http://www.fundinguniverse.com/company-histories/Fannie-Mae-Company-History.html.
17. Loutskina and Strahan (2009).
18. Frame and White (2005).

Other lending activities also used securitization as a financing technique. Automobile loans were first securitized in 1985; credit card loans followed in 1986.[19] By 2006, approximately 55 percent of all mortgages, 45 percent of all credit card loans, and 16 percent of nonrevolving loans (many of which are auto installment loans) were securitized.[20] Over time, these networks of firms and investors displaced traditional lenders. For a more complete discussion of the history of securitization, see the definitive treatise by Frankel (2006).

Assessing the Social Welfare Implications of Securitization

Much attention has been focused on the way in which changes in financial intermediation, especially in mortgages, have influenced the national and global economy.[21] The difficulty with assessing the impacts of securitization, however, stems from the many different elements associated with this class of innovations. These elements include, but are not limited to, the following:

- The sale of a loan from the original lender to another investor(s).
- The bundling of loans from a single or multiple lenders, with subsequent sale to investors.
- The standardization of the underlying assets encouraged by parties putting together or guaranteeing pools.
- The guaranteeing of assets, fully or partly, by government or private parties.
- Other credit enhancement, for example, through overcollateralization.
- The tranching of claims to create multiple securities differentiated by credit or prepayment risk.
- The creation of stand-alone loan originators (mortgage brokers) who tended not to have an economic interest in the long-term viability of originated loans.
- The creation of stand-alone servicers with a complex set of incentives as agents of diffuse shareholders.
- The creation of securitized structures using other securitized structures (or derivatives) as underlying assets.
- The creation of securitized structures using high-risk (subprime) loans as the underlying assets.
- The use of and reliance upon credit ratings that may fail to take into account the level of risks in some of these structures.

19. *Asset Securitization Comptroller's Handbook* (1997).

20. Mortgage data from Rosen (2007). Revolving and nonrevolving debt data from Federal Reserve Statistical Release, Series G19, http://www.federalreserve.gov/releases/g19/Current/.

21. Ashcraft and Schuermann (2008); Berndt and Gupta (2008); Coval, Jurek, and Stafford (2009); Hoffman and Nitschka (2008); Mayer, Pence, and Sherlund (2009); Mian and Sufi (2008); Purnanandam (2009); and Shiller (2008).

Critiques of the innovation of "securitization" must acknowledge that a pass-through securitization of prime mortgages originated by banks is quite a different phenomenon from a CDO-squared issue where the underlying asset is a low-ranked tranche of a different CDO, whose underlying assets, in turn, are a portfolio of no-documentation subprime loans originated by mortgage brokers.

The second challenge with analyzing the welfare impacts of securitization, say of home mortgages or student loans, is to assess the appropriate outcome metric. There are a variety of legitimate measures. For example, some early studies suggest that the first decades of securitization led to lower interest rates for borrowers (see Hendershott and Shilling 1989; Sirmans and Benjamin 1990; and Jameson, Dewan, and Sirmans 1992). Others point to the wider availability of credit, leading in turn to considerably higher home ownership rates, which rose from about 62 percent in 1960 to almost 69 percent in 2004, with the strongest gains among nonwhite American households.[22] Against these positive metrics of lower rates, expanded credit availability, and broader homeownership, we must consider the cost of higher levels of foreclosure, especially among subprime borrowers, putatively the primary beneficiaries of this increased lending.

We can quantify the benefits of lower costs of financing, but how would one quantify the benefits of having an additional 1 percent of households owning homes, or the costs of 1 percent of homeowners losing their homes through foreclosure? Neither direct measurements nor a counterfactual approach can overcome the problem of multiple metrics, some of which do not lend themselves to quantitative measurement.

Identifying Counterfactual Alternatives

Which counterfactual history might we use to compare against the actual past where securitization has financed much consumer debt? Using the mortgage market as the primary research site, these alternatives might include the following:

- Depositories continue to originate and hold mostly prime loans, with limited whole loan sales to other depositories or to specialized mortgage investors.
- Depositories continue to originate mostly prime loans, but some are bundled in the form of pass-through (single class) MBS securities.
- Depositories continue to originate mostly prime loans, but some are bundled in the form of either pass-through (single class) securities or multiclass (CMO) structures—but not more complex structures (synthetic CDOs or CDO-squared structures).

22. See http://www.census.gov/hhes/www/housing/hvs/annual05/ann05t12.html for the national figures and http://www.census.gov/hhes/www/housing/hvs/annual05/ann05t20.html for the breakdowns by race.

Of course, these three alternatives are just a few of the nearly unlimited number of counterfactuals, which could be made by layering on (a) subprime lending; (b) the originate-to-distribute model using independent mortgage brokers, reimbursed through yield spread premia; (c) the existence of less-optimistic credit ratings by rating organizations. Even beyond these variants, we would need to consider even broader counterfactuals. For example, a world in which depositories originate and hold mortgages would likely operate quite differently in a setting where branching and intrastate banking were prohibited versus one in which national banking organizations could create diversified portfolios of loans by virtue of their scope.

Comparing the first (no pooling) to the second (pass-through MBS) and third (simple CMO) phases of securitization, there is some evidence that the early securitization gave rise to measurable benefits. As noted before, mortgage rates fell in the first phase, and homeownership rose from 61.9 percent to 64.4 percent from 1960 to 1980, and then to 66.2 percent in 2000. Elmer and Seelig (1998) document and study the general rise in foreclosure rates from 1950 through 1997. They examine the empirical determinants of this time series, and conclude that securitization and the ancillary activity of third-party servicing does not explain the trend in foreclosures. (Rather, they find that measures of household debt and savings are better predictors of foreclosures.) While this evidence is far from complete, it is suggestive that the roughly first three decades of securitization were not likely welfare-reducing. Indeed, having deep pools of capital to fund national mortgage markets was a likely improvement over local mortgage lenders.

The more recent history of securitization is probably a different matter. It is not clear that the economy unambiguously benefited from ever more complex structures, higher-risk underwriting of subprime borrowers, sloppier underwriting standards in general, and an increasing role for mortgage originators with few long-term incentives. In almost textbook fashion, we see an innovation more widely diffused, used by a new population (riskier borrowers) in new ways (in securitizations of securitizations), and purchased by less experienced investors (relying on ratings).

While determining social welfare implications of securitization is difficult, even establishing simpler facts about the phenomenon is not simple. A large body of papers, including a number of recent working papers, examine aspects of securitization and attempt to measure the direct impacts of the practice. For example, studies reach contradictory conclusions about whether riskier banks use securitization, whether they have lower funding costs, or whether securitization increases loan supply. (For a summary of some of these studies see Panetta and Pozzolo [2010], who study credit risk transfer in over 100 countries.) For example, a recent working paper, using propensity scoring techniques to try to determine a counterfactual (had banks not chosen to securitize) finds that after controlling for whether banks choose to securitize, there is no statistically significant impact of securitiza-

tion on banks' funding costs, credit exposure, or profitability (Sarkisyan, et al. 2010). While the authors frame the work in terms of a counterfactual, it addresses a far narrower question: How would banks have performed had they not used securitization (but implicitly assuming that securitization exists and is used by other institutions)? Even so, using similar data but an instrumental variables approach using bank size as an instrument, Jiangli and Pritsker (2008) conclude that securitization played a positive role in reducing insolvency risk among banks. There are numerous papers that empirically analyze the effect of securitization on bank stability, as measured by *Z*-scores, systemic risk, and other measures. Not surprisingly, they, too, reach contradictory results. For a recent survey—and evidence of a negative relationship between securitization and bank financial soundness—see Michalak and Uhde (2010).

The extant literature largely attempts to address how securitization affects individual banks, but to assess the social welfare implications of this innovation, one needs a broader frame. Gorton and Metrick (2010) summarize the reasons for the growth of modern securitization (reduction in bankruptcy costs, tax advantages, reduction in moral hazard, reduced regulatory costs, transparency, and customization). They, along with Adrian and Shin (2010), highlight how securitization was part of a larger set of innovations that constitute the so-called shadow banking system, in which market-based financial intermediaries replaced traditional banks. These other elements include money market mutual funds and repo contracts. Together, these papers demonstrate another challenge with analyzing the innovation of securitization: it is closely linked to a network of innovations, so it is difficult, if not impossible, to separate their effects.

Where does this leave us? Certainly, the existing work on securitization, even if ambiguous, provides a useful first step to understanding this innovation. The precise details of securitization, in conjunction with other trends that make up the shadow banking system, will probably thwart any definitive scientific study of the phenomenon. However, one can imagine projects, similar to Fogel's, with all of the same critiques, that consider the following counterfactuals:

- What if only prime mortgages had been securitized?
- What if no "no-doc" mortgages would have been allowed to be securitized?
- What if rating agencies would have rated more poorly (or refused to rate) certain highly structured transactions?

For example, Fogel examined access to nonrail transportation modes to understand the constraints on trade had there been no railroads. In theory, one could examine the holders of various securitized products and their investment charter restrictions to determine what fraction of holdings realistically could have been placed into the market were the issues not rated.

While other institutions might have emerged with an appetite for unrated securities, the exercise would provide a meaningful boundary on the problem. Similarly, if one were to constrain the securitized pools by excluding subprime and no-doc loans, what would the pro forma default rates have been and how might they have rippled through the economy? We suspect that a thoughtful, step-by-step counterfactual approach, inspired by Fogel's 250 page masterpiece, would provide many insights not available from more traditional studies. While a counterfactual approach does not simplify matters much, it has the tangible benefit of forcing us to focus on which elements of securitization are most problematic.

11.4 Conclusions and Other Research Directions

As we have highlighted here, while existing empirical evidence and conceptual frameworks can tell us much about financial innovation, there are substantial unanswered questions. In this final section, we discuss some of the promising avenues for future research. While no method is without problems, these approaches complement one another.

The first approach is to examine settings where there are constraints on financial innovation. The exploitation of exogenous constraints is by now a well-accepted technique in empirical economic research. In particular, a classic example of such constraints that might present an opportunity for careful study is Islamic finance, particularly as practiced in Saudi Arabia and the Persian Gulf. As commonly interpreted, sharia-compliant financial structures exclude the use of debt and multiple classes of equity. Such a setting may provide a "natural experiment" for gauging impact of financial innovation or its absence. Unfortunately, while these economies may have fewer financial innovations that relate to those more common in Western economies, other differences may preclude them from providing the type of natural experiments that would sharply identify the impacts of innovation.

A second avenue may be the greater exploitation of experimental techniques. A number of efforts have attempted to gauge the consequences of new securities, with an almost exclusive focus on those geared toward the developing countries' poor. Examples of such experimental studies have included assessments of new products such as rainfall insurance (Giné, Townsend, and Vickery 2007; Cole et al. 2009), novel rules for institutions (such as Giné and Karlan's [2009] analysis of microcredit lending rules), and new institutions (for instance, Bhattamishra's 2008 study of rain banks). The focus on such innovations is easy to understand: one can gain statistically meaningful results for a very modest investment. But the methodology could be more generally applied, particularly if researchers were to work in conjunction with financial institutions. One problem with such methodologies, however, is that small-scale experiments are almost surely unable to measure

the systemic costs or benefits that we just highlighted, and are likely to focus primarily on the experience of early adopters.

The same concern—an inability to assess broader externalities—is likely to be a barrier to our third suggested avenue as well: to apply the tools of structural estimation of the social impact of new products to financial innovations. While these models have assessed many classes of product innovations, financial innovations have been largely neglected. But complex dynamics just outlined may make such empirical assessments challenging.

Detailed histories or case studies of financial innovation can offer additional evidence to help uncover the social welfare implications of systemically important new products. By judicious selection of research sites, we can put appropriate attention on innovations that had major impacts on society. The historical or case study approach forces us to examine each innovation in its entirety, both in terms of the full time span of its adoption and the many ripples in the economy.

Finally, the use of counterfactuals—where we invent our own data—perversely may discipline us to be explicit about our implicit assumptions and metrics. The decades of debates over counterfactuals has sensitized us to the need to think in terms of general equilibrium rather than partial effects, to consider complementaries and path dependencies, and to carefully measure outcomes. Despite all of these problems, we believe that this less "scientific" method may add new insights into understanding financial innovation.

References

Adrian, T., and H. Shin. 2010. "The Changing Nature of Financial Intermediation and the Financial Crisis of 2007–09." Federal Reserve Bank of New York Staff Report no. 439. New York: Federal Reserve Bank of New York. http://papers.ssrn.com/so13/papers.cfm?abstract_id=1576590.

Alexander, J. 1991. "Do the Merits Matter? A Study of Settlements in Securities Class Actions." *Stanford Law Review* 43:497–598.

Allen, F., and D. Gale. 1994. *Financial Innovation and Risk Sharing.* Cambridge, MA: MIT Press.

Andrade, G., and S. Kaplan. 1998. "How Costly is Financial (Not Economic) Distress? Evidence from Highly Leveraged Transactions That Became Distressed." *Journal of Finance* 53:1443–93.

Asset Securitization Comptroller's Handbook. 1997. Washington: US Department of Treasury.

Ashcraft, A. B., and T. Schuermann. 2008. "Understanding the Securitization of Subprime Mortgage Credit." Federal Reserve Bank of New York Staff Report no. 318. New York: Federal Reserve Bank of New York.

Axelson, U., P. Strömberg, T. Jenkinson, and M. Weisbach. 2009. "Leverage and Pricing in Buyouts: An Empirical Analysis." EFA 2009 Bergen Meetings Working Paper. http://ssrn.com/abstract=1344023.

Barber, B., and T. Odean. 2000. "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *Journal of Finance* 55:773–806.

———. 2001a. "Boys will be Boys: Gender, Overconfidence, and Common Stock Investment." *Quarterly Journal of Economics* 116:261–92.

———. 2001b. "The Internet and the Investor." *Journal of Economic Perspectives* 15 (1): 41–54.

———. 2002. "Online Investors: Do the Slow Die First?" *Review of Financial Studies* 15:455–87.

———. 2004. "Are Individual Investors Tax Savvy? Evidence from Retail and Discount Brokerage Accounts." *Journal of Public Economics* 88:419–42.

Ben-Horim, M., and W. Silber. 1977. "Financial Innovation: A Linear Programming Approach." *Journal of Banking and Finance* 1:277–96.

Bergstresser, D., J. Chalmers, and P. Tufano. 2009. "Assessing the Costs and Benefits of Brokers: A Preliminary Analysis of the Mutual Fund Industry." *Review of Financial Studies* 22:4129–56.

Berndt, A., and A. Gupta. 2008. "Moral Hazard and Adverse Selection in the Originate-to-Distribute Model of Bank Credit." http://ssrn.com.ezp-prod1.hul.harvard.edu/abstract=1290312.

Berry, S., J. Levinsohn, and A. Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63:841–90.

Bhattamishra, R. 2008. "Grain Banks: An Institutional and Impact Analysis." PhD diss., Cornell University.

Bloom, N., R. Sadun, and J. Van Reenen. 2009. "Do Private Equity-Owned Firms Have Better Management Practices?" In *Globalization of Alternative Investments Working Papers Volume 2: Global Economic Impact of Private Equity 2009,* edited by A. Gurung and J. Lerner, 1–23. New York: World Economic Forum USA.

Bresnahan, T. 1986. "Measuring the Spillovers from Technical Advance: Mainframe Computers in Financial Services." *American Economic Review* 76:742–55.

Bresnahan, T., and M. Trajtenberg. 1995. "General Purpose Technologies 'Engines of Growth'?" *Journal of Econometrics* 65:83–108.

Bunzl, M. 2004. "Counterfactual History: A User's Guide." *American Historical Review* 109:845–58.

Burt, R. 1987. "Social Contagion and Innovation: Cohesion versus Structural Equivalence." *American Journal of Sociology* 92:1287–335.

Campbell, J., H. Jackson, B. Madrian, and P. Tufano. 2010. "The Regulation of Consumer Financial Products: An Introductory Essay with Four Case Studies." http://ssrn.com/abstract=1649647.

Cameron, R., ed. 1972. *Banking and Economic Development.* New York: Oxford University Press.

Cameron, R., O. Crisp, H. Patrick, and R. Tilly. 1967. *Banking in the Early Stages of Industrialization.* New York: Oxford University Press.

Carr, E. 1987. *What is History?* London: Vintage.

Central Intelligence Agency. 2009. *CIA Factbook.* Washington, DC: Government Printing Office.

Chevalier, J. 1995. "Capital Structure and Product Market Competition: An Empirical Study of Supermarket LBOs." *American Economic Review* 85:206–56.

Cole, S., X. Giné, J. Tobacman, P. Topalova, R. Townsend, and J. Vickery. 2009. "Barriers to Household Risk Management: Evidence from India." Federal Reserve Bank of New York Staff Report no. 373. New York: Federal Reserve Bank of New York.

Coleman, J., E. Katz, and H. Menzel. 1966. *Medical Innovation: A Diffusion Study.* New York: Bobbs-Merrill.

Coval, J., J. Jurek, and E. Stafford. 2009. "The Economics of Structured Finance." *Journal of Economic Perspectives* 23:3–25.
Cowan, R., and D. Foray. 2002. "Evolutionary Economics and the Counterfactual Threat." *Journal of Evolutionary Economics* 12:539–62.
Coy, P. 2009. "Financial Innovation Under Fire." *Business Week,* September 16.
Crane, D., K. A. Froot, S. P. Mason, A. F. Perold, R. C. Merton, Z. Bodie, E. R. Sirri, and P. Tufano. 1995. *The Global Financial System.* Boston: Harvard Business School Press.
Dasgupta, P., and J. Stiglitz. 1980. "Industrial Structure and the Nature of Innovative Activity." *Economic Journal* 90:266–93.
David, P. 1969. "Transport Innovation and Economic Growth: Professor Fogel on and off the Rails." *The Economic History Review* 22 (3): 506–25.
Davis, G. 1994. *A History of Money.* Cardiff: University of Wales Press.
Davis, S., J. Haltiwanger, R. Jarmin, J. Lerner, and J. Miranda. 2008. "Private Equity and Employment." In *Globalization of Alternative Investments Working Papers Volume 1: Global Economic Impact of Private Equity 2008,* edited by A. Gurung and J. Lerner, 43–64. New York: World Economic Forum USA.
———. 2009. "Private Equity, Jobs and Productivity." In *Globalization of Alternative Investments Working Papers Volume 2: Global Economic Impact of Private Equity 2009,* edited by A. Gurung and J. Lerner, 25–44. New York: World Economic Forum USA.
Elmer, P., and S. Seelig. 1998. "The Rising Long-Term Trend of Single-Family Mortgage Foreclosure Rates." Federal Deposit Insurance Corporation Working Paper 98-2. http://ssrn.com/abstract=126128 or doi:10.2139/ssrn.126128.
Fang, L., V. Ivashina, and J. Lerner. 2010. "Unstable Equity? Combining Banking with Private Equity Investing." Unpublished Working Paper. Harvard University.
Ferguson, N., ed. 1997. *Virtual History: Alternatives and Counterfactuals.* London: Picador.
Fergusson, Niall. 2000. *Virtual History: Alternatives and Counterfactuals.* New York: Basic Books.
Fink, M. 2008. *The Rise of Mutual Funds: An Insider's View.* New York: Oxford University Press.
Fogel, R. 1964. *Railroads and American Economic Growth: Essays in Econometric History.* Baltimore, MD: Johns Hopkins Press.
Frame, S., and L. White. 2005. "Fussing and Fuming About Fannie and Freddie: How Much Smoke, How Much Fire?" *Journal of Economic Perspectives* 19: 159–84.
Frame, W., and L. White. 2004. "Empirical Studies of Financial Innovation: Mostly Talk and Not Much Action?" *Journal of Economic Literature* 42:116–44.
Frankel, T. 2006. *Securitization,* 2nd ed. New York: Fathom Publishing Company.
French, K. 2008. "Presidential Address: The Cost of Active Investing." *Journal of Finance* 63:1537–73.
Gennaioli, N., A. Shleifer, and R. Vishny. 2010. "Financial Innovation and Financial Fragility." NBER Working Paper no. 16068. Cambridge, MA: National Bureau of Economic Research, June.
Giné, X., and D. Karlan. 2009. "Group versus Individual Liability: Long Term Evidence from Philippine Microcredit Lending Groups." Unpublished Working Paper. World Bank.
Giné, X., R. Townsend, and J. Vickrey. 2007. "Patterns of Rainfall Insurance Participation in Rural India." World Bank Policy Research Working Paper 4408.
Goetzmann, W., and G. Rouwenhorst, eds. 2005. *The Origins of Value: The Financial*

Innovations That Created Modern Capital Markets. New York: Oxford University Press.
Goldin, C. 1995. "Cliometrics and the Nobel." *The Journal of Economic Perspectives* 9 (2): 191–208.
Gorton, G., and A. Metrick. 2010. "Regulating the Shadow Banking System." Yale School of Management Working Paper. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1676947.
Granovetter, M. 1978. "Threshold Models of Collective Behavior." *American Journal of Sociology* 83:420–43.
Greenspan, A. 1995. "Remarks at a Conference on Risk Measurement and Systemic Risk." Board of Governors of the Federal Reserve System, Washington, DC.
Hausman, J. 1997. "Valuing the Effect of Regulation on New Services in Telecommunications." *Brookings Papers on Economic Activity: Microeconomics* 26:1–39.
Hellmann, T., and M. Puri. 2002. "Venture Capital and the Professionalization of Start-Up Firms: Empirical Evidence." *Journal of Finance* 57:69–97.
Helpman, E., ed. 1998. *General Purpose Technologies and Economic Growth.* Cambridge, MA: MIT Press.
Hendershott, P., and J. Shilling. 1989. "The Impact of Agencies on Conventional Fixed-Rate Mortgage Yields." *Journal of Real Estate Finance and Economics* 2:101–15.
Hoberg, G., B. Goldfarb, D. Kirsch, and A. Triantis. 2009. "Does Angel Participation Matter? An Analysis of Early Venture Financing." Robert H. Smith School Research Paper no. RHS 06-072. http://ssrn.com/abstract=1024186.
Hoffman, M., and T. Nitschka. 2008. "Securitization of Mortgage Debt, Asset Prices, and International Risk Sharing." Working Paper Series, Institute for Empirical Research in Economics, University of Zurich.
Jameson, M., S. Dewan, and C. Sirmans. 1992. "Measuring Welfare Effects of 'Unbundling' Financial Innovations: The Case of Collateralized Mortgage Obligations." *Journal of Urban Economics* 31:1–13.
Jensen, M. 1989. "The Eclipse of the Public Corporation." *Harvard Business Review* 67:61–74.
Jiangli, W., and M. Pritsker. 2008. "The Impacts of Securitization on US Bank Holding Companies." SSRN Working Paper Series. http://ssrn.com/abstract=1102284.
Johnson, S., and J. Kwok. 2009. "More Financial Innovation." http://baselinescenario.com/2009/06/17/more-financial-innovation.
Kane, E. 1977. "Good Intentions and Unintended Evil: The Case Against Selective Credit Allocation." *Journal of Money, Credit and Banking* 9:55–69.
Kaplan, S., and A. Schoar. 2005. "Private Equity Performance: Returns, Persistence and Capital Flows." *Journal of Finance* 60:791–823.
Kaplan, S., and J. Stein. 1993. "The Evolution of Buyout Pricing and Financial Structure in the 1980s." *Quarterly Journal of Economics* 108:313–57.
Khandani, A., A. Lo, and R. Merton. 2009. "Systemic Risk and the Refinancing Ratchet Effect." Harvard Business School Working Paper no. 10-023.
Khorana, A., H. Servaes, and P. Tufano. 2005. "Explaining the Size of the Mutual Fund Industry Around the World." *Journal of Financial Economics* 78:145–85.
Kindleberger, C. 1984. *A Financial History of Western Europe.* London: George Allen and Unwin.
Kortum, S., and J. Lerner. 2000. "Assessing the Contribution of Venture Capital to Innovation." *Rand Journal of Economics* 31:674–92.
Krugman, P. 2007. "Innovating Our Way to Financial Crisis." *New York Times,* December 3.

Lamoreaux, N., M. Levenstein, and K. Sokoloff. 2007. "Financing Invention During the Second Industrial Revolution: Cleveland, Ohio, 1870–1920." In *Financing Innovation in the United States, 1870 to the Present,* edited by N. Lamoreaux and K. Sokoloff, 39–84. Cambridge, MA: MIT Press.

Lerner, J. 2002. "Where Does *State Street* Lead? A First Look at Finance Patents, 1971–2000." *Journal of Finance* 57:901–30.

———. 2006. "The New New Financial Thing: The Origins of Financial Innovations." *Journal of Financial Economics* 79:233–55.

———. 2010. "The Litigation of Financial Innovations." *Journal of Law and Economics* 53:807–31.

Lerner, J., M. Sorensen, and P. Stromberg. 2008. "Private Equity and Long-Run Investment: The Case of Innovation." NBER Working Paper no. 14623. Cambridge, MA: National Bureau of Economic Research, December.

Litan, R. 2010. "In Defense of Much, But Not All, Financial Innovation." Unpublished Working Paper. Brookings Institution.

Loutskina, E., and P. Strahan. 2009. "Securitization and the Declining Impact of Bank Finance on Loan Supply: Evidence from Mortgage Originations." *Journal of Finance* 64:861–89.

Mason, S., R. Merton, A. Perold, and P. Tufano. 1995. *Cases in Financial Engineering: Applied Studies of Financial Innovation.* Englewood Cliffs, NJ: Prentice Hall.

Mayer, C., K. Pence, and S. Sherlund. 2009. "The Rise in Mortgage Defaults." *Journal of Economic Perspectives* 23:27–50.

McClelland, P. 1968. "Railroads, American Growth, and the New Economic History: A Critique." *The Journal of Economic History* 28 (1): 102–23.

Merton, R. 1992. "Financial Innovation and Economic Performance." *Journal of Applied Corporate Finance* 4:12–22.

Mian, A., and A. Sufi. 2008. "The Consequences of Mortgage Credit Expansion: Evidence from the 2007 Mortgage Default Crisis." NBER Working Paper no. 13936. Cambridge, MA: National Bureau of Economic Research, April.

Michalak, T., and A. Uhde. 2010. "Credit Risk Securitization and Banking Stability: Evidence from the Micro-Level for Europe." University of Bochum Working Paper Series. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1383469.

Michalopoulos, S., L. Laeven, and R. Levine. 2010. "Financial Innovation and Endogenous Growth." Unpublished Working Paper. Brown University.

Miller, M. 1986. "Financial Innovation: The Last Twenty Years and the Next." *Journal of Financial and Quantitative Analysis* 21:459–71.

Mishra, P. 2010. "Financial Innovation in Financial Markets: A Reassessment." Unpublished Working Paper. Indian Institute of Technical Education and Research.

Mollica, M., and L. Zingales. 2007. "The Impact of Venture Capital on Innovation and the Creation of New Business." Unpublished Working Paper. University of Chicago.

Murray, G. 1998. "A Policy Response to Regional Disparities in the Supply of Risk Capital to New Technology-Based Firms in the European Union: The European Seed Capital Fund Scheme." *Regional Studies* 32:405–19.

Nerlove, M. 1966. "Review: Railroads and American Economic Growth." *The Journal of Economic History* 26 (1): 107–15.

Odean, T. 1999. "Do Investors Trade Too Much?" *American Economic Review* 89: 1279–98.

Panetta, F., and A. Pozzolo. 2010. "Why Do Banks Securitize Their Assets? Bank-Level Evidence from Over One Hundred Countries." Bank of Italy Working Paper. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1572983.

Persons, J., and V. Warther. 1997. "Boom and Bust Patterns in the Adoption of Financial Innovations." *Review of Financial Studies* 10:939–67.
Petrin, A. 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy* 110:705–29.
Purnanandam, A. 2009. "Originate-to-Distribute Model and the Subprime Mortgage Crisis." Paper presented at the American Finance Association meeting. Atlanta, Georgia, January 3–5, 2010. http://ssrn.com.ezp-prod1.hul.harvard.edu/abstract=1167786.
Rogers, E. 1962. *Diffusion of Innovations.* New York: Free Press.
Rosen, R. 2007. "The Role of Securitization in Mortgage Lending." *Chicago Fed Letter: Essays on Issues,* no. 244, November.
Ross, S. 1976. "Options and Efficiency." *Quarterly Journal of Economics* 90:75–89.
Ryan, A., G. Trumbull, and P. Tufano. 2010. "A Brief Post-War History of Consumer Finance." Unpublished Manuscript.
Sarkisyan, A., B. Casu, A. Clare, and S. Thomas. 2010. "Securitization and Bank Performance." Sir John Cass Business School Working Paper Series, City University London. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1490925.
Schwarcz, S. 2008. "Systemic Risk." *Georgetown Law Journal* 97:193–249.
Shiller, R. 2008. *The Subprime Solution: How Today's Global Financial Crisis Happened and What to Do About It.* Princeton, NJ: Princeton University Press.
Silber, W., ed. 1975. *Financial Innovation.* Lexington: Lexington Books.
———. 1983. "The Process of Financial Innovation." *American Economic Review Papers and Proceedings* 73:89–95.
Sirmans, C., and J. Benjamin. 1990. "Pricing Fixed Rate Mortgages: Some Empirical Evidence." *Journal of Financial Services Research* 4:191–202.
Sirri, E., and P. Tufano. 1995. "The Economics of Pooling." In *The Global Financial System,* edited by D. Crane, 81–128. Boston: HBS Press.
Slater, R. 1997. *John Bogle and the Vanguard Experiment.* Chicago: Irwin.
Stein, J. 1997. "Internal Capital Markets and the Competition for Corporate Resources." *Journal of Finance* 52:11–33.
Strömberg, P. 2008. "The New Demography of Private Equity." In *Globalization of Alternative Investments Working Papers Volume 1: Global Economic Impact of Private Equity 2008,* edited by A. Gurung and J. Lerner, 3–26. New York: World Economic Forum USA.
Sylvan, D., and S. Majeski. 1998. "A Methodology for the Study of Historical Counterfactuals." *International Studies Quarterly* 42:79–108.
Tetlok, P., and R. Lebow. 2001. "Poking Counterfactual Holes in Covering Laws: Cognitive Styles and Historical Reasoning." *American Political Science Review* 95:829–43.
Trajtenberg, M. 1990. "A Penny for Your Quotes: Patent Citations and the Value of Inventions." *Rand Journal of Economics* 21:172–87.
Tufano, P. 1989. "Financial Innovation and First-Mover Advantages." *Journal of Financial Economics* 25:213–40.
———. 1995. "Securities Innovations: A Historical and Functional Perspective." *Journal of Applied Corporate Finance* 7:90–103.
———. 2003. "Financial Innovation." In *Handbook of the Economics of Finance, Volume 1a: Corporate Finance,* edited by G. Constantinides, M. Harris, and R. Stulz, 307–36. New York: Elsevier.
Tufano, P., and B. Kyrillos. 1994. "Leland O'Brien Rubinstein Associates Incorporated: SuperTrust." Case Study 294-050: Harvard Business School.